

# Modellbasierte Poseschätzung von Menschen aus dichten Volumendaten

Diplomarbeit  
zur Erlangung des Grades  
DIPLOM-INFORMATIKER  
im Studiengang Computervisualistik

vorgelegt von

Benjamin Knopp

Betreuer: Dipl.-Inform. Tobias Feldmann, Institut für Anthropomatik,  
Arbeitsgruppe Human Motion Analysis, KIT  
Erstgutachter: Dr.-Ing. Dipl. Math. Annika Wörner, Institut für Anthropomatik,  
Arbeitsgruppe Human Motion Analysis, KIT  
Zweitgutachter: Prof. Dr.-Ing. Dietrich Paulus, Institut für Computervisualistik,  
Arbeitsgruppe Aktives Sehen, Universität Koblenz-Landau

Koblenz, im Mai 2010



## Kurzfassung

Im Rahmen dieser Diplomarbeit wird ein Verfahren zur markerlosen Pose-Rekonstruktion vorgestellt. Die Modellierung des menschlichen Körpers geschieht auf Basis der Starrkörperphysik. Mittels eines probabilistischen Ansatzes wird das Modell in eine volumetrische Rekonstruktion der Szene eingepasst. Die hierfür zu bewältigende Suche in hochdimensionalen Zustandsräumen wird mittels eines *Partikelfilters* in Kombination mit *Simulierter Abkühlung* vorgenommen. Eine Berücksichtigung anthropometrischer Besonderheiten sowie kinematischer Grenzen wird zur weiteren Stützung des Verfahrens vorgenommen. Die vollständige Umsetzung des Verfahrens durch dreidimensionale Beobachtungs- und Messmodelle führt zu einer exakten Poserekonstruktion und vermeidet Mehrdeutigkeiten während der Auswertung.

## Abstract

This diploma thesis is introducing a method for markerless tracking. The human body is described by a physics based rigid body chain. Using a probabilistic approach the model is fitted into a volumetric reconstruction of the scene. To cope with the high dimensional search space, a particle filter in combination with simulated annealing is used. Anthropometric characteristics and kinematic limits will also be incorporated. The complete implementation is based on a three dimensional observation and measurement model, which leads to a more accurate pose reconstruction and avoids ambiguities.



## Erklärung

Ich versichere, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat und von dieser als Teil einer Prüfungsleistung angenommen wurde. Alle Ausführungen, die wörtlich oder sinngemäß übernommen wurden, sind als solche gekennzeichnet.

Die Vereinbarung der Arbeitsgruppe für Studien- und Abschlussarbeiten habe ich gelesen und anerkannt, insbesondere die Regelung des Nutzungsrechts.

Mit der Einstellung dieser Arbeit in die Bibliothek bin ich einverstanden. ja  nein

Der Veröffentlichung dieser Arbeit im Internet stimme ich zu. ja  nein

Koblenz, den 31. März 2010

# Inhaltsverzeichnis

Abbildungsverzeichnis	5
Tabellenverzeichnis	9
Algorithmenverzeichnis	10
<b>1 Einführung</b>	<b>13</b>
1.1 Aufbau dieser Arbeit . . . . .	15
1.2 Arbeitsthese . . . . .	16
<b>2 Grundlagen</b>	<b>17</b>
2.1 Starrkörperphysik . . . . .	17
2.1.1 Gelenkanschlüsse . . . . .	18
2.1.2 Inverse Kinematik . . . . .	19
2.2 Volumetrische Rekonstruktion aus Kameradaten . . . . .	19
2.2.1 Vordergrund- & Hintergrundmodelle . . . . .	20
2.2.2 3D-Rekonstruktion . . . . .	20
2.3 Bayes Filter . . . . .	21
2.3.1 Bayes-Theorem . . . . .	21
2.3.2 Problemformulierung . . . . .	21
2.3.3 Kalman-Filter . . . . .	23
2.3.4 Partikelfilter . . . . .	23
2.4 Komponenten eines Pose-Schätzers . . . . .	25
<b>3 Stand der Technik</b>	<b>29</b>
3.1 Menschmodelle . . . . .	30
3.1.1 Dynamische Aspekte . . . . .	32
3.1.2 Modelladaption . . . . .	32
3.2 Prädiktion und Modelleinpassung . . . . .	33
3.2.1 Optimierungen . . . . .	34
3.3 Beobachtungsraum und Sensordaten . . . . .	36
3.4 Taxonomie markerloser Trackingverfahren . . . . .	36
<b>4 Eigener Ansatz zur Pose-Rekonstruktion</b>	<b>39</b>

4.1	Anwendungskontext und Vorüberlegungen . . . . .	39
4.2	Aufbau des Verfahrens . . . . .	40
4.3	Körpermodell: volumetrische Starrkörper . . . . .	42
4.3.1	Koordinatensysteme . . . . .	43
4.3.2	Rasterisierung . . . . .	45
4.3.3	Behandlung von Winkeln und Gelenkanschlügen . . . . .	46
4.4	Simulierte Abkühlung und der Annealed ParticleFilter . . . . .	46
4.4.1	Glättung der Gewichtsfunktionen . . . . .	47
4.4.2	Ablauf eines Zeitschritts . . . . .	47
4.4.3	Laufzeitparameter . . . . .	50
4.4.4	Übergang zum nächsten Zeitschritt . . . . .	52
4.4.5	Diffusionsstrategien zur Partikelbewegung . . . . .	53
4.4.6	Suche in baumartigen Topologien . . . . .	54
4.5	Weitere Verfahren zur Reduktion des Suchraumes . . . . .	54
4.5.1	Harte Suchraumpartitionierung . . . . .	54
4.5.2	Optimierte Partikelstreuung . . . . .	55
4.6	Dreidimensionales Belegungsgitter als Beobachtungsraum . . . . .	56
4.6.1	Symmetrisches Messmodell . . . . .	57
<b>5</b>	<b>Experimente und Validierung</b>	<b>59</b>
5.1	Evaluation von Einzelkomponenten des Partikelfilters . . . . .	59
5.1.1	Diffusionsstrategien . . . . .	60
5.1.2	Einfluss der Abstandsfunktion . . . . .	61
5.1.3	Partikelselektion . . . . .	63
5.1.4	Suchraumpartitionierung . . . . .	63
5.2	Wahl der Laufzeit-Parameter . . . . .	64
5.2.1	Überlebensrate . . . . .	64
5.2.2	Layer und Partikelzahl . . . . .	65
5.3	Verfahren zur Messung und Bewertung der Übereinstimmung . . . . .	67
5.3.1	Visuelle Übereinstimmung als Goldstandard . . . . .	68
5.3.2	Vergleich von Gelenkpositionen . . . . .	68
5.4	Ergebnisse der Pose-Rekonstruktion . . . . .	69
5.5	Experimente zur Modellinitialisierung . . . . .	72
<b>6</b>	<b>Schlussbetrachtung</b>	<b>73</b>
6.1	Fazit . . . . .	73
6.2	Ausblick . . . . .	74
<b>A</b>	<b>Implementierung</b>	<b>77</b>
A.1	Verwendete Vorarbeiten . . . . .	77
A.2	Architektur . . . . .	77

A.3	Arbeitsumgebung und Werkzeuge . . . . .	79
A.3.1	Die Arbeitsumgebung . . . . .	79
A.3.2	Der Modell-Editor . . . . .	80
<b>B</b>	<b>Mathematische Grundlagen</b>	<b>81</b>
B.1	Allgemein . . . . .	81
B.1.1	Boltzmann-Statistik . . . . .	81
B.1.2	Markov-Kette . . . . .	81
B.1.3	Metropolis-Algorithmus . . . . .	82
<b>C</b>	<b>Glossar</b>	<b>83</b>
	<b>Weiterführende Literatur zu wissenschaftlichen Ausarbeitungen</b>	<b>87</b>





# Abbildungsverzeichnis

1.1	Markerbasiertes Tracking: Turnerin mit VICON <sup>©</sup> -Markern. . . . .	15
2.1	Starrkörperkette: schematische Darstellung einiger Geometrie- und Transformationsknoten . . . . .	18
2.2	Zusammenhang: Zustandsvektor – Gelenkkoordinaten . . . . .	19
3.1	Skelett-Modell des menschlichen Körpers. Quelle: Wikimedia, Hubert Ludwig, Schul - Naturgeschichte 1891 . . . . .	33
4.1	Volumenrekonstruktion: Aufbau eines Belegungsgitters mittels Rückprojektion der Vordergrund-Masken in den Raum. . . . .	40
4.2	Datenflussdiagramm: Rekonstruktion, Vorhersage und Evaluation. . . .	41
4.3	Modell-Editor: Aufbau einer Starrkörperkette und Einstellung der Freiheitsgrade . . . . .	43
4.4	Aufbau der Starrkörper-Kette . . . . .	44
4.5	Vergleich: Starrkörpermodell und Voxeldarstellung . . . . .	45
4.6	Zeitschritt eines Annealed Particle Filters: Verlauf der Partikelstreuung, Grafik aus Deutscher und Reid [DR03]. . . . .	48
4.7	Partikelbewegung durch Simulierte Abkühlung, 2 rotatorische Freiheitsgrade des 1. Starrkörpers (x- und y-Achse). . . . .	50
4.8	Partitionierungsschema zur Vermeidung von iterativen Fehlern. . . . .	55
4.9	Gegenüberstellung verschiedene Beobachtungsräume . . . . .	56
5.1	Darstellung des multimodalen Suchraums: Rotation des Modells um die (a) x- und (b) y-Achse. . . . .	60
5.2	Gegenüberstellung der Diffusionsstrategien: kovarianzbasierte, varianzbasierte und konstant (abnehmende) Diffusion. . . . .	61
5.3	Abstandsmaße ohne (a-c) und mit Überlappungserkennung (d-f), von links nach rechts: Partikelgewichtung, finale Pose, Gewichtungsexponenten	62
5.4	Gegenüberstellung zweier Strategien zur Partikelselektion: konventionelle Selektion und Selektion mit zusätzlichem Transfer der vorhergehenden Optima. . . . .	63
5.5	Vergleich von globaler Pose-Optimierung und partitionierter Suche. . .	64
5.6	Gewichtsfunktion in Abhängigkeit von der gewählten Abkühlrate. . . .	65

5.7	Gewichtung, Gewichtungsexponent und Überlebensrate für $(p = 1000, l = 20)$ , $(p = 500, l = 20)$ , $(p = 250, l = 40)$ sowie $(p = 125, l = 80)$ , erwartete Überlebensrate= 0.25. . . . .	66
5.8	Detektionsergebnisse Karatekämpfer: Frame 0 – 1000 . . . . .	69
5.9	Chronologische Darstellung der bestimmten Modell-Konfigurationen. . . . .	70
5.10	Frame 450: Fehlzuordnung und Überlappung der Beine. . . . .	70
5.11	Ergebnisse des markerbasierten Vergleiches, Visualisierung der Gelenkpunkte (links) und Abweichungen pro Gelenkpunkt in Zentimetern (rechts). . . . .	71
5.12	Erfolgreiche Modellausrichtung bei schlechter Initialpose . . . . .	72
A.1	UML-Klassendiagramm der Starrkörper-Komponenten . . . . .	78
A.2	Projektansicht der Wombat-Arbeitsumgebung . . . . .	79
A.3	Projektansicht der Wombat-Arbeitsumgebung . . . . .	80

# Tabellenverzeichnis

2.1	Gelenkwinkelausmaße nach [Ste09] . . . . .	18
3.1	Verschiedene Einsatzgebiete für Trackinganwendungen. . . . .	29
3.2	Übersicht Tracking-Verfahren, nach aufsteigender Komplexität sortiert. . . . .	31
3.3	Taxonomie wegweisender Trackingverfahren . . . . .	37



# Algorithmenverzeichnis

1	Zwischenschritte eines SIR-Filters für einen diskreten Zeitschritt . . . .	24
2	Iterationsablauf bei Simulierter Abkühlung . . . . .	49



# Kapitel 1

## Einführung

### Kurzbeschreibung

Die Zahl der Szenarien, in welchen die Beziehung Mensch - Computer nicht mehr auf die eines Instruierenden und einer ausführenden Maschine beschränkt ist, steigt seit Jahren kontinuierlich. In unterschiedlichsten Einsatzgebieten ist ein genaues Erfassen und Verstehen des menschlichen Bewegungsapparates (Motion-Capturing) erforderlich. So werden im medizinischen Kontext häufig in Form von Videosequenzen aufgezeichnete Bewegungsabläufe auf Gelenkfehlstellungen hin untersucht, während im Augmented-Reality Bereich sowie in der Robotik die Übertragbarkeit menschlicher Bewegungen auf virtuelle Modelle oder Roboter im Vordergrund steht. Auch im Gebiet der Mensch-Maschine Interaktion spielt Gesten- und Bewegungserkennung eine zunehmend bedeutende Rolle.

Im Rahmen dieser Arbeit wird ein Verfahren zur Rekonstruktion der menschlichen Pose vorgestellt, das auf einem modellbasierten, probabilistischen Ansatz beruht. Es werden volumetrische Datensätze verwendet, welche die zu rekonstruierende Pose über die Zeit enthalten. In vielen der Eingangs erwähnten Anwendungsfällen entspricht die Verwendung von markerbasierten Systemen dem Stand der Technik. Hierbei werden leicht interpretierbare Messpunkte am Körper befestigt, welche z. B. optisch oder elektrisch ausgelesen werden können. Dieses Vorgehen beschränkt die Anwendbarkeit auf im Vorfeld entsprechend präparierte Aufnahmesituationen und führt zu einer Abhängigkeit von komplexen Hardware- / Software-Systemen. Weiterhin sind die erzielten Ergebnisse von der exakten Positionierung der verwendeten Marker abhängig. Steigende Rechenleistung und die zunehmende Erschließung von Szenarien, in welchen die Anbringung von Markern nicht erwünscht oder technisch nicht umsetzbar ist, lassen Verfahren zum markerlosen Tracking attraktiver erscheinen.

Das allgemeine Vorgehen besteht darin, ein parametrisierbares Modell des menschlichen Körpers mit Beobachtungen im Bildraum derart zur Übereinstimmung zu bringen, dass die Unterschiede zwischen Modell und Beobachtung minimiert werden.



Für die Modellierung des menschlichen Körpers wird häufig auf die Simulation von Starrkörperketten zurückgegriffen, was sich zum Quasi-Standard in der hier beschriebenen Domäne entwickelt hat.

Der in dieser Arbeit gewählte Ansatz verfolgt das Ziel, sowohl die Pose-Schätzung als auch die Abstandsfunktion vollständig im dreidimensionalen Raum anzusiedeln.

Die hierfür notwendigen 3-D-Volumendaten können auf verschiedene Art und Weise generiert werden. Während die verwendeten Testdaten durch ein Mehrkammersystem gewonnen wurden, ist auch der Einsatz von Laserscannern und den immer populärer werdenden *time-of-flight* Kameras möglich.

Da auch bei entsprechend hoher zeitlicher Auflösung eine deterministische Rekonstruktion komplexer Bewegungsabläufe - ohne den Bewegungsablauf einschränkende Annahmen - kaum zu gewährleisten ist, wird im Rahmen dieser Arbeit ein probabilistischer Ansatz verfolgt.

Das verwendete Körpermodell basiert auf [Ste09] und erlaubt eine genaue, anatomisch korrekte Rekonstruktion, sowohl der auftretenden Gelenkwinkel, als auch der hieraus resultierenden Kräfte. Es liefert eine wirklichkeitsgetreue Modellierung der verschiedenen menschlichen Gelenktypen (Drehgelenke, Kugelgelenke etc.), welche die zugehörigen Bewegungsräume exakt abbilden, und ermöglicht daher eine intuitive räumliche Ausrichtung.

Um das Modell mit einem Volumendatensatz vergleichen zu können, ist zusätzlich eine Beschreibung der Volumina einzelner Starrkörper notwendig. Üblicherweise werden hierfür einfache geometrische Beschreibungen wie Zylinder oder polygonale Netze verwendet. Durch die dichte Rekonstruktion<sup>1</sup> ist es möglich, auf dreidimensionale Überlappungsmaße zurückzugreifen und die volumetrische Überschneidung zu messen. Ein möglichst allgemeingültiges Menschmodell, welches die Freiheitsgrade der entsprechenden Körperteile wirklichkeitsgetreu abbildet, besitzt eine derart hohe Dimensionalität, dass eine globale Betrachtung des hieraus resultierenden Suchraumes einen beachtlichen Aufwand zur Folge hat. Da die Aufenthaltswahrscheinlichkeit nicht für jede denkbare Parametrisierung ermittelt werden kann, werden die Modi der zugrundeliegenden Verteilungsfunktion mittels eines Partikelfilters [IB98] ermittelt. Die Anzahl der benötigten Partikel und die Dimension des betrachteten Suchraums stehen in einem direkten Zusammenhang, so dass im konkreten Fall eine harte Partitionierung des eigentlichen Raumes in Subräume niedriger Dimensionalität vorgenommen wird, um eine deutliche Reduktion der benötigten Partikel zu ermöglichen [BB09]. Hieraus resultiert eine hierarchische Suche der optimalen Modellposition, wobei der Einfluss unterschiedlicher Traversierungsreihenfolgen zu untersuchen ist. (So führt z.B. eine unzureichende Ausrichtung des Rumpfes zwangsläufig auch zu einer fehlerhaften Ausrichtung der Arme etc.)

Zum einen bietet der gewählte Ansatz den Vorteil, dass Ambiguitäten, wie sie durch die Auswertung in einer geringen Anzahl von 2D Ansichten auftreten können [CH04], vermieden werden, zum anderen ermöglicht es der Vergleich dichter volu-

---

<sup>1</sup>im Vergleich z. B. zur konvexen Hülle, wird der Körper als durchgehend gefüllt betrachtet

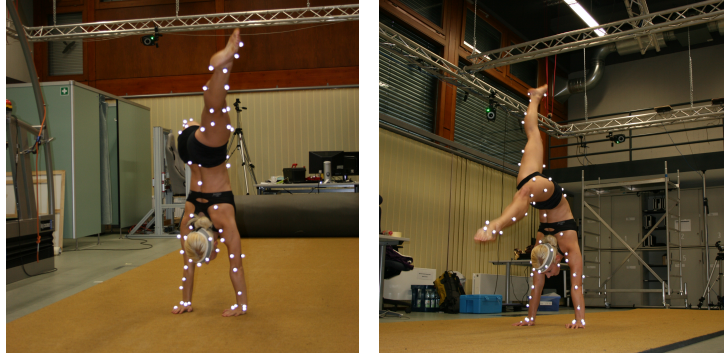


Abbildung 1.1: Markerbasiertes Tracking: Turnerin mit VICON<sup>®</sup>-Markern.

metrischer Daten auf die fehleranfällige Betrachtung von Silhouetten oder Kanten zur Ähnlichkeitsbestimmung zu verzichten. Zur Evaluation der rekonstruierten Bewegungsabläufe wird auf modellierte Idealposen und Aufnahmesequenzen zurückgegriffen, welche unter zusätzlicher Verwendung optischer Marker angefertigt wurden (siehe Abbildung 1.1). Während die hierdurch erhaltenen Zusatzinformationen für den eigentlichen Rekonstruktionsprozess unberücksichtigt bleiben, bilden die mittels eines VICON<sup>®</sup>-Systems errechneten Modellposen eine Vergleichsbasis.

Die manuell erstellten Idealposen erlauben eine genauere Messung des durch das Vorhersageverfahren bedingten Fehlers, da Abweichungen in der Volumenrekonstruktion oder der Markerpositionierung unberücksichtigt bleiben. Die genaue Kenntnis der absoluten Gelenkpositionen sowie auch der jeweiligen Gelenkwinkel erlaubt eine fundierte abschließende Beurteilung der erzielten Ergebnisse. Neben dem Aspekt, inwiefern sich der Suchaufwand durch die Integration von Vorwissen reduzieren läßt, werden auch die Auswirkungen verschiedener Abstandsmaße und Volumenmodelle auf die geometrische Genauigkeit der Ergebnisse betrachtet.

## 1.1 Aufbau dieser Arbeit

Die Arbeit ist im Weiteren wie folgt organisiert: In Kapitel 2 werden die notwendigen Grundlagen zu den erforderlichen Einzelkomponenten eines Frameworks zur Pose-Rekonstruktion vorgestellt. Hierzu zählen insbesondere die Modelle zur Beobachtung und Vorhersage des ebenfalls an dieser Stelle eingeführten Mensch-Modells. Kapitel 3 gibt einen Überblick bezüglich einiger in diesem Bereich geleisteten Vorarbeiten und stellt deren Vorgehensweisen und Besonderheiten vor. Das Kapitel schließt mit einer Taxonomie, welche die Verfahren anhand der verwendeten Komponenten klassifiziert. Der in dieser Arbeit gewählte Ansatz wird im folgenden Kapitel 4 vorgestellt. Die Gemeinsamkeiten und Unterschiede zu den zuvor beschriebenen Verfahren werden erörtert und die Wahl der Einzelkomponenten begründet. Die erzielten Ergebnisse werden im 5. Kapitel präsentiert. Neben einer Analyse der Laufzeit- und Modellparameter werden qualitative und quantitative Daten zur Beurteilung der ermittelten Modellkonfigura-

trionen vorgestellt sowie die hierzu durchgeführten Experimente erörtert. Abschließend werden in Kapitel 6 die Ergebnisse zusammengefasst und beurteilt. Im Anhang sind die wesentlichen Begrifflichkeiten in Form eines Glossars aufgeführt. Weiterhin werden mathematische Hintergründe sowie Details zur Implementation der im Rahmen dieser Arbeit entstandenen Entwicklungs- und Evaluationsplattform *wombat* (**w**orking environment for **m**odel **b**ased **t**racking) präsentiert.

## 1.2 Arbeitsthesen

1. Wie in Kapitel 3 zu sehen sein wird, entspricht die Verwendung von zweidimensionalen Kamerabildern bei der Pose-Rekonstruktion dem heutigen Standard. Der Modellvergleich findet in der Bildebene statt – bei Mehrkamarasystemen durch Fusion der Einzelbewertungen. Untersuchungen zu der Anzahl verwendeter Kamerapositionen haben gezeigt, dass bei dieser Vorgehensweise mehr als drei Kameras keinen signifikanten Mehrwert bringen. Der in dieser Arbeit gewählte volumetrische Modellvergleich arbeitet sowohl mit Kamera-basierten Volumenrekonstruktionen, welche durch eine erhöhte Kamerazahl genauere Ergebnisse liefern, als auch prinzipiell mit anderen 3-D Datenquellen, wie *time-of-flight* Kameras oder Laserscannern zusammen. *Diese Vorgehensweise bietet Vorteile bezüglich der Genauigkeit und Robustheit des Verfahrens.*
2. *Die Integration von a priori Modellwissen kann helfen, den Suchaufwand erheblich zu reduzieren.* Hierzu können sowohl statische Informationen, wie Gelenkwinkelanschlüsse, als auch dynamische Informationen, wie Schranken für zeitliche Veränderungen herangezogen werden.
3. *Neben der Reduktion der Wertebereiche ergibt sich durch die Zerlegung des Ausgangsproblems in weniger komplexe Teilprobleme weiteres Einsparpotential.* Die Vor- und Nachteile einer in diesem Zusammenhang betrachteten Suchraumpartitionierung werden untersucht.
4. Darüber hinaus wird in Form von *Simulierter Abkühlung* ein heuristisches Suchverfahren in Kombination mit Partikelfiltern eingesetzt. *Hieraus ergibt sich eine deutliche Reduzierung der Partikelmenge.* Die Auswirkungen dieser Vorgehensweise sowie eventuell auftretende Probleme werden betrachtet. Eine Auswahl an Verfahren zur Steuerung der Abkühlrate, sowie Vorschläge zur Wahl der Laufzeitparameter werden präsentiert.

# Kapitel 2

## Grundlagen

Die Grundkomponenten, welche sowohl die Basis des in Kapitel 4 vorgestellten Ansatzes bilden, als auch beim Verständnis der Bestandsaufnahme in Kapitel 3 helfen, werden in den folgenden Abschnitten vorgestellt.

### 2.1 Starrkörperphysik

Mehr als 100 Gelenke und 750 Muskeln halten die Knochen des menschlichen Körpers zusammen – entsprechend komplex gestaltet sich die Modellierung. Die Ansatzpunkte der Muskulatur liegen außerhalb der Gelenkzentren, zudem sind schon an einfachsten Bewegungen eine Vielzahl von Muskelgruppen beteiligt. Da sowohl das aus starren Elementen bestehende Skelett, als auch die komplette Muskulatur nach außen hin nicht sichtbar sind, sind allein aufgrund visueller Beobachtungen die tatsächlichen Muskelbewegungen nicht rekonstruierbar. Für viele Einsatzzwecke hat sich die Modellierung durch Starrkörperketten als hinreichend gute Approximation durchgesetzt. Grundprinzip der Starrkörperphysik ist es, komplexe Körpermodelle aus einer Reihe starrer Körper zu modellieren, welche topologisch geordnet und durch Gelenkpunkte miteinander verbunden sind.

Für die Modellierung von menschlichen Körpern erweist sich eine Baumstruktur, wie sie auszugsweise in Abbildung 2.1 zu sehen ist, als optimale Topologie. Das Modell wird über die Transformation des Wurzelementes in der Welt positioniert, die Gelenkwinkel entsprechen den jeweiligen Rotationen an entsprechender Stelle des Baumes. Innerhalb des Baumes wird von einer festen Bindung der einzelnen Starrkörper aneinander ausgegangen, die Translation der Starrkörper zueinander wird folglich als fest angenommen. Die Position jedes Festkörpers ergibt sich durch Akkumulation der übergeordneten Transformationsknoten.

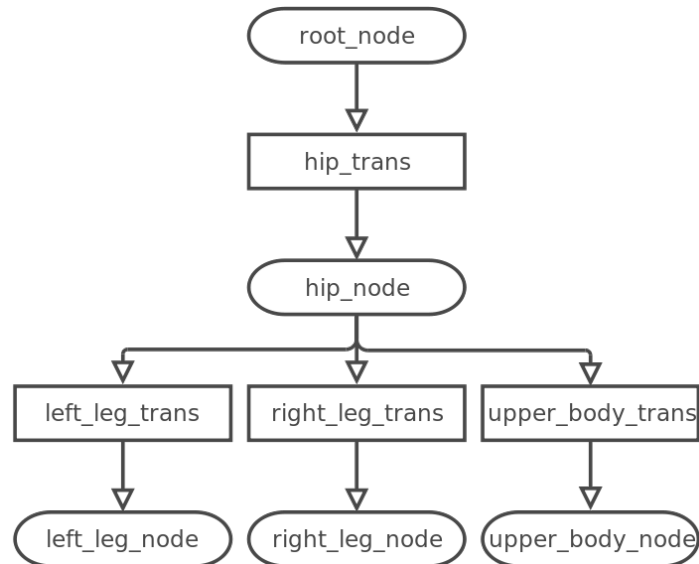


Abbildung 2.1: Starrkörperkette: schematische Darstellung einiger Geometrie- und Transformationsknoten

### 2.1.1 Gelenkanschläge

Würde man jedes Starrkörper-Gelenk mit vollen drei rotatorischen Freiheitsgraden modellieren, so ließen sich Modell-Konfigurationen erzeugen, welche der menschliche Körper nicht einnehmen kann. Weiterhin würde auch die Dimensionalität des Parameterraumes unnötig stark steigen. Betrachtet man etwa Ellenbogen oder Kniegelenke, so ergibt sich nur ein einziger Freiheitsgrad, jeweils mit einer Rotationsachse orthogonal zur Längsachse des Körpers. Neben der Anzahl der Freiheitsgrade sind auch die Gelenkwinkelanschläge von entscheidender Bedeutung. Entlang der meisten Dimensionen ist, wie in Tabelle 2.1 zu sehen ist, eine Bewegung um weniger als 180 Grad möglich.

<b>Gelenk</b>	<b>1. Achse</b>	<b>2. Achse</b>	<b>3. Achse</b>
Hüftgelenk	145°	70°	85°
Kniegelenk	130°	-	-
Sprunggelenk	70°	45°	-
Schlüsselbein (innen)	40°	40°	-
Schlüsselbein (außen)	40°	40°	40°
Schulter	155°	130°	90°
Ellenbogen	160°	180°	-
Handgelenk	120°	50°	-

Tabelle 2.1: Gelenkwinkelausmaße nach [Ste09]

## 2.1.2 Inverse Kinematik

Die *inverse Kinematik*, oft auch als Rückwärtstransformation bezeichnet, dient der Ermittlung von Modellkonfigurationen anhand gegebener Gelenkpositionen. Zu einer bekannten Pose der vollständigen Starrkörperkette ist also die Parametrisierung des Modells zu ermitteln. Diese umfasst sowohl die globale Position als auch die einzelnen Gelenkwinkel.

Der Zusammenhang zwischen Zustandsvektor und Gelenkkoordinaten ist in Abbildung 2.2 zu sehen. Sämtliche Freiheitsgrade eines Menschmodells in Form von globaler Position und Orientierung, sowie allen einstellbaren Gelenkwinkel sind Bestandteil des Zustandsvektors. Hieraus ergeben sich durch Traversierung der Starrkörperkette die jeweiligen Gelenkkoordinaten. Im Allgemeinen werden zur Berechnung der inversen Kinematik die letzten Glieder der kinematischen Ketten, die so genannten *Endeffektoren*, korrekt ausgerichtet und im Folgenden die verbleibenden Glieder, entsprechend ihrer Freiheitsgrade ausgerichtet. Diese Aufgabe kann sowohl mit geometrischen, analytischen als auch numerischen Verfahren gelöst werden. Da aufgrund von Messungenauigkeiten häufig keine exakten Gelenkpositionen vorliegen, erfolgt die Lösung meist durch Formulierung eines Minimierungsproblems.

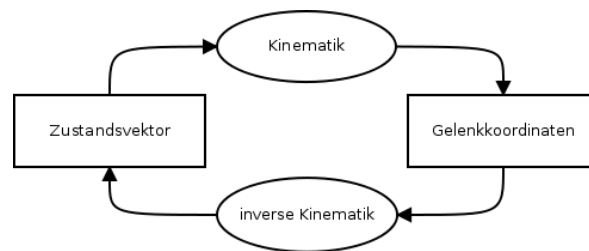


Abbildung 2.2: Zusammenhang: Zustandsvektor – Gelenkkoordinaten

## 2.2 Volumetrische Rekonstruktion aus Kameradaten

Unabhängig davon ob es sich um ein kommerzielles System oder einen Prototyp handelt, 2D Kameradaten sind bei nahezu allen optischen Tracking-Systemen die Grundlage sämtlicher Beobachtungen. Eine Projektion der aufgenommenen Szene geht jedoch mit einem erheblichen Informationsverlust einher. Neben dem Mangel an Tiefeninformationen stellen partielle Verdeckungen innerhalb der Szene ein Problem dar. Um diesem Effekt entgegenzuwirken, werden häufig Mehrkamarasysteme verwendet, welche die Szene aus möglichst unterschiedlichen Blickwinkeln beobachten.

Die so gewonnenen Ansichten werden innerhalb eines Beobachtungsmodells genutzt, um die Übereinstimmung einer (Mensch-)Modellkonfiguration mit diesen Bildinformationen zu messen. Dies geschieht häufig durch eine Projektion des dreidimensionalen Modells in die einzelnen Bildebenen. Hier werden dann Bild- und Modellmerkmale miteinander verglichen, indem z. B. die Überlappung der Silhouette des Modells mit

einer Vordergrundmaske gemessen wird. Ein geometrisch kalibriertes Mehrkamerasystem liefert die für die Projektion benötigten Informationen. Neben der Möglichkeit, ein Modell in die Kameraebenen zu projizieren, kann die Bildinformation auch genutzt werden, um eine 3D-Rekonstruktion der Szene zu erstellen. Hierzu wird gewöhnlich von fest installierten Kameras und einem vornehmlich statischen Hintergrund sowie kontrollierten Aufnahmebedingungen ausgegangen, etwa in Form konstanter Beleuchtung. Diese Randbedingungen entsprechen weitestgehend den Einschränkungen, welche auch für die Verwendung kommerzieller markerbasierter Systeme gelten.

### 2.2.1 Vordergrund- & Hintergrundmodelle

Mittels Hintergrundsubtraktion ist es möglich, den Bildvordergrund vom Bildhintergrund zu trennen. Im einfachsten Fall wird davon ausgegangen, dass sich nur das zu verfolgende Objekt im Vordergrund befindet, in komplexeren Szenen wird eine weitere Segmentierung des Vordergrundes vorgenommen. Die Trennung erfolgt anhand mindestens eines – meist statistischen – Hintergrundmodells, optional erfolgt eine explizite Modellierung des Vordergrundes. Übliche Verfahren zur statistischen Modellierung sind etwa Gaußmischverteilung (*mixture of gaussian*) [SG99] oder Kerndichteschätzung (*kernel density estimation*)[EDD03]. Hierbei werden die zu erwartenden Hintergrundmodalitäten pixelweise erfasst und durch mehrere Gaußverteilungen modelliert. Das so gewonnene Hintergrundmodell wird über die Zeit hinweg aktualisiert und Abweichungen werden mittels Schwellwertverfahren als Vordergrund klassifiziert. Die resultierenden Silhouetten können anschließend für die Pose-Rekonstruktion verwendet werden.

### 2.2.2 3D-Rekonstruktion

Die Vordergrundmasken erlauben Rückschlüsse darüber, entlang welches Sichtstrahles sich Vordergrundobjekte innerhalb der Szene befinden. Allerdings lässt sich aus einer einzelnen Ansicht die notwendige Tiefeninformation nicht rekonstruieren. Stellt man das Sichtvolumen, welches im Erfassungsbereich aller Kameras befindet, als äquidistant aufgeteilten Voxel-Raum vor, so lassen sich die belegten Voxel mittels Rückprojektion der Silhouetten im Raum ermitteln. Hierzu wird jedes Voxel, welches sich auf dem Sichtstrahl durch Kamerazentrum und Silhouetten-Pixel befindet, als potentiell belegt markiert. Abschließend wird für jedes Voxel festgestellt, durch wie viele Kameras es als belegt markiert wurde, um mittels Schwellwert zwischen belegten und unbelegten Voxeln zu unterscheiden.

Hieraus ergibt sich ein Voxelraum, welcher durch Farbinformationen und Belegungswahrscheinlichkeiten weiter angereichert und mittels morphologischer Operatoren geglättet werden kann.

## 2.3 Bayes Filter

Ist ein dynamischer Prozess lediglich durch Beobachtungen in Form von Stichproben bekannt, so stellen stochastische Verfahren zur Zustandsschätzung eine Möglichkeit dar, relevante Parameter der zugrundeliegenden Verteilungsfunktion zu ermitteln. Stochastische Zustandsschätzer unterscheiden sich hinsichtlich ihrer Annahmen bezüglich Linearität und Modalität des betrachteten Prozesses.

Eine allgemeine Form stellen die sequentiellen **Monte-Carlo-Methoden** (auch Partikelfilter, SIR-Filter oder Condensation Algorithmus) dar. Sie sind in der Lage nichtlineare, multimodale Verteilungen zu erfassen. Ziel dieser Methoden ist es, die *a posteriori* Wahrscheinlichkeitsdichte basierend auf allen verfügbaren Informationen und durchgeführten Messungen zu ermitteln. In vielen Anwendungsfällen ist es notwendig, zu jedem Zeitschritt Messungen durchzuführen. Ein **rekursiver Aufbau** des Filters ist in solchen Fällen die übliche Vorgehensweise. Hierbei erfolgt die Systemmodellierung so, dass nur der jeweils letzte Zeitschritt zur Vorhersage notwendig ist (Markov-Kette).

### 2.3.1 Bayes-Theorem

Das Bayes-Theorem beschreibt den Zusammenhang bedingter Wahrscheinlichkeiten zweier Ereignismengen  $A$  und  $B$ . Aus den Grundwahrscheinlichkeiten  $P(A)$  und  $P(B)$  sowie der bedingten Wahrscheinlichkeit  $P(B|A)$  erlaubt es die Berechnung von  $P(A|B)$ .

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \quad (2.1)$$

### 2.3.2 Problemformulierung

Die Sequenz der Zustände  $\vec{x}_t$  mit  $\{\vec{x}_t, t \in \mathbb{N}\}$  ist gegeben durch:

$$\vec{x}_t = f_t(\vec{x}_{t-1}, \vec{v}_{t-1}) \quad (2.2)$$

wobei  $f_t : \mathbb{R}^{n_x} \times \mathbb{R}^{n_v} \rightarrow \mathbb{R}^{n_x}$  ein potentiell nicht lineares **Zustandsübergangsmodell** ist. Jeder Folgezustand ergibt sich also mittels des Zustandübergangsmodells aus dem vorherigen Zustand und einem zusätzlichen Rauschen. Die Größen  $n_x, n_v$  geben die Dimension des Zustandsvektors respektive des Prozessrauschens  $\vec{v}_{t-1}$  an.

Ziel eines jeden Tracking-Verfahrens ist es nun  $\vec{x}_t$  (rekursiv) aus einer Reihe von Messungen  $\vec{z}_t$  zu ermitteln.

$$\vec{z}_t = h_t(\vec{x}_t, \vec{n}_t) \quad (2.3)$$



Das potentiell nicht lineare Messmodell  $h_t : R^{n_x} \times R^{n_n} \rightarrow R^{n_z}$  überführt den Zustandsvektor  $\vec{x}_t$  zusammen mit dem Messrauschen  $\vec{n}_t$  in die Messung  $\vec{z}_t$ . Die Dimensionen von Messrauschen und Messung sind hierbei durch  $n_n$  und  $n_z$  gegeben.

Die Aufgabe eines Bayes-Filters ist es, die Wahrscheinlichkeit einer Parametrisierung des Vektors  $\vec{x}_t$  zum Zeitpunkt  $t$  unter Berücksichtigung der Messungen  $\{\vec{z}_1, \dots, \vec{z}_t\}$  zu bestimmen. Die initiale Wahrscheinlichkeitsdichte  $p(\vec{x}_0|\vec{z}_0) = p(\vec{x}_0)$  wird als bekannt vorausgesetzt. Unter diesen Voraussetzungen kann die Wahrscheinlichkeitsdichtefunktion  $p(\vec{x}_t|\vec{z}_1 : t)$  in zwei Schritten ermittelt werden:

### 1. Vorhersage:

Unter Verwendung des Systemmodells (Gleichung 2.2) wird mittels der Chapman–Kolmogorov Gleichung (2.4) die Wahrscheinlichkeitsdichtefunktion zum Zeitpunkt  $t$  ermittelt. Da 2.2 eine Markov-Kette (siehe B.1.2) erster Ordnung beschreibt, gilt hierbei  $p(\vec{x}_t|\vec{x}_{t-1}, \vec{z}_{1:t-1}) = p(\vec{x}_t|\vec{x}_{t-1})$ . Die *a posteriori* Wahrscheinlichkeitsdichte des vorhergehenden Zeitschrittes  $p(\vec{x}_{t-1}|\vec{z}_{1:t-1})$  wird als bekannt vorausgesetzt.

Die bivariate Wahrscheinlichkeitsverteilung von  $\vec{x}_t$  und  $\vec{x}_{t-1}$  ergibt sich bei bekannten  $\vec{z}_{1:t-1}$  aus:

$$\begin{aligned} p(\vec{x}_t, \vec{x}_{t-1}|\vec{z}_{1:t-1}) &= p(\vec{x}_t|\vec{x}_{t-1}, \vec{z}_{1:t-1})p(\vec{x}_{t-1}|\vec{z}_{1:t-1}) \\ &= p(\vec{x}_t|\vec{x}_{t-1})p(\vec{x}_{t-1}|\vec{z}_{1:t-1}) \end{aligned}$$

Integration über  $\vec{x}_{t-1}$  ergibt die Chapman–Kolmogorov Gleichung und damit den Vorhersageschritt:

$$p(\vec{x}_t|\vec{z}_{1:t-1}) = \int p(\vec{x}_t|\vec{x}_{t-1})p(\vec{x}_{t-1}|\vec{z}_{1:t-1})d\vec{x}_{t-1} \quad (2.4)$$

### 2. Aktualisierung:

Aus der nun bekannten *a priori* Verteilung  $p(\vec{x}_t|\vec{z}_{1:t-1})$  sowie dem Beobachtungsmodell  $p(\vec{z}_t|\vec{x}_t)$  kann die *a posteriori* Verteilung mittels der Bayes'schen Regel berechnet werden.

$$p(\vec{x}_t|\vec{z}_t, \dots, \vec{z}_1) = \frac{p(\vec{z}_t|\vec{x}_t)p(\vec{x}_t|\vec{z}_{1:t-1})}{p(\vec{z}_t|\vec{z}_{1:t-1})} \quad (2.5)$$

Das zuvor beschriebene Verfahren stellt eine allgemeine Form des Bayes-Filters dar, jedoch ist eine rekursive Ermittlung der vollständigen Wahrscheinlichkeitsdichte in der Praxis meist nicht möglich. Unter gewissen Randbedingungen ist jedoch eine analytische Lösung etwa mit Hilfe des Kalman-Filters möglich. Gelten die hierzu erforderlichen Annahmen nicht, so stehen approximative Verfahren in Form von erweiterten Kalman-Filtern oder Partikelfiltern zu Verfügung.

### 2.3.3 Kalman-Filter

Kann von einem linearen Prozessmodell ausgegangen werden, welches lediglich eine Modalität besitzt und daher unter Berücksichtigung von Rauschen durch Mittelwert und Varianz modelliert werden kann, so ist ein Kalman-Filter [Kal60; WB06] das Mittel der Wahl. Der Kalman-Filter ist ein *Minimum-Varianz-Schätzer* und bietet den Vorteil, dass die aufwändige Auswertung mehrerer Hypothesen entfällt. Sofern  $p(\vec{x}_{t-1}|\vec{z}_{1:t-1})$  gaußverteilt ist, kann unter folgenden Annahmen gezeigt werden, dass Gleiches für  $p(\vec{x}_t|\vec{z}_t, \dots, \vec{z}_1)$  gilt:

1.  $\vec{v}_{t-1}$  und  $\vec{n}_t$  entstammen beide gaußverteilten Zufallsgrößen.
2.  $f_t(\vec{x}_{t-1}, \vec{v}_{t-1})$  und  $h_t(\vec{x}_t, \vec{n}_t)$  sind bekannte lineare Funktionen

Unter diesen Voraussetzungen können die Gleichungen 2.2 und 2.3 wie folgt umgeschrieben werden:

$$\vec{x}_t = F_t \vec{x}_{t-1} + \vec{z}_{t-1} \quad (2.6)$$

$$\vec{z}_t = H_t \vec{x}_t + \vec{n}_t \quad (2.7)$$

$F_t$  und  $H_t$  sind bekannte Matrizen (System- und Messwertmatrix), welche einen linearen funktionalen Zusammenhang definieren.

$$p(\vec{x}_{t-1}|\vec{z}_{1:t-1}) = N(\vec{x}_{t-1}; m_{t-1|t-1}, P_{t-1|t-1}) \quad (2.8)$$

$$p(\vec{x}_t|\vec{z}_{1:t-1}) = N(\vec{x}_t; m_{t|t-1}, P_{t|t-1}) \quad (2.9)$$

$$p(\vec{x}_t|\vec{z}_{1:t}) = N(\vec{x}_t; m_{t|t}, P_{t|t}) \quad (2.10)$$

In [DNBB99] wird im Rahmen von Experimenten nachgewiesen, dass das Personen-tracking häufig mit multimodalen, nicht gaußverteilten Wahrscheinlichkeitsdichten einhergeht.

Falls die zuvor genannten Randbedingungen nicht zutreffen, werden approximative Verfahren eingesetzt. Hierzu zählen sowohl verschiedene Kalman-Filter Derivate wie etwa *Extended Kalman Filter* oder *Unscented Kalman Filter*.

Während ersteres das Modell durch eine lokale Linearisierung vereinfacht, wird im zweiten Fall die *Unscented Transform* benutzt, um sowohl Mittelwert als auch Kovarianz unter nichtlinearer Transformation vorherzusagen [JU04].

### 2.3.4 Partikelfilter

Partikelfilter sind Monte-Carlo-Methoden, welche auf einer Punktmengen-Repräsentation der Wahrscheinlichkeitsdichte basieren. Sie können auf jeden Zustandsraum (*state space*) angewendet werden und stellen diesbezüglich eine

Verallgemeinerung des Kalman Filters dar. Die Dynamik des Systems (Zustandes) wird für diskrete Zeitschritte ermittelt, so dass sich das Problem als Ermittlung der Zustandsvektoren für eine zeitlich zusammenhängende Folge beschreiben lässt. Da die Wahrscheinlichkeitsdichte-Funktion sowohl im kontinuierlichen, als auch im diskreten Fall mit steigender Dimensionalität nicht – oder nur sehr aufwändig – ausgewertet werden kann, verfolgen Partikelfilter den Ansatz, die Dichtefunktion durch die Wahl einiger geeigneter Stichproben (Partikel) zu repräsentieren. Partikelfilter (auch SIR Filter = Sampling Importance Resampling Filter) wurden erstmals von Andrew Blake und Michael Isard 1998 in Form des *Condensation* Algorithmus im Kontext von visuellem Personentracking eingesetzt [IB98]. Die Dichtefunktion wird durch eine Menge gewichteter Partikel wie folgt repräsentiert:

$$p(\vec{x}) \approx \sum_{i=1}^N w_i \delta(\vec{x} - \vec{x}_i) \quad (2.11)$$

$$w_i \propto \frac{\pi(\vec{x}_i)}{q(\vec{x}_i)} \quad (2.12)$$

$w_i$  entspricht hierbei dem normalisierten Gewicht des  $i$ -ten Partikels.

Die *a posteriori* Wahrscheinlichkeitsdichte  $p$  zum Zeitpunkt  $t$  ergibt sich hierbei durch Berücksichtigung der aktuellen Konfigurationen  $\vec{x}$  und aller Messungen  $\vec{z}$  bis hin zum Zeitpunkt  $t$ . Jeder Partikel besteht aus einer Parametrisierung (dem Zustandsvektor  $s_t$ ) und der dazugehörigen Gewichtung  $\pi_t$ .

Die Gewichte sind so normiert, dass weiterhin gilt:

$$\sum_{t=0}^N \pi_t^{(n)} = 1 \quad (2.13)$$

Die folgende Darstellung(1) zeigt eine Iteration des Partikelfilters:

---

**Algorithmus 1** Zwischenschritte eines SIR-Filters für einen diskreten Zeitschritt

---

Eine Iteration des Partikelfilters für einen Zeitpunkt  $t$  besteht aus den folgenden drei Schritten.

1. **Resampling:** Eine bereits gewichtete Partikelmenge des vorhergehenden Zeitschrittes geht durch „Ziehen mit Zurücklegen“ in die Partikelmenge des aktuellen Zeitschrittes über. Die Wahrscheinlichkeit, einen Partikel zu ziehen, entspricht hierbei seiner Gewichtung.
  2. **Diffusion:** Die Partikel werden (stochastisch) – entsprechend des gewählten Rausch- oder Bewegungsmodells – bewegt.
  3. **Weighting:** Die Partikel werden durch das verwendete Messmodell gewichtet und repräsentieren somit die *a posteriori* Wahrscheinlichkeitsdichte.
-

Der tatsächliche Zustand (zum Zeitpunkt  $t$ ) kann im Folgenden sowohl als gewichtetes Mittel als auch als Maximum dieser Verteilung angenommen werden.

Eine nicht endliche Partikelzahl vorausgesetzt, ergäbe sich eine exakte Abbildung der Wahrscheinlichkeitsverteilung. Da in der Praxis jedoch verschiedenste Faktoren die tatsächliche Partikelzahl begrenzen, stellt diese den limitierenden Faktor bezüglich der Genauigkeit der Annäherung dar.

## Dimensionalität und Partikelzahl

Die Anzahl benötigter Partikel, um eine hinreichend genaue Abdeckung des Suchraumes zu erhalten, wächst proportional mit der Dimensionalität des Raumes, was insbesondere im hier betrachteten Kontext schnell zu Einschränkungen führen kann. MacCormick und Black haben bereits 2000 in [MI00] gezeigt, dass Dimensionalität und Partikelanzahl wie folgt korrelieren:

$$N \geq \frac{D_{min}}{\alpha^d} \quad (2.14)$$

$N$  entspricht der Anzahl der *benötigten Partikel*,  $d$  ist die Anzahl der Dimensionen. Sowohl  $\alpha$  als auch  $D_{min}$  sind Konstanten, welche passend zu ermitteln sind, wobei  $\alpha \ll 1$  gilt.

Nach [DR03] lassen sich mittels des *Condensation*-Algorithmus im Kontext des Starrkörper Trackings hinreichend gute Ergebnisse für niedrig dimensionale Räume bis hin zu einer Grenze von etwa 10 Dimensionen erreichen.

Die in niedrigeren Räumen gebräuchliche Methode, das Optimum durch äquidistante Abtastung zu ermitteln, stößt im Bereich der Pose-Rekonstruktion schnell an ihre Grenzen. Um die Pose eines 30 dimensionalen Modells durch Abtastung mit einer Genauigkeit von  $5^\circ$  zu ermitteln ( $180^\circ$  durchschnittliche Bewegungsfreiheit vorausgesetzt), wäre eine Menge von  $4 * 10^{37}$  Auswertungen notwendig.

Ein auf Partikelfiltern und Suchraum-Unterteilung beruhender Ansatz, welcher durch kontinuierliche Optimierung der Partikel und Zerlegung des Zustandsraumes die zuvor angesprochenen Probleme berücksichtigt, wird als Teil dieser Arbeit in Kapitel 4 vorgestellt.

## 2.4 Komponenten eines Pose-Schätzers

Betrachtet wird im Folgenden ein Verfahren zur markerlosen Rekonstruktion der Pose eines Menschmodells ohne die Zurhilfenahme von künstlichen Markern.

Unter Pose-Rekonstruktion versteht man die Ermittlung jener Parameter, welche die Position und Orientierung eines Körpers im Raum eindeutig beschreiben. Für einen festen Körper im Raum handelt es sich hierbei um je drei Translations- sowie Rotationsparameter. Im Kontext der Starrkörperphysik 2.1 sind jedoch nicht zwangsläufig für jedes Kettenglied alle Parameter frei wählbar. Da in der Regel alle Körper entlang einer Starrkörperkette – mit Ausnahme des Wurzelements – eine (körper-)feste Lage besitzen und sich damit die Freiheitsgrade aus maximal drei Rotationsparametern ergeben. Neben der Bestimmung der Lage zu einem festen Zeitpunkt  $t$  besteht eine weitere Herausforderung in der Verfolgung des Körpers während zukünftiger Beobachtungszeitpunkte. Hierbei wird in den meisten Fällen versucht, die freien Parameter  $p$  zum Zeitpunkt  $t + 1$  aus den vorhergehenden Informationen zu ermitteln.

$$p_{t+1} = f(p_t, p_{t-1}, p_{t-2}, \dots, p_0) \quad (2.15)$$

Häufig ist es möglich, den Prozess als Markov-Kette darzustellen, so dass sich 2.16 wie folgt vereinfachen lässt:

$$p_{t+1} = f(p_t) \quad (2.16)$$

Die zur Rekonstruktion benötigten *Tracking-Algorithmen* lassen sich allgemein durch eine Reihe von Grundbausteinen beschreiben. Das prinzipielle Vorgehen besteht darin, in zeitdiskreten Abständen eine Szene zu beobachten und gewählte Merkmale des Beobachtungsraums mit einer Modellannahme zu vergleichen. Unter dieser Betrachtung zerfällt das Vorgehen in die folgenden Bestandteile.

1. **Beobachtungsraum:** Der Beobachtungsraum bestimmt, welche Merkmale (z. B. aus dem Kamerabild) verwendet werden und wie diese gegebenenfalls extrahiert werden.
2. **Körpermodell:** Das Körpermodell enthält das *a priori* Wissen über den zu lokalisierenden Körper. Hierzu können etwa kinetische oder kinematische Angaben und personenspezifische Merkmale, wie die Längen einzelner Körperglieder, gehören.
3. **Messmodell:** Das Messmodell (auch *Beobachtungsmodell*) beschreibt, wie die beobachteten Merkmale und das Körpermodell mit einander verglichen werden können. Dies ist notwendig, da die zu vergleichenden Einheiten in unterschiedlichen Repräsentationen mit abweichender Semantik vorliegen können.
4. **Vorhersagemodell:** (auch *Systemmodell*, *Bewegungsmodell*) Im Fall der (hier vornehmlich betrachteten) probabilistischen Verfahren ist es erforderlich, basierend auf bisherigen Beobachtungen eine Vorhersage über eine oder mehrere zukünftige Modellkonfigurationen machen zu können.

Dem Beobachtungsraum wird ein Messvektor entnommen, welcher eine verrauschte Beobachtung des tatsächlichen Zustandes ist. Im generellen Fall kann davon ausgegan-

gen werden, dass der Messvektor eine deutlich niedrige Dimensionalität besitzt als das tatsächlich beobachtete Modell. Wird etwa für ein Modell mit 30 Freiheitsgraden die Überlappung im Bildraum gemessen und ein skalares Gewicht zurückgeliefert, wird die Diskrepanz besonders deutlich. Die Prozesse, welche das Rauschen der Messung bedingen, sind im Allgemeinen nicht vollständig bekannt, bzw. nur schwerlich zu modellieren. Verzerrungen durch das Kamerasystem, falsch positionierte Messpunkte sowie zeitliche Ungenauigkeiten sind nur einige der auftretenden Schwierigkeiten.



# Kapitel 3

## Stand der Technik

Auf den folgenden Seiten wird ein Überblick bezüglich der aktuell verfügbaren Verfahren zur Pose-Rekonstruktion präsentiert. Wesentliche Bestandteile und Methoden werden identifiziert und hinsichtlich ihrer Möglichkeiten und Grundannahmen klassifiziert.

Eine Kategorisierung bestehender Tracking-Verfahren gestaltet sich prinzipiell schwierig, da eine Vielzahl möglicher Schlüssel denkbar ist. Umfangreiche Expertisen zu markerlosen und teils auch markerbasierten Trackingverfahren existieren etwa in Form von [Gav99; Gra01; YJS06; ZH04]. Während Yilmaz eine mehrdimensionale Klassifizierung vornimmt und sowohl nach technischen als auch anwendungsorientierten Aspekten unterscheidet, stellt Gavrilin Dimensionalität und Form des Menschmodells in den Vordergrund. Verschiedenste Szenarien von Sportlertracking, Überwachungs- und Sicherheitssystemen bis hin zur Robotik bieten sich als Einsatz- und Forschungsgebiete an. Eine Grobübersicht und Zusammenfassung üblicher Kategorien zeigt Tabelle 3.1. Die Anforderungen an das einzusetzende System unterscheiden sich teils deutlich und können einer ersten Kategorisierung dienen.

Kategorie	Beschreibung
Bewegungserkennung	Identifikation von Bewegungsmustern
Überwachung	Erkennung von Abweichungen und außergewöhnlichen Ereignissen
Mensch-Maschine Interaktion	Eye-Tracking, Gestenerkennung

Tabelle 3.1: Verschiedene Einsatzgebiete für Trackinganwendungen.

Neben diesem **pragmatischen, zweckorientierten Ansatz** erscheint eine Betrachtung der eingesetzten **Technologien** sinnvoll. Im ersten Schritt ergibt sich hieraus eine Unterteilung in markerbasierte und nicht markerbasierte Systeme. Die erstgenannte Gruppe unterteilt sich im Folgenden weiter anhand der eingesetzten



Markertechnologien. Streng genommen verzichten markerfreie Systeme lediglich auf synthetische Messpunkte und beschränken sich auf die Beobachtung natürlicher Eigenschaften (im Bildraum). Hier bietet sich also eine Unterscheidung anhand der gewählten Merkmale an.

Auf **konzeptioneller Ebene** ergibt sich eine klare Trennung sowohl auf Basis der eingesetzten Vorhersageverfahren als auch bezüglich der Körpermodelle. Die Pose-Rekonstruktion kann unter der Prämisse erfolgen, dass eine deterministische Vorhersage möglich ist, sich also die Parameter bis auf unvermeidbare Rauschterme mit hinreichender Genauigkeit durch bisherige Beobachtung und ein Zustandsübergangsmodell ermitteln lassen. Dem gegenüber steht ein probabilistischer Ansatz, welcher versucht, aus bisherigen Ergebnissen und der aktuellen Beobachtung die Plausibilität einer Parametrisierung zu beurteilen. Bezüglich der verwendeten Körper- bzw. im konkreten Fall Menschmodelle bietet sich ein sehr breites Spektrum. Die eingesetzten Verfahren reichen von punktbasierten Verfahren hinweg über einfache Linienmodelle bis hin zu komplexen dreidimensionalen geometrischen Oberflächennetzen.

## 3.1 Menschmodelle

Die Detektion eines Menschen innerhalb einer Szene geschieht durch das Auffinden von Merkmalen, die spezifische Eigenschaften des Menschen beschreiben. Neben den Sensordaten, welche die Szenebeschreibung liefern, ist daher eine Beschreibung des Menschen notwendig. Eine Einteilung anhand der verschiedenen Detailgrade führt zur folgender Unterscheidung:

### 1. Modellfreie Systeme:

Die Detektion erfolgt auf Basis von Einzelmerkmalen, deren Anordnung Rückschlüsse über Anwesenheit und gegebenenfalls Position einer Person zulassen. Beispielsweise kann die Detektion von Körperregionen anhand von Form- und Farbmerkmalen geschehen. So detektiert Lee in [LC03] Hände und Kopf anhand einer kombinierten Kontur- und Hautfarbenbeschreibung. Alternativ werden auch Punktmerkmale oder Texturinformationen eingesetzt.

### 2. Indirekte Modelle:

In diese Kategorie fallen Modelle, welche keine direkte Parametrisierung zulassen. Dies können z. B. Konturbeschreibungen sein, welche Bewegungszwischenschritte oder Handlungen wie Greifen, Laufen etc. beschreiben. Auch Merkmalsdatenbanken lassen sich dieser Klasse zuschreiben.

### 3. Direkte Modelle:

Modelle dieser Kategorie erlauben eine direkte Parametrisierung der Modelle. Insbesondere für die Verwendung in probabilistischen Filtern ist diese Art der Modellierung erforderlich. Moeslund [Gra01] schätzte den Anteil der modellbasierten Verfahren im Jahr 2000 auf rund 40%, ausgehend von einer Begutachtung von rund 150 Veröffentlichungen.

Der (direkte) modellbasierte Ansatz folgt dem Schema „*predict - match - update*“ (Vorhersage - Übereinstimmung - Aktualisierung). Basierend auf bisherigen Informationen erfolgt eine Vorhersage, deren Übereinstimmung mit den aktuellen Beobachtungen anschließend bewertet wird. Aufgrund dieser Bewertung erfolgt letztendlich eine Aktualisierung des Vorhersagemodells. Die Bewertung geschieht mittels Analyse durch Synthese (*analysis-by-synthesis*). Das Vorhersagemodell wird hierbei so synthetisiert (abstrahiert, interpretiert) dass sich eine mit dem Beobachtungsraum vergleichbare Basis ergibt.

Zur Synthetisierung des Modells gibt es verschiedene Methoden. Die meisten Ansätze basieren auf einer Überführung von Modelleigenschaften in den Beobachtungsraum. Dies kann etwa in Form einer Projektion von Modellmerkmalen in den Bildraum geschehen. Beispielsweise projiziert Deutscher [DBR09] sowohl Silhouette als auch Konturpunkte in ein Vordergrundmodell bzw. Kantenbild der Aufnahme. Dieses kombinierte Vorgehen soll die Schwächen der einzelnen Ansätze kompensieren. Eine reine Betrachtung der Kanten respektive deren Abstände zueinander erweist sich als anfällig gegenüber Bildrauschen und sich ändernden Bildinhalten. Die Verwendung einer mittels Vordergrundextraktion gewonnenen Silhouette ist diesbezüglich robuster, kann aber Überlappungen und Verdeckungen nicht auflösen.

Zusätzlich zur Integration mehrerer Bildeigenschaften schlägt Lee [LC03] vor, die Detektion durch sogenannte *auxiliary measurements* zu stützen. Hierzu wird, basierend auf Kontur- und Farbmerkmalen, die Position von Kopf und Händen grob lokalisiert, um etwaige Körperkonfigurationen zu verifizieren.

Kategorie	Unterkategorie	Beschreibung
2D-Modelle	Blob-Tracker	Tracking unartikulierter Modelle, keine Starrkörperkette
	Skelett- und Stickmodelle	einfache Starrkörperkette, Vergleich mittels Thinning-Algorithmen
	Silhouetten und Konturpunkte	Vergleich mit Kantenbild oder Vordergrundmaske
3D-Modelle	geometrische Hüllkörper	Volumenmodellierung mittels Ellipsoiden, Quadrics, Zylindern etc.
	Oberflächennetze	topologische Netze, Triangulierung, visuelle Hülle

Tabelle 3.2: Übersicht Tracking-Verfahren, nach aufsteigender Komplexität sortiert.

### 3.1.1 Dynamische Aspekte

Neben den statischen Eigenschaften eines Körpers kann auch Wissen über seine Dynamik explizit modelliert werden. So wird in [BB09] neben absoluten Gelenkanschlüssen auch die zeitliche Veränderung limitiert, so dass die Winkelgeschwindigkeiten experimentell ermittelte Grenzen nicht überschreiten. Hierzu wird basierend auf der Frame-rate die maximale Winkelveränderung in Bezug auf das Optimum des vorhergehenden Zeitschritts bestimmt. Ein trinanguliertes Netz wird zusätzlich um die Starrkörper gelegt, welches in Gelenkbereichen adaptiv gestaltet ist und sich so an die jeweilige Körperstellung anpasst. Dieser Ansatz vermeidet harte Kanten und führt zu einer besseren Silhouette im Bereich der Gelenkpunkte.

### 3.1.2 Modelladaption

Über die Laufzeitparameter hinaus existieren fixe Größen, wie etwa die Segmentlängen und Volumina der einzelnen Starrkörper, deren fallspezifische Bestimmung entscheidenden Einfluss auf die Effizienz des Verfahrens haben kann. Eine Parametrisierung der Starrkörper kann sowohl manuell als auch automatisiert erfolgen. So können die Segmentlängen als weitere Freiheitsgrade betrachtet und bei bekannter Initialpose automatisch durch das Partikelfilter-Framework optimiert werden. Erfolgt die Anpassung durch den Anwender, ist sowohl ein Zugriff auf den vollen Parametersatz denkbar, als auch auf einen reduzierten Satz an Stellparametern. Diese können mittels Hauptkomponentenanalyse aus einer Menge von 3D-Modellen extrahiert werden [BB09; HSS<sup>+</sup>09] und spezifische Merkmale wie Geschlecht, Größe, Alter oder Gewicht widerspiegeln. Diese Anpassung ist zwingend erforderlich, da ein schlecht eingepasstes Modell die Rekonstruktion nachhaltig beeinträchtigt. So schwankt etwa das Verhältnis von Schulterbreite zu Hüftbreite (siehe Abbildung 3.1) zwischen Männern und Frauen um etwa 15% [Das03], was bei einem allgemeinen Modell zwangsläufig zu Fehlern in der Position der Hüft- oder Oberarmgelenke führen würde.

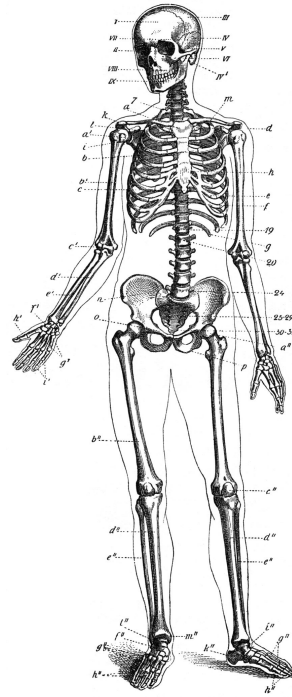


Abbildung 3.1: Skelett-Modell des menschlichen Körpers. Quelle: Wikimedia, Hubert Ludwig, Schul - Naturgeschichte 1891

## 3.2 Prädiktion und Modelleinpassung

Die Vorhersage und Bewertung von Modellkonfigurationen basiert auf Annahmen, welche die Beschaffenheit des Suchraumes und seine zeitliche Entwicklung betreffen. Sminichisescu [ST03] unterteilt die Suchstrategien für hochdimensionale Räume in drei Kategorien:

1. **lokale Optimierung, lokaler Abstieg:** Eine initiale Pose wird kontinuierlich durch deterministische Optimierungsmethoden inkrementell verbessert. Hierzu zählen z.B. Newton-Iteration, Levenberg-Marquardt oder Iterative Closest Point (ICP). Der letztgenannte Algorithmus wird etwa in *VooDoo*<sup>1</sup>, einem OpenSource-System zur Starrkörper-Verfolgung verwendet, welches auf diesem Wege Starrkörper in 3D Laserdaten einpasst[KVD07].
2. **gleichmäßige, planmäßige Abtastung:** Es erfolgt eine Auswertung der Kostenfunktion an vordefinierten (meist äquidistanten) Stellen des Wertebereichs, wie etwa in Form eines orthogonalen Gitters [GD96].
3. **stochastische Abtastung:** Die Abtastung des Suchraumes geschieht durch stochastische Partikelbewegung. Als eine der bekanntesten Veröffentlichungen in diesem Bereich dürfte die Vorstellung des *Condensation* Algorithmus durch Isaard and Blake[IB98] gelten.

---

<sup>1</sup><http://www.iain.ira.uka.de/users/knoop/VooDoo/doc/html/>

Während die meisten Verfahren zum Tracking artikulierter hochdimensionaler Modelle Partikelfilter einsetzen, existiert mit [SMC01] ein abweichender Ansatz, welcher ein 27 dimensionales Handmodell mittels des Unscented Kalman Filters über 100 Frames erfolgreich verfolgt. Knoop [KVD07] verwendet ein zylindrisches Menschmodell, welches mittels ICP in die beobachtete Datenmenge eingepasst wird. Ein weiterer interessanter Aspekt besteht in der Möglichkeit, neben herkömmlichen 2D Daten auch direkt Laserscans verarbeiten zu können.

### 3.2.1 Optimierungen

Wird eine Suchraumabtastrung vorgenommen, so folgt aus dieser die Auswertung einer Vielzahl von Hypothesen. Im Sinne einer effizienten Umsetzung ist es diesbezüglich wünschenswert, die benötigte Anzahl an Auswertungen möglichst gering zu halten.

#### Lokale Optimierung und Partitionierung

Die Formulierung des Anfangsproblems als eine Summe von Teilproblemen (Partitionierung) stellt ebenso wie eine kontinuierliche Verbesserung von bereits durchgeführten Bewertungen eine mögliche Strategie dar. Gavrilu [Gav99] sowie MacCormick [MI00] verwenden eine Partitionierung des Suchraums, um die Anzahl der benötigten Partikel zu reduzieren. Bandouch und Beetz stellen in [BEB08] einen kombinierten Ansatz vor, welcher Suchraumpartitionierung und lokale Optimierung in Form von Simulierter Abkühlung nutzt. Beide Verfahren werden dort getrennt betrachtet und ihre Vorteile sowie Risiken herausgearbeitet:

- Wurden Suchraumpartitionierung und lokale Optimierung mit der gleichen Zahl an Partikelauswertungen durchgeführt, so führte die Partitionierung zu einem geringen maximalen Fehler. Allerdings stellte sich auch heraus, dass die mittels Partitionierung erzielten Ergebnisse erheblich von der gewählten Aufteilung des Suchraums abhängen. Weiterhin ist kritisch anzumerken, dass eine Partitionierung die Gefahr multiplikativer Fehler birgt. Wird etwa in einer frühzeitig bearbeiteten Partition der Torso falsch positioniert, so besteht für die untergeordneten Körperteile keine Chance zur korrekten Positionierung.
- Eine deutliche Reduktion der Partikelzahl im Vergleich zu konventionellen Partikelfiltern wurde auch mittels Simulierter Abkühlung erreicht. Die hierbei erfolgte simultane Optimierung aller 51 Freiheitsgrade erforderte mindestens 20 Iterationen, um befriedigende Ergebnisse zu erhalten.

Eine Partitionierung des Suchraums führt häufig, wie bereits erörtert, zu einem multiplikativen Fehler in der Parameterermittlung. Diesem Phänomen wird in [BB09] systematisch entgegengewirkt, indem verschiedene Partitionierungsschemata gegeneinander antreten. Es werden nebenläufig mehrere Schemata ausgeführt und die Optima dieser Partitionierungen anschließend fusioniert. Der Rechenaufwand steigt hierbei linear mit der Anzahl der Schemata.

## Motion Templates

Während viele auf Motion Templates basierende Verfahren darauf ausgelegt sind, einen kontextspezifischen Bewegungstyp zu erkennen, ermittelt Urtasun [UF03] mittels Hauptkomponentenanalyse auf einer Multiaktivitätsdatenbank jene Parameter, welche geeignet sind, eine Vielzahl von Bewegungen klar voneinander zu unterscheiden. Häufig wird auch der Suchraum eingegrenzt, indem man sich auf einen problemspezifischen Teilbereich beschränkt. Diese Einschränkung kann z. B. darin bestehen, dass lediglich eine begrenzte Menge an Bewegungen eindeutig erkannt werden soll. Sollen etwa nur Laufbewegungen erkannt werden, so werden diese im Vorfeld gelernt, um signifikante Bewegungsmuster zu identifizieren. Während dies im Bereich der Gestenerkennung ein vielversprechender Ansatz ist, besteht im Kontext der Bewegungsanalyse die Gefahr, dass relevante Abweichungen von einem idealen Bewegungstyp durch diese Verfahren eliminiert werden und so z. B. Gelenkfehlstellungen nicht erkannt werden.

## Weitere Ansätze

Deutscher[DR03] stellt 2005 eine Fortsetzung seiner Arbeit[DBR09] vor, welche den Suchraum mittels einer *weichen Partitionierung* und einem genetischen Algorithmus zur Partikelmutation effektiver bearbeitet. Das vorgestellte Partitionierungsschema ist als Alternative zu einer harten Suchraumunterteilung zu verstehen und arbeitet ohne *a priori* Wissen über Modellzusammenhänge. Der Mutationsoperator nimmt eine Kreuzung gut bewerteter Partikel vor, um deren Parametrisierung zu mischen.

### 3.3 Beobachtungsraum und Sensordaten

Marker- oder sensorbasierte Systeme stellen aufgrund ihrer langjährigen Verfügbarkeit und tendenziell geringen Auswertungskomplexität in vielen Bereichen den Stand der Technik dar. Je nach Bereich werden optische, akustische oder mechanische Marker und Sensoren eingesetzt. Einen kurzen Überblick gewährt die folgende Auflistung:

- **optisch:** Visuelle Marker, sowohl natürliche als auch künstlich angebrachte Merkmale
- **akustisch:** Ultraschallsensoren, Abstandsmessung
- **mechanisch:** mechanische Ermittlung von Gelenkwinkeln, z. B. Datenhandschuhe
- **elektromagnetisch:** auf Radio- oder Mikrowellen basierende Techniken
- **elektronisch:** Intertia-Sensoren, Gyroscope

Neben den genannten Vorteilen existieren jedoch zwei wesentliche Nachteile, welche sowohl Akzeptanz als auch Genauigkeit betreffen. So werden z. B. beim Sportler-Tracking die Probanden durch angebrachte Marker in ihrem natürlichen Bewegungsfreiraum eingeschränkt oder durch die spürbaren Marker und Sensoren irritiert. Beides führt zu Abweichungen im Bewegungsablauf und damit zu einer Verfälschung der Ergebnisse. Zusätzlich ist die Anbringung von Markern auf Landmarken schwierig, da sich die flexible Haut relativ zu diesen Fixpunkten verschiebt und damit die Markerposition nicht wie angenommen fest ist. Auch die Masseträgheit der Sensoren selbst begünstigt diese Verschiebung zusätzlich. Während bei markerbasierten Systemen der Beobachtungsraum durch die gewählte Technologie klar definiert ist, bieten sich bei sichtbasierten, markerlosen Systemen verschiedene Bildeigenschaften zur Beobachtung an.

### 3.4 Taxonomie markerloser Trackingverfahren

In der folgenden Tabelle (3.3) sind exemplarisch die Erkenntnisse dieses Kapitels zusammengefasst. Die wesentlichen Veröffentlichungen zu den zuvor genannten Verfahren werden aufgeführt. Weiterhin sind alle wesentlichen Bestandteile der jeweiligen Umsetzung der Tabelle zu entnehmen.

Quelle	Körpermodell	Vorhersage	Meßmodell	Einschränkungen und Besonderheiten
Articulated Body Motion Capture by Stochastic Search [DBR09]	Starrkörperkette, Zylinder	Partikelfilter, Simulierte Abkühlung	Kanten, Silhouette	weiche Partitionierung, Partikel-Mutation
3D Human Body Tracking using Deterministic Temporal Motion Models [UF03]	nur zeitliches Bewegungsmodell	-	Bewegungsdatenbank, PCA	Multiaktivitätsdatenbank
A Human Body Model for Articulated 3D Pose Tracking [KVD07]	Zylindermodell	-	Einpassung mittels ICP	Integration von 2D und 3D Daten möglich
A Model-Based Approach for Estimating Human 3D Poses in Static Images [LC06]	generalisierte Zylinder, Starrkörperkette	Partikelfilter	Silhouette	<i>auxiliary measurements</i>
Accurate Human Motion Capture Using an Ergonomics-Based Anthropometric Human Model [BEB08]	ergonomisches Modell (Ramsis)	Partikelfilter, Simulierte Abkühlung	Mesh-Projektion, symmetrischer Silhouettenvergleich	adaptive Meshes, kinetische Einschränkungen
Model-Based Hand Tracking Using an Unscented Kalman Filter [SMC01]	Handmodell, Quadrics	Unscented Kalman Filter	Konturvergleich	unimodaler Ansatz

Tabelle 3.3: Taxonomie wegweisender Trackingverfahren





# Kapitel 4

## Eigener Ansatz zur Pose-Rekonstruktion

In den vorherigen Kapiteln wurden die einzelnen Problemfelder bei der Konzeption einer Trackinganwendung vorgestellt. Verschiedene Lösungsansätze wurden präsentiert und ein Überblick bezüglich deren aktueller wissenschaftlichen Diskussion gegeben. Die sorgfältige Auswahl der Einzelkomponenten bestimmt wesentlich die Effizienz des Tracking-Verfahrens und ist von der jeweiligen Zielsetzung abhängig. Im Folgenden werden die gewählten Bestandteile aufgeführt und die Entscheidungen begründet. Den Anfang bildet eine Vorstellung des Anwendungskontextes, welcher die anschließenden Abschnitte motiviert.

### 4.1 Anwendungskontext und Vorüberlegungen

Im Rahmen einer Kooperation der Arbeitsgruppe *Human Motion Analysis* des Karlsruher Instituts für Technologie mit dem Institut für Sport und Sportwissenschaft ist innerhalb der letzten vier Jahre (2006 - 2010) am BioMotion Center in Karlsruhe eine umfangreiche Sammlung von Bewegungsaufnahmen entstanden. Mittels eines Mehrkameranysystems mit bis zu acht Hochgeschwindigkeitskameras und einem VICON<sup>©</sup> Aufnahmesystem wurden diverse Bewegungsabläufe aufgezeichnet.

Die Marker wurden in einem weiteren Arbeitsschritt Modellpunkten zugeordnet und so halbautomatisch Bewegungstrajektorien dieser Punkte erzeugt. In [FDW09] wurde ein adaptives Verfahren zur Vordergrund- / Hintergrundsegmentierung vorgestellt, welches mittels probabilistischer Fusion eine Volumenrekonstruktion vornimmt. Basierend auf diesem Verfahren wurde aus den Aufnahmen eine räumliche Darstellung erzeugt, welche die Szene in Form eines Belegungsgitters entspricht.

Wie in Abbildung 4.1 zu sehen ist, erfolgt eine *dichte Volumenrekonstruktion*. Dies bedeutet, dass alle Voxel entlang eines Sichtstrahls als belegt angenommen werden,

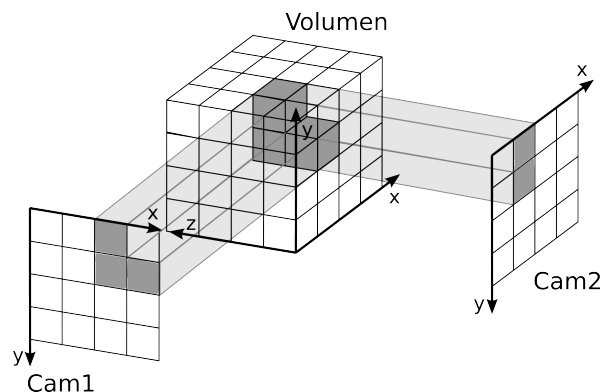


Abbildung 4.1: Volumenrekonstruktion: Aufbau eines Belegungsgitters mittels Rückprojektion der Vordergrund-Masken in den Raum.

solange keine weitere Rückprojektion hierzu im Widerspruch steht. Auf diese Weise wird nicht nur die visuelle Hülle rekonstruiert, sondern auch das Innere eines Körpers als belegt markiert.

Neben der Möglichkeit einer volumetrischen Rekonstruktion unterscheiden sich die vorliegenden Aufnahmen in zwei weiteren Aspekten von vielen verfügbaren Bewegungsdatenbanken. Die erfassten Bewegungsabläufe sind, verglichen mit einfachen Laufbewegungen oder Greif- und Zeigegesten, äußerst komplex. Dies gilt insbesondere mit Blick auf die auftretenden Winkelgeschwindigkeiten und Selbstverdeckungen. Weiterhin wurden alle Aufnahmen konsequent sowohl mit einem markerbasierten System als auch einem kalibrierten Mehrkameranystem für Hochgeschwindigkeitsaufnahmen durchgeführt. Diese Aufnahmen bilden die Datengrundlage der vorliegenden Arbeit, welche als reines 3D-Verfahren auf den rekonstruierten Volumen arbeitet.

## 4.2 Aufbau des Verfahrens

Ziel dieser Arbeit ist die Pose-Rekonstruktion mittels geeignetem Körpermodell und den zuvor beschriebenen Daten ist.

*Hierzu wird ein multimodaler, probabilistischer Ansatz verwendet, welcher ein artikuliertes Starrkörpermodell und Volumendaten mit einem Messmodell im dreidimensionalen Raum verknüpft.*

Ein Großteil der bereits vorgestellten Verfahren mit dreidimensionalem Menschmodell basiert auf einer Projektion dieses Modells in den Bildraum, welcher gleichzeitig als Messraum dient, um den Abstand zwischen Modell und Beobachtungen im Bildraum zu messen. Dieser Ansatz bietet den Vorteil einer hohen Effizienz, da einfach umzusetzende und erprobte Abstandsmaße für den zweidimensionalen Bildraum existieren. In dieser Arbeit wird jedoch ein Ansatz verfolgt, welcher eine

Modellprojektion in einen Raum niedrigerer Dimensionalität vermeidet. Eine genauere Messung und einfachere Integration von Sensordaten sind hierbei das Ziel.

Eine deterministische Vorhersage der genauen Modellparametrisierung ist aufgrund des multimodalen Zustandsraumes [WR05; DR03; ST03] nicht zu erwarten. Diese Multidimensionalität ist bedingt durch Mess- oder Rekonstruktionsfehler sowie Mehrdeutigkeiten in der Modellausrichtung und Zuordnung von Merkmalen. Die Diskrepanz der Dimensionalitäten von Zustands- und Beobachtungsraum verstärken das Problem zusätzlich. Die in den Kapiteln 2 und 3 diskutierten Probleme im Umgang von Partikelfiltern mit hochdimensionalen Räumen werden durch den Einsatz eines Annealed Particle Filters [DBR09] in Kombination mit einer Suchraumpartitionierung [MI00] berücksichtigt.

Auf ein explizites Bewegungsmodell wird zugunsten einer stochastischen Partikelbewegung verzichtet. Selbst bei den vorliegenden Videosequenzen mit hoher Framerate birgt die Extrapolation der Gelenkparameter durch Winkelgeschwindigkeiten und Kräfte Risiken, da sich Fehler aus vorhergehenden Zeitschritten aufsummieren. Insbesondere Verfahren mit lokaler Linearisierung leiden unter der Unterabtastung der auftretenden Kräfte. Das verwendete Menschmodell realisiert eine auf einer Voxel-Darstellung basierende Starrkörper-Repräsentation, welche erst die gewünschte Vergleichsmethode (siehe Abschnitt 4.6) ermöglicht. Es bietet weiterhin durch eine hohe Kompatibilität zu etablierten Analysewerkzeugen eine Möglichkeit zur Evaluation der erzielten Ergebnisse.

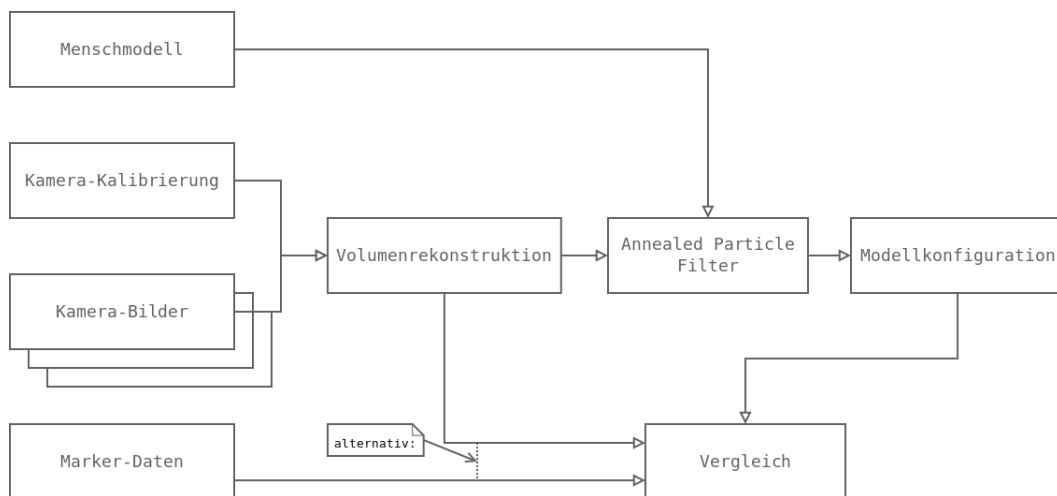


Abbildung 4.2: Datenflussdiagramm: Rekonstruktion, Vorhersage und Evaluation.

In Abbildung 4.2 wird der Ablauf des gesamten Verfahrens dargestellt: Aus den Kalibrierinformationen und Bildfolgen der einzelnen Kameras wird ein Volumen rekonstruiert. Mittels eines Annealed Particle Filters wird das Menschmodell möglichst exakt in dieses Volumen eingepasst. Ein abschließender Vergleich mit Markerdaten oder dem rekonstruierten Volumen liefert Aufschluss über die erzielte Genauigkeit.

Die folgenden Abschnitte stellen die einzelnen Komponenten des Verfahrens vor. Ihre Abfolge orientiert sich am Datenfluss, wie er dem gezeigten Diagramm zu entnehmen ist. Die Rekonstruktion des Volumens aus Kameradaten erfolgt wie in Kapitel 2 beschrieben und wird daher nicht gesondert betrachtet.

### 4.3 Körpermodell: volumetrische Starrkörper

Im Rahmen des Sonderforschungsbereiches 588 – *Humanoide Roboter – Lernende und kooperierende multimodale Roboter* entstand an der Universität Karlsruhe das Simulationswerkzeug *MkdTools* als Bestandteil der Dissertation „*Zur Modellierung und Simulation biomechanischer Mehrkörpersysteme*“ [Ste09].

Das mit Hilfe von Matlab<sup>©</sup> realisierte Werkzeug erlaubt die Bewegungssimulation von Starrkörperketten, sowohl mit vor- als auch rückwärtiger Kinematik. Hierbei werden nicht nur die reinen Lageinformationen sondern auch die sich aus der Bewegung ergebenden Geschwindigkeiten, Beschleunigungen und Kräfte berechnet. Ferner existiert die Möglichkeit das verwendete Starrkörper-Modell automatisch mittels VICON<sup>©</sup>-Markern bezüglich der Segmentlängen zu optimieren, was eine einfache und exakte Anpassung auf verschiedene Aufnahmesequenzen erlaubt. Die Markerpositionen sind als körperfeste Bestandteile modelliert und erlauben daher einen effektiven Vergleich von ermittelter und tatsächlicher Position und Ausrichtung.

Für erste Experimente wurde eine Matlab-C++-Schnittstelle implementiert, welche eine Nutzung des Modells innerhalb der verwendeten Entwicklungsumgebung ermöglichte. Hierbei wurde neben der Möglichkeit, das Modell mittels C++ zu parametrisieren und die sich ergebenden Gelenk-Posen auszulesen, auch darauf geachtet, möglichst viel Modellwissen verfügbar zu machen. Auch wenn niedrige Filterlaufzeiten nicht im Fokus dieser Arbeit stehen, stellten sich die aus dieser Umsetzung resultierenden Laufzeiten als unpraktikabel heraus. Die Berechnungszeit einer einzelnen Modellpose (ohne Vergleich mit den Beobachtungsdaten) liegt auf einem aktuellen System bei rund 200 Millisekunden, wobei 90% der Zeit auf die Socket-Kommunikation zwischen Matlab und C++ zurückzuführen sind.

Unter Berücksichtigung der verwendeten Partikelzahlen ergeben sich hieraus für die reine Ausrichtung des Modells bei 6000 Partikeln schon 20 Minuten Rechenzeit, den eigentlich rechenintensiven Part – in Form des volumetrischen Vergleichs – noch nicht berücksichtigt.

Eine native C++-Implementierung zur Modellierung von Starrkörperketten wurde daher umgesetzt, welche alternativ zu der bestehenden Lösung eingesetzt werden kann. Die Umsetzung ist hinsichtlich der Vorwärtskinematik, also der Ausrichtung des Modells anhand eines Zustandsvektors, vollständig kompatibel, so dass eine Evaluation der Ergebnisse innerhalb der Matlab-Umgebung zu einem späteren Zeitpunkt problemlos möglich ist.

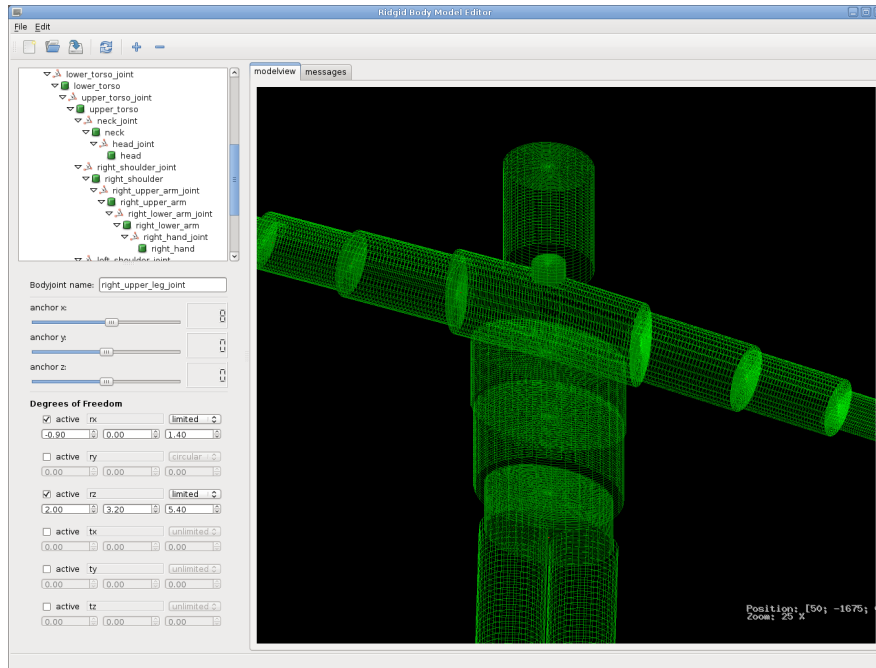


Abbildung 4.3: Modell-Editor: Aufbau einer Starrkörperkette und Einstellung der Freiheitsgrade

Das Körpermodell ist durch einen Szenegraphen definiert, welcher sich aus Geometrie- und Transformationsknoten in wechselnder Reihenfolge zusammensetzt. Die Definition des Modells erfolgt hierbei nicht statisch, sondern kann problemspezifisch erfolgen. Hierzu wurde ein Modellierungswerkzeug implementiert, welches die grafische Erzeugung und Manipulation von Starrkörperketten erlaubt (siehe Abbildung 4.3). Zur Modellierung der Starrkörper stehen unterschiedliche geometrische Primitive zur Verfügung, welche entsprechend parametrisiert werden können. Die Beschreibung der Gelenke erfolgt durch Angabe des Gelenkkordinatensystems und ihrer Freiheitsgrade. Einzelne Rotationsachsen und Translationsrichtungen können gezielt freigegeben oder gesperrt werden. Mit Blick auf den eingesetzten Partikelfilter ist eine Angabe des Wertebereiches möglich. Diese ermöglicht die korrekte Errechnung von Abständen sowie eine angepasste Randbehandlungen während der Partikelbewegung.

### 4.3.1 Koordinatensysteme

Im Rahmen dieser Arbeit werden ausschließlich karthesische, rechtshändige Koordinatensysteme verwendet. Jeder Starrkörper wird hierbei in einem eigenen Koordinatensystem definiert. Die Definition der Gelenkpunkte zu ihren Kindelementen (untergeordnete Knoten) erfolgt im selben lokalen Koordinatensystem, wobei die Ankerpunkte (Ausgangspunkt des Starrkörpers) generell volle 6 Freiheitsgrade besitzen. Der Koordinatenursprung kann sowohl in seiner Translation  $\vec{t} = [t_x, t_y, t_z]^T$ , als auch in seiner Rotation  $\vec{r} = [r_x, r_y, r_z]^T$  frei definiert werden. Im Gelenkpunkt selbst befindet sich das Gelenkkordinatensystem, welches die späteren Freiheitsgrade beherbergt. In Ab-

bildung 4.4 ist schematisch die Kopplung zweier Starrkörper aneinander zu sehen. Zu beachten ist, dass sich der Starrkörper selbst nicht, wie gezeigt, entlang der  $y$ -Achse ausdehnen muss – sowohl Rotation als auch Translation sind frei wählbar.

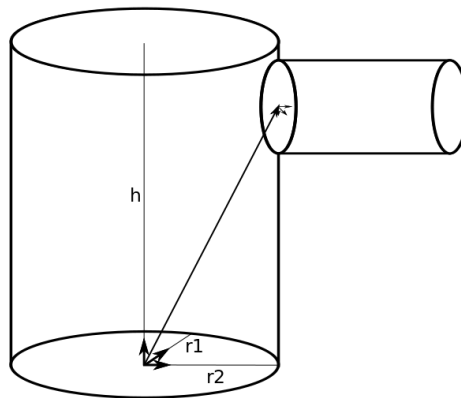


Abbildung 4.4: Aufbau der Starrkörper-Kette

Die Akkumulation der Eulerwinkel erfolgt hierbei nach DIN 9330, in der Reihenfolge  $z, y, x$  (Yaw-Pitch-Roll). Skalierungen und Scherungen des Koordinatensystems werden bewusst nicht ermöglicht. Gelenke mit einem translatorischen Freiheitsgrad, wie sie in [Ste09] vorgesehen sind, wurden in der aktuellen Implementation nicht berücksichtigt. Eine potentielle Ungenauigkeit, welche sich aus der minimalen Verschiebung der Starrkörperelemente ergeben könnte, liegt außerhalb der Genauigkeit, welche sich mit der momentanen Volumendiskretisierung erreichen ließe. Eine Erweiterung um beliebige Gelenkdarstellungen, wie sie etwa für komplexe Wirbelsäulenmodelle nötig wären, ist jederzeit möglich. Gelenkanschlüsse lassen sich als obere und untere Schranken für die Wertebereiche der drei Rotationsparameter definieren. Die Werte sind dabei so zu wählen, dass mindestens alle physiologisch möglichen Gelenkstellungen innerhalb des gültigen Wertebereiches liegen.

### 4.3.2 Rasterisierung

Ist die geometrische Beschreibung des Starrkörpers festgelegt, erfolgt eine Transformation dieser Form in ein Belegungsgitter. Hierzu wird eine einmalige Rasterisierung während einer Initialisierungsphase vorgenommen, da die Parameter des Starrkörpers zur Laufzeit als fest angesehen werden können. Jedes Belegungsgitter besitzt eine globale Position und Orientierung in Bezug auf ein festes gemeinsames Weltkoordinatensystem. Zusätzlich zu obigen Angaben besitzt jeder Voxelaum eine feste Größe (räumliche Ausdehnung) sowie einen Skalierungsfaktor, welcher die Umrechnung in eine definierte Größeneinheit erlaubt. Ein solches Belegungsgitter wird im weiteren Verlauf als Voxelaum bezeichnet. In Abbildung 4.5 ist eine frontale Sicht, sowohl auf das Starrkörpermodell, als auch die sich ergebende Voxeldarstellung gegeben.

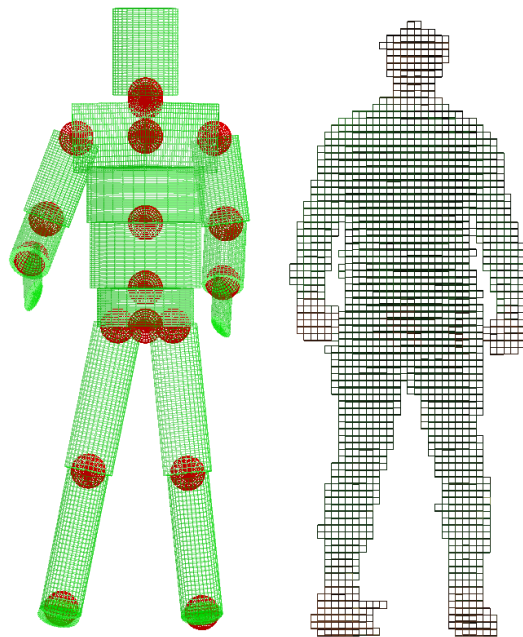


Abbildung 4.5: Vergleich: Starrkörpermodell und Voxeldarstellung

Die Anwendung einer Modellpose auf die Belegungsgitter erfolgt erst während einer Visualisierungs- oder Vergleichsphase. Dies geschieht durch eine affine Transformation des Voxelaums in das jeweilige Zielkoordinatensystem, welches sich durch Akkumulation aller Transformationen entlang des zugehörigen Szenegraphen ergibt.

Zur Modellierung der einzelnen Starrkörper werden vornehmlich gerade Zylinder mit elliptischer Grundfläche genutzt. Diese Darstellung führt zu einer Parametrisierung des Körpers mittels zweier Radien ( $r_1, r_2$ ) sowie der Zylinderhöhe ( $h$ ). Die binäre Voxelbelegung (0=unbelegt, 1=belegt) für  $x, y, z \in \mathbb{N}$  ergibt sich nach Gleichung 4.1.

$$v(x, y, z) = \begin{cases} 1 & (\frac{x^2}{r_1^2} + \frac{y^2}{r_2^2} \leq 1) \& (z \leq h) \\ 0 & \text{sonst} \end{cases} \quad (4.1)$$



### 4.3.3 Behandlung von Winkeln und Gelenkanschlügen

Die Verwendung eines zyklischen Zahlenraums, wie er bei Winkeln auftritt  $[0^\circ \dots 360^\circ]$  bedarf einer gesonderten Behandlung, wenn statistische Größen ermittelt werden sollen. Sämtliche Skalare des Parametervektors, entsprechen – mit Ausnahme der globalen Translation des Modells – den Gelenkwinkeln der Starrkörperkette. Die Winkel werden in absoluten Werten angegeben, beziehen sich also immer auf die Ruhelage des Modells. Wird zu einem Winkel durch das Zustandsübergangsmodell ein weiterer Winkel hinzuaddiert, so ist eine Überschreitung des Wertebereichs möglich. Eine Abbildung des Ergebnisses in den Wertebereich ist also an dieser Stelle erforderlich. Soll, etwa für die Berechnung der Varianz, die Winkeldifferenz  $wd$  zweier Winkel berechnet werden, so ist jeweils der kleinere der beiden sich ergebenden Drehwinkel zu verwenden  $wd(\alpha, \beta) = \min(|\alpha - \beta|, 360 - |\alpha - \beta|)$ . Diese einheitliche Betrachtung führt zu einer konsistenten Berechnung aller betrachteter Größen und entspricht gleichzeitig dem Wunsch, mit einer möglichst geringen Veränderung eine Modellausrichtung vorzunehmen. Um die genannten Aspekte und ebenso die notwendigen Gelenkanschlüge möglichst transparent in den Partikelfilter zu integrieren, wurden spezielle Datentypen implementiert, welche die erforderlichen Funktionalitäten umsetzen. Details hierzu sind in Anhang A zu finden.

## 4.4 Simulierte Abkühlung und der Annealed ParticleFilter

Die Simulierte Abkühlung (*simulated annealing*) ist ein iteratives, heuristisches Optimierungsverfahren, welches sich insbesondere für hochdimensionale Optimierungsprobleme eignet, die sich durch ein kombinatorisches Vorgehen nicht mit akzeptablem Aufwand lösen lassen. Der algorithmische Ansatz der Simulierten Abkühlung geht auf [KGV07] zurück. Eine Integration dieses Verfahrens in einen Partikelfilter wurde erstmals von Jonathan Deutscher et al. durchgeführt [DBR09]. Das Verfahren ähnelt in seinem Vorgehen den bekannten SIR-Filtern (auch Condensation, Partikelfilter etc.), da es ebenfalls einen stochastischen Ansatz verfolgt. Die im Rahmen der Simulierten Abkühlung erzeugten und bewegten Partikel sind jedoch nicht entsprechend der zugrunde liegenden (unbekannten) Verteilungsfunktion gestreut, so dass das Verfahren **nicht mit dem Bayes-Theorem konform** ist.

Mit der Simulierten Abkühlung wird ein natürlicher Abkühlungsprozess nachempfunden, wie er z. B. beim Abkühlen von Metallen auftritt. Hierbei müssen sich die Atome während der Abkühlphase so neu ausrichten, dass ein möglichst energiearmer (optimaler) Zustand erreicht wird. Übertragen auf algorithmische Probleme ergibt sich hieraus eine Bewegung der Partikel in mehreren zeitdiskreten Schritten, wobei die monoton fallende Temperatur der Wahrscheinlichkeit entspricht, dass sich die Ergebnisse des aktuellen Zeitschrittes auch verschlechtern dürfen. Dieser Ansatz bewirkt, dass das Verfahren in der Lage ist, lokale Optima auch wieder zu verlassen

und ermöglicht gleichzeitig durch eine sich verlangsamende Partikelbewegung, das Bereiche nahe tatsächlicher Optima mit einer zunehmenden Genauigkeit untersucht werden.

#### 4.4.1 Glättung der Gewichtsfunktionen

Die prinzipielle Funktionsweise eines Partikelfilters bleibt auch im Falle des Annealed Particle Filters erhalten und entspricht daher der Beschreibung in Abschnitt 2.3.4.

Zu jedem Zeitschritt wird jedoch nicht nur eine einmalige Gewichtung der Partikel vorgenommen sondern in Form eines iterativen Verfahrens wiederholt. Jede Iteration entspricht einer weiteren Abkühlphase, so dass die Partikelbewegung kontinuierlich abnimmt. Hierzu wird eine Reihe  $m = 0 \dots M$  von Wahrscheinlichkeitsverteilungen erzeugt, deren Wahrscheinlichkeitsdichten durch  $p_0(\vec{x}) \dots p_M(\vec{x})$  gegeben sind. Entscheidend ist, dass sich jede Wahrscheinlichkeitsdichte  $p_m(\vec{x})$  nur geringfügig von ihrem Nachfolger  $p_{m+1}(\vec{x})$  unterscheidet. Seien  $w_0(\vec{z}, \vec{x}), \dots, w_M(\vec{z}, \vec{x})$  Gewichtsfunktionen, wobei  $w_0$  eher schmalbandig mit spitz zulaufenden Modi und  $w_M$  eher geglättet ist. Erreicht wird dies durch

$$w_m(\vec{z}, \vec{x}) = w(\vec{z}, \vec{x})^{\beta_m} \quad (4.2)$$

wobei  $w(\vec{z}, \vec{x})$  die Originalgewichtungsfunktion bezeichnet, und  $\beta_0 \leq \beta_1 \leq \dots \leq \beta_M$  die entsprechenden Exponenten hierzu sind. Ziel dieses Vorgehens ist es, gezielt Partikel auszuwählen, welche in eine nächste Iteration übertragen werden. Die Partikelauswahl geschieht durch stochastisches Ziehen mit Zurücklegen, so dass die Partikelgewichtung, gesteuert durch die Gewichtsfunktion, der entscheidende Faktor für das *Überleben* eines Partikels ist.

#### 4.4.2 Ablauf eines Zeitschritts

Im Folgenden wird der Ablauf eines Filterdurchlaufs beschrieben. Dies umfasst sämtliche Schritte von einer ungewichteten Menge von Zuständen bis hin zu einer gewichteten, neu verteilten Menge, welche in die folgende Iteration übertragen wird. Sowohl die Initialisierung des Verfahrens als auch die Bestimmung eines Optimums werden separat betrachtet.

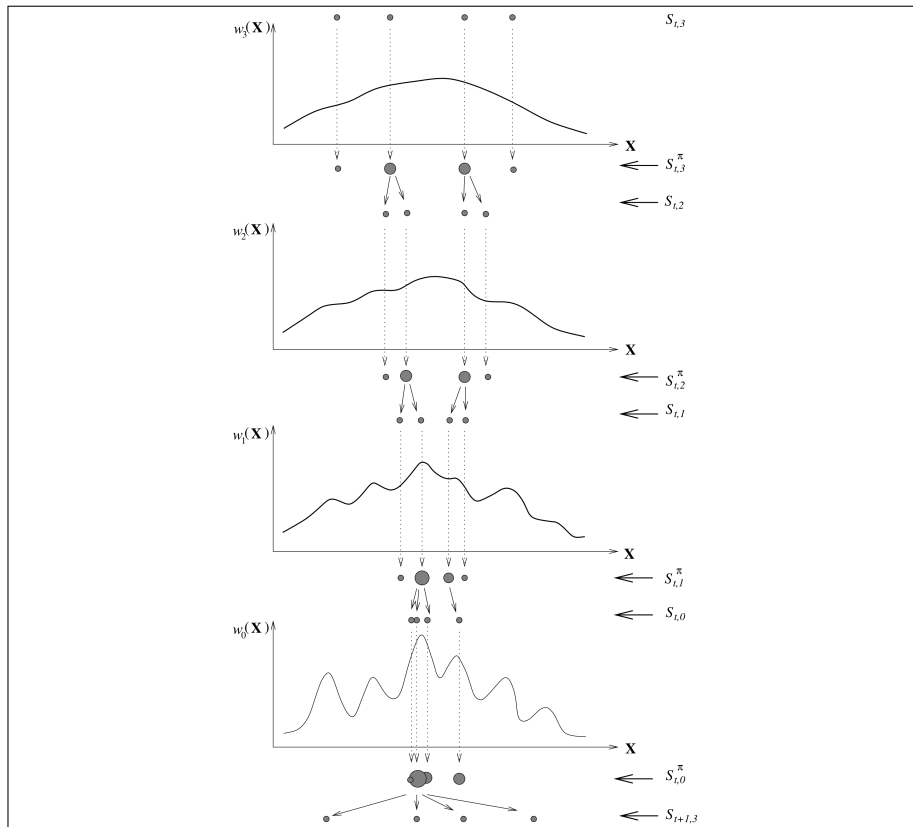


Abbildung 4.6: Zeitschritt eines Annealed Particle Filters: Verlauf der Partikelstreuung, Grafik aus Deutscher und Reid [DR03].

Abbildung 4.6 zeigt schematisch die Partikelbewegung innerhalb eines Zeitschritts über vier Abkühlphasen hinweg. Zu beachten ist die absteigende Reihenfolge der Indizes, begonnen wird also mit der Gewichtsfunktion  $w_3$ . Während anfänglich eine Auswahl der Partikel über einen sehr großen Abschnitt des Definitionsbereichs hinweg erfolgt, konzentriert sich im weiteren Verlauf die Auswahl zunehmend auf die lokalen Maxima der Funktion. Formell wird das Vorgehen, welches dieses Verhalten motiviert, in der folgenden Darstellung des Algorithmus erklärt (2):

---

**Algorithmus 2** Iterationsablauf bei Simulierter Abkühlung

---

1. Für jeden Zeitschritt  $t_t$ : Es wird in Layer  $M$  gestartet, mit  $m = M$ .
2. Jeder Layer wird mit einer ungewichteten Partikelmenge  $S_{t,m}$  initialisiert.
3. Jedem dieser Partikel wird ein Gewicht zugewiesen:

$$\pi_{t,m}^{(i)} \propto w_m(z_t, s_{t,m}^{(i)}) \quad (4.3)$$

wobei die Gewichte so normalisiert sind, dass  $\sum_N \pi_{t,m}^{(i)} = 1$  gilt. Die so generierte Menge wird mit  $S_{t,m}^\pi$  bezeichnet.

4. Es werden  $N$  Partikel aus  $S_{t,m}^\pi$  mit Zurücklegen gezogen, wobei die Wahrscheinlichkeit des Ziehens ihrer aktuellen Gewichtung entspricht.

Jeder Partikel geht hierbei wie folgt in seinen Nachfolger über:

$$s_{t,m-1}^{(n)} = s_{t,m}^{(n)} + B_m \quad (4.4)$$

$B_m$  ist hierbei eine multivariate gaussverteilte Zufallsvariable, mit Varianz  $P$  und Mittelwert  $0$ .

5. Das auf diesem Wege erzeugte Partikel-Set  $S_{t,m-1}$  wird als initialer Layer für den folgenden Annealing-Schritt  $m - 1$  genutzt. **Dieser Prozess wird wiederholt, bis  $S_{t,0}^\pi$  erreicht wird.**
6. Die optimale Pose wird aus  $S_{t,0}^\pi$  gemittelt:

$$\chi_t = \sum_{i=1}^N s_{t,0}^{(i)} \pi_{t,0}^{(i)} \quad (4.5)$$

7. Die initiale Partikelmenge des nächsten Partikelfilter-Schritts wird erzeugt durch:

$$s_{k+1,M}^{(n)} = s_{k,0}^{(n)} + B_0 \quad (4.6)$$

$B_0$  bezeichnet hierbei wie gehabt ein additives Rauschen.

---

Die Partikelbewegung, welche sich durch den zuvor beschriebenen Abkühlprozess ergibt, ist in Abbildung 4.7 dargestellt. Die Partikelmenge wird über 10 Layer hinweg optimiert, die dargestellten Freiheitsgrade entsprechen der globalen Rotation um die x- bzw y-Achse.

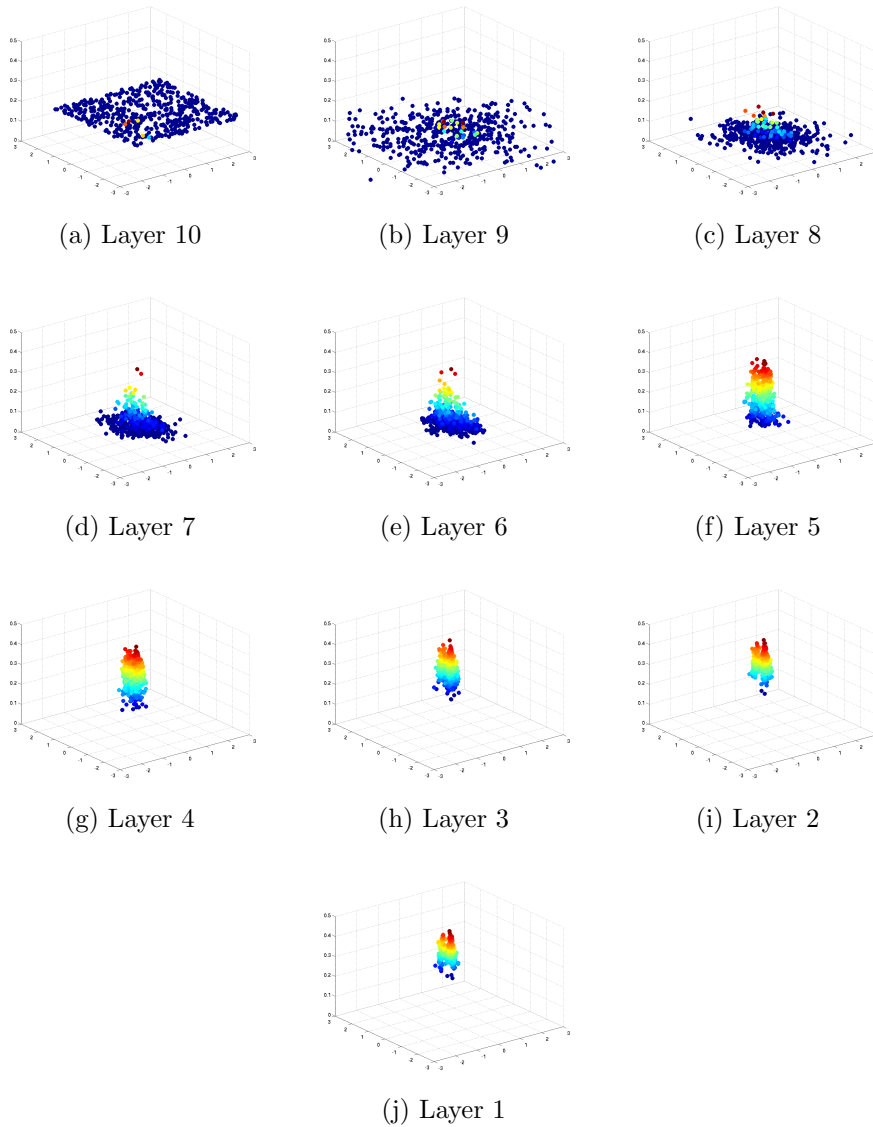


Abbildung 4.7: Partikelbewegung durch Simulierte Abkühlung, 2 rotatorische Freiheitsgrade des 1. Starrkörpers (x- und y-Achse).

### 4.4.3 Laufzeitparameter

In der bisherigen Betrachtung wurde die Bestimmung der Laufzeitparameter, welche die verwendeten Gewichtsfunktion festlegen, nicht vorgenommen.

Hierzu zählen sowohl die Exponenten der Gewichtsfunktion als auch die *Überlebensrate*,

welche die Wahl der Exponenten beeinflusst. Die Parameter werden im Folgenden vorgestellt, sowie Hinweise zu deren Auswahl gegeben.

### Überlebensrate $\alpha$

Die Überlebensrate (*survival rate*)  $\alpha$  ist ein Maß für die Partikelzahl, welche die Partikelwahl (resampling-Schritt) einer Abkühlphase überlebt. Zur Berechnung der Überlebensrate wird zuerst die *survival diagnostic*  $D$  berechnet:

$$D = \left( \sum_{i=1}^N \pi_i^2 \right)^{-1} \quad (4.7)$$

Die Bedeutung von  $D$  lässt sich an zwei einfachen Beispielen herleiten:

Liegt lediglich eine perfekte Hypothese vor  $((x_1, 1))$ , während die restlichen Partikel mit 0 gewichtet sind  $(\{(\vec{x}_i, 0)_{i=2}^N\})$ , so gilt  $D = 1$ . Dies würde zu dem Überleben eines einzelnen Partikels führen.

Ist jedoch die Wahrscheinlichkeit über alle Partikel gleich verteilt, gilt also  $\pi_i = \frac{1}{N} \forall i \in \{1, \dots, N\}$ , so ergibt sich:

$$D = \left( \sum_{i=1}^N N \frac{1}{N^2} \right)^{-1} = \left( N * \frac{1}{N^2} \right)^{-1} = N$$

Folglich würden alle Partikel den Auswahlprozess überleben. Eine Normierung der Gleichung führt letztendlich zu:

$$\alpha = \frac{D}{N} \quad (4.8)$$

Hieraus resultiert für die Überlebensrate ein Wertebereich zwischen 0 und 1.

### Die Gewichtungsexponenten $\beta$

Der Wert  $\beta$  bestimmt die Glattheit der Wahrscheinlichkeitsdichtefunktion im jeweiligen Annealing-Schritt und wird allgemein als Abkühlrate bezeichnet. Große Werte für  $\beta$  produzieren eine Kurve mit stärkeren Ausschlägen, kleine Werte haben den gegenteiligen Effekt.

Die Wahl von  $\beta$  erfolgt unter der Prämisse, eine festgelegte Überlebensrate erreichen zu wollen. Es kann gezeigt werden [DBR09], dass  $D(\beta)$  in Abhängigkeit von  $\beta$  monoton fallend ist, so dass die Gleichung

$$D(\beta) = N\alpha \quad (4.9)$$

eine eindeutige Lösung besitzt. Mittels Gradientenabstieg kann die Fehlerfunktion

$$\Delta_\alpha(\beta) = (\alpha_m - \alpha_{init}(\beta)) \quad (4.10)$$

gelöst und das gesuchte  $\beta_m^t$  bestimmt werden.

Als Startwert für jede Abkühlphase kann der ermittelte Wert des vorherigen Zeitschritts verwendet werden. Die Gewichte der einzelnen Partikel müssen nicht mit jeder Änderung von  $\beta_m^t$  neu berechnet werden. Da  $w_m(\vec{z}, \vec{x}) = w(\vec{z}, \vec{x})_m^\beta$  gilt, können einfach die Werte  $w(\vec{z}, \vec{x}) = s_{t,m}^{(i)}$  für  $i = 1 \dots N$  für jedes Tupel  $S_{t,m}$  gespeichert werden.

## Die Anzahl der Annealing-Layer

Die Anzahl der auszuwertenden Zustandsvektoren ergibt sich als Produkt aus der Anzahl an Partikeln pro Iteration und der Gesamtzahl der Iterationen. Verschiedene Parameterkombinationen mit einer gleichen Anzahl an Auswertungen sind daher denkbar. In [DBR09] wird beschrieben, dass bis zu einem gewissen Punkt, eine Verdopplung der Iterationen eine Halbierung der Partikelzahl zulässt. Generell ist jedoch die Wahl beider Parameter äußerst problemspezifisch. Die benötigte Anzahl an Partikeln pro Iteration ist stark von der Dimensionalität des Suchraums anhängig, wobei jedoch hochgradig multimodale Beobachtungen generell eine höhere Partikelzahl benötigen. Da eine Verbesserung der Ergebnisse ausschließlich durch wiederholtes Iterieren und das hiermit verbundene Abtasten interessanter Bereiche erreicht wird, ist die Iterationsanzahl verantwortlich für die Exaktheit der Ergebnisse.

### 4.4.4 Übergang zum nächsten Zeitschritt

Auch nach mehreren Annealing-Iterationen besteht keine Garantie, dass das tatsächliche Optimum dem Optimum der Partikelmenge entspricht. Es empfiehlt sich folglich ein defensives Vorgehen bei der Generierung der initialen Partikelmenge für den nächsten Zeitschritt. Würden lediglich die höchstgewichteten Partikel in den nächsten Zeitschritt übertragen, besteht die Gefahr, dass selbst unter Ausschöpfung der kinetischen und kinematischen Grenzen sowie der maximal möglichen Diffusion keine Partikel in den Wertebereich des tatsächlichen Optimums gestreut werden. Daher empfiehlt es sich, zusätzlich zu Partikeln in der Nähe des zuvor ermittelten Optimums weitere Partikel mit einer erhöhten Streuung einzufügen.

### 4.4.5 Diffusionsstrategien zur Partikelbewegung

Die Veränderung der Zustandsvektoren beim Übergang einer Abkühl-Iteration zur nächsten erfolgt unter Verwendung eines Rauschmodells. Die neue Position ergibt sich also aus der ursprünglichen Position und einem additiven Rauschterm.

Das Rauschen wird als normalverteilt vorausgesetzt, so dass eine Parametrisierung mittels Varianz und Mittelwert erfolgt. Werden weiterhin mittelwertsfreie Rauschterme vorausgesetzt, reduziert sich der Umfang einer Strategie zur Partikelbewegung (*Diffusionsstrategie*) auf die Ermittlung geeigneter Varianzen.

#### Konstante Reduzierung der Diffusion

Die einfachste Vorgehensweise besteht in einer konstanten Reduktion der Varianz. Die Varianz der ersten Iteration wird entsprechend der erwarteten maximalen Modellbewegung gewählt. Die Varianz aller folgenden Iterationen ergibt sich als Produkt der vorhergehenden Varianz und einem konstanten Faktor. Dieser Parameter muss problemspezifisch ermittelt werden, da eine zu schnelle Abnahme der Varianz zu einer unzureichenden Exploration des Suchraumes führt. Eine zu langsame Abnahme hingegen erhöht die Anzahl der benötigten Iterationen.

#### Varianzbasierte Diffusion

Die Varianz aller Partikel gibt Aufschluss darüber, wie stark die Partikel entlang der einzelnen Dimensionen des Zustandsraumes gestreut sind. Eine hohe Streuung kann als Indikator für eine unzureichende Ermittlung der Modi der beobachteten Wahrscheinlichkeitsdichte dienen. Hieraus leitet sich wiederum die Notwendigkeit ab, weitere Bereiche des Suchraumes zu untersuchen. Erst wenn gehäuft Partikelkonzentrationen auftreten und die Varianz sinkt, wird eine Reduzierung der Partikelstreuung vorgenommen.

#### Kovarianzbasierte Diffusion

Der vorherige Ansatz bietet den Vorteil, dass eine problemspezifische Wahl des Rauschens erfolgt, jedoch wird von einem unkorrelierten Rauschen der einzelnen Dimensionen ausgegangen. Abhilfe schafft hier die Verwendung einer Kovarianzmatrix, welche sich aus allen Partikeln errechnet. Die Initialisierung der Kovarianzmatrix erfolgt in Form einer Diagonalmatrix, welche die maximale Bewegung des Modells für einen Zeitschritt enthält.



#### 4.4.6 Suche in baumartigen Topologien

Werden alle Freiheitsgrade des Modells auf einmal ermittelt, so ergeben sich zwangsläufig Abhängigkeiten der Parameter untereinander.

Aufgrund der baumartigen Strukturierung der Starrkörperkette sind z.B. die Gelenkwinkel der Hände nur unter Berücksichtigung der Parametrisierung von Arm und Rumpf interpretierbar. Sind die zuvor genannten Körpersegmente also noch nicht annähernd optimal ausgerichtet ist keine sinnvolle Interpretation der Hand-Pose möglich. Je nach verwendetem Messmodell kann es sogar zu Fehlern kommen, welche das Filterverhalten negativ beeinflussen: Wird z.B. der Arm so ausgerichtet, dass es zu einer vollständigen Überlappung mit Rumpf oder Beinen kommt, kann das resultierende, verhältnismäßig hohe Überlappungsmaß dazu führen, dass der Filter sich in lokalen Maxima verfängt, welche aber auch in weiteren Schritten nicht zielführend genutzt werden können.

### 4.5 Weitere Verfahren zur Reduktion des Suchraumes

Als globales Optimierungsverfahren leidet die Simulierte Abkühlung unter einer Erhöhung der Suchraumdimensionalität. Eine rapide steigende Partikelzahl ist die Folge. Im Rahmen dieser Arbeit wird dem Problem auf zweierlei Art entgegengewirkt. Zum einen wird eine harte Partitionierung des Suchraums vorgenommen (Abschnitt 4.5.1), zum anderen wird versucht die verwendete Partikelzahl möglichst effizient einzusetzen.

#### 4.5.1 Harte Suchraumpartitionierung

Das Ziel einer Suchraumpartitionierung (*partitioned sampling*) [MI00] ist es, ein Problem in eine Summe von Teilproblemen zu zerlegen. Hierzu ist sowohl eine Zerlegbarkeit des Zustandsraumes als auch der Gewichtungsfunktion erforderlich.

Im Detail bedeutet dies, dass sich die Gesamtgewichtung aus einer Reihe bekannter Teilgewichtungen ergibt, deren Zusammenhang mit den Parametern des Zustandsvektors bekannt sein muß.

Übertragen auf den Kontext der Starrkörperphysik ergibt sich hieraus die Notwendigkeit, die Passgenauigkeit jedes Starrkörpers separat zu ermitteln. Bei der Zugehörigkeit von Parametern zu Starrkörpern ist der Kettencharakter des Modells zu beachten, da die Parametrisierung der übergeordneten Elemente immer die Position des aktuell betrachteten Körpers entscheidend beeinflusst. Eine hierarchische Suche, wie sie Gavrilla in [Gav99] verwendet, trägt dieser Tatsache Rechnung. In [MI00] wurde ebenfalls gezeigt, dass sich die benötigte Partikelzahl  $N$  auf eine Größe reduzieren lässt, welche lediglich von der Dimensionalität der größten Teilpartition abhängt.

Während eine Suchraumpartitionierung den Vorteil bietet, den Suchaufwand weitestgehend von der Dimensionalität des Suchraumes zu entkoppeln, da sich die Partikelzahl nur an der Größe der Teilräume orientiert, ist die Abhängigkeit der Ergebnisse innerhalb der einzelnen Partitionen als deutlicher Nachteil zu verbuchen. Eine Fehldetektion innerhalb einer der initialen Partitionen führt zwangsläufig zu schlechteren Ergebnissen innerhalb aller abhängigen Partitionen. Daher ist darauf zu achten, dass die Partitionierung so vorgenommen wird, dass die Lokalisation von Optima innerhalb dieser initialen Partition möglichst ohne Mehrdeutigkeiten möglich ist.

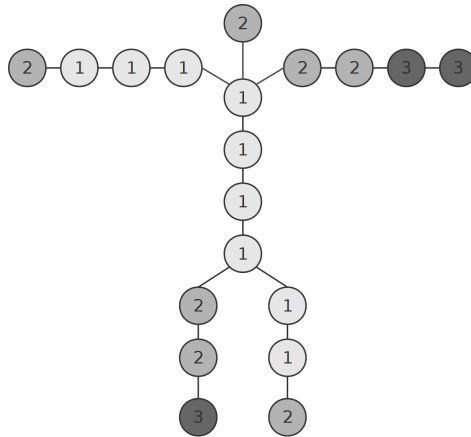


Abbildung 4.8: Partitionierungsschema zur Vermeidung von iterativen Fehlern.

Um das Risiko einer Fehlparametrisierung in frühen Partitionen zu reduzieren, wird ein Partitionierungsschema verwendet, welches Fehlinterpretationen vermeidet. Hierzu werden Teile der Extremitäten mit in die erste Partition integriert, eine mögliche Aufteilung ist in Abbildung 4.8 zu sehen. Das Erfordernis, sowohl einen Teil des Armes als auch eines Beines direkt korrekt auszurichten, lässt eine grobe Fehlpositionierung des Torsos unwahrscheinlich werden.

## 4.5.2 Optimierte Partikelstreuung

In Ermangelung eines kinematisch motivierten Bewegungsmodells erfolgt die Partikelbewegung aufgrund rein stochastischer Größen. Mangelnde Kenntnis der Umgebung eines Partikels führt dazu, dass sich Partikel durch den Diffusionsschritt wieder vom Optimum weg bewegen können. Auch wenn diese potentielle Verschlechterung von Ergebnissen ein elementarer Vorteil des Verfahrens sein kann, da so lokale Optima überwunden werden, ist es dennoch unvorteilhaft, die ursprünglichen, höher bewerteten Partikel vollständig in zukünftigen Iterationen zu ignorieren. Daher wird der Ansatz des stochastischen Ziehens um eine gewichtungs-basierte Partikelauswahl ergänzt. Hierzu wird ein gewisser Prozentsatz der initialen Partikelmenge der folgenden Abkühliteration aus den höchstgewichteten Partikeln der aktuellen Phase generiert. Im Gegensatz zu dem etwa in [DR03] vorgestellten Ansatz, welcher hierbei eine Kreuzung der Partikel untereinander vornimmt, erfolgt die Übernahme der Partikel

ohne Modifikation. Da besonders zu Anfang die Partikel aus verschiedenen Modi stammen, führt eine Kreuzung, welche dem *survival of the fittest* Prinzip genetischer Algorithmen entspricht, zu einer Vermischung der Modalitäten und damit zu Partikeln welche keinem der verwendeten lokalen Optima zuzuschreiben sind.

## 4.6 Dreidimensionales Belegungsgitter als Beobachtungsraum

Der Vergleich von Modell und Beobachtung im Bildraum erscheint aus zweierlei Gründen sinnvoll. Zum einen stehen hierfür etablierte Methoden der (zweidimensionalen) Bildverarbeitung zur Verfügung, zum anderen liefert das Beobachtungsmodell, unabhängig von den gewählten Deskriptoren, nur zweidimensionale Informationen.

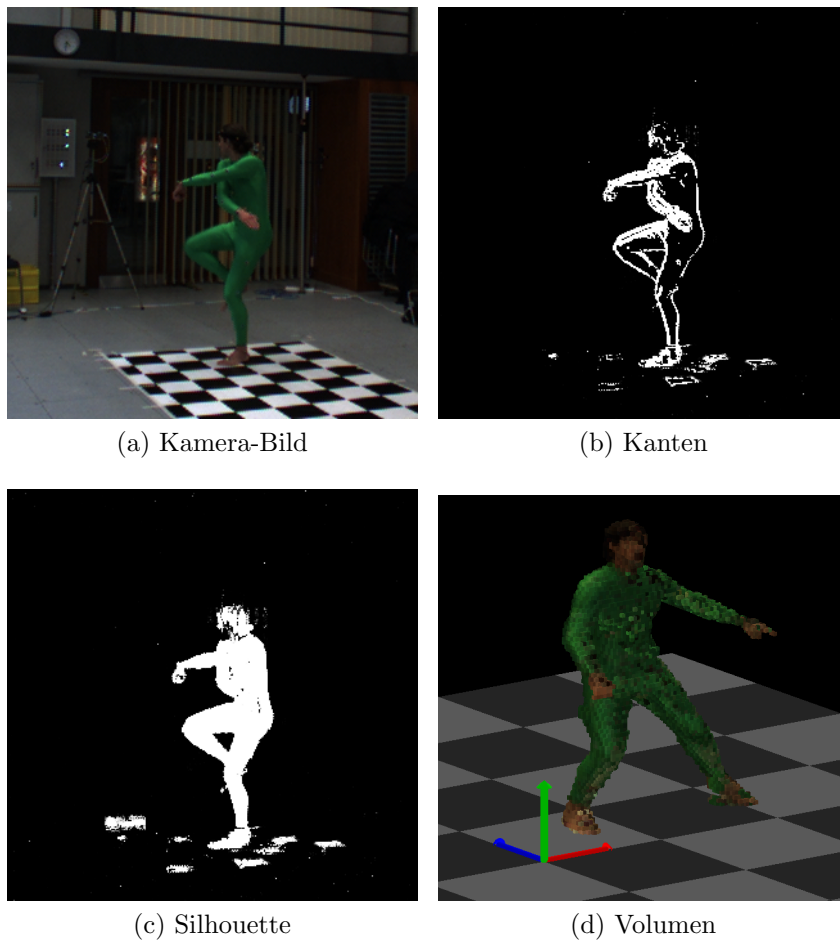


Abbildung 4.9: Gegenüberstellung verschiedene Beobachtungsräume

### 4.6.1 Symmetrisches Messmodell

Wird ein Starrkörpermodell durch ein Stickmodell, Konturpunkte oder z.B. eine linienbasierte Kontur beschrieben, so stellt sich die Frage nach einem geeigneten Vergleichsverfahren zur Messung der Übereinstimmung mit Beobachtungen im Bildraum. Hierzu werden häufig die Abstände der projizierten Modellpunkte zu Merkmalen im Bildraum gemessen. Bestimmen lediglich die erfolgreich projizierten Modellpunkte die Güte der Überlappung, so können prinzipiell mehrere Modellpunkte auf einen Bildpunkt abgebildet werden. Dies mag in einigen Fällen korrekte Ergebnisse liefern, kann aber auch degenerierte Posen begünstigen. Ein häufig anzutreffendes Phänomen ist die Rotation aller Extremitäten in das Volumen des Rumpfes. Da sich hier die größte Fläche befindet – und damit häufig auch die meisten Kanten oder Merkmale – liefert eine Abbildung in diesen Bereich schnell eine hohe Gewichtung. Ein Ansatz zur Vermeidung dieser Problematik ist die Verwendung von Silhouetten, sowohl als Vordergrund-Repräsentation (Kameradaten), als auch für die Modellbeschreibung. Mittels XOR-Differenz kann der Abbildungsfehler genauer bestimmt werden, da der nur durch eine Silhouette belegte Bereich möglichst klein ausfallen sollte. Da bei diesem Ansatz jedoch unmöglich entschieden werden kann, wie viele Starrkörper zur Belegung eines Silhouetten-Pixels beitragen, entstehen nicht auflösbare Mehrdeutigkeiten, wenn sich aus Kamerasicht z.B. der Arm vor dem Brustkorb befindet. Eine Strategie zur Abschwächung dieser Problematik ist die Fusion von Kanten und Silhouetten zur Auswertung der Gewichtsfunktion, was z. B. in [DR03] umgesetzt wird.

Erfolgt die Bewertung der Pose durch Projektion in mehrere Kamerasichten, ist eine anschließende Fusion der Bewertungen aller Einzelansichten nötig. Dies gestaltet sich schwierig, da die üblichen statistischen Maße, etwa in Form von Mittelwert, Minimum, Maximum etc., kaum Rückschlüsse auf die tatsächliche Lage zulassen. Eine geometrisch korrekte Interpretation unterschiedlicher Sensorantworten liefert die gewünschten Ergebnisse, ist jedoch nur aufwändig zu modellieren.

Die Voxelraum-Repräsentation sowohl des Beobachtungsraums als auch des Menschmodells ermöglicht die Realisierung einer Abstandsfunktion, welche die zuvor beschriebenen Probleme systembedingt vermeidet. Durch eine Fusion der Kamera-Daten **vor** der Auswertung der Abstandsfusion werden widersprüchliche Abstandsmaße vermieden. Der Wegfall einer Projektion führt zu einer eindeutigen Aussage bezüglich der Überlappungen. Die Schnittbildung zweier Voxelräume kann entweder innerhalb des Referenzkoordinatensystems eines der beiden Räume oder innerhalb eines dritten Raumes erfolgen. Einmalig muss für die zu schneidenden Räume die Transformationsmatrix zur Überführung der Koordinaten in den Vergleichsraum ermittelt werden. Anschließend werden Bounding-Boxen der Räume aneinander geklippt und ebenfalls der Transformation ins Zielkoordinatensystem unterworfen. Auf diese Weise kann der Raum, welcher für etwaige Schnittpunkte zu traversieren ist, erheblich reduziert werden. Es erfolgt eine Abtastung des Zielraumes. Eine Interpolation über die Ur-

sprungsräume ist möglich, wird jedoch zugunsten eines verringerten Aufwands für die Partikelgewichtung nicht verwendet.

Die Überlappung zweier Voxelräume entspricht der Menge der in beiden Voxelräumen als belegt markierten Voxel und dient als Maß für deren Übereinstimmung.

# Kapitel 5

## Experimente und Validierung

Das im vorherigen Kapitel vorgestellte Verfahren zur Exploration des Suchraums beschreibt das generelle Vorgehen zur Detektion der optimalen Modellkonfiguration. Eine konkrete Umsetzung des Verfahrens erfordert eine Reihe von problemspezifischen Entscheidungen, deren Effekte in diesem Kapitel dargestellt werden.

Innerhalb des nächsten Abschnittes (5.1) werden jene Bestandteile evaluiert, welche durch sich unterscheidende Strategien umgesetzt werden können. Hierzu zählen die Abstands- oder Gewichtungsfunktion sowie die Strategien zur Partikelstreuung und -selektion. In Abschnitt 5.2 werden die Auswirkungen der drei essenziellen Laufzeitparameter – der Abkühlrate sowie der Layer- und Partikelzahlen – betrachtet. Die Ergebnisse der durchgeführten Testläufe werden in Abschnitt 5.4 präsentiert. Den Abschluss bildet eine Betrachtung des Initialisierungsproblems, welches bei vollkommen unbekannter Modelausrichtung auftritt.

### 5.1 Evaluation von Einzelkomponenten des Partikelfilters

Die Multimodalität des Suchraums wurde bereits mehrfach als inhärente Eigenschaft der Pose-Rekonstruktion vorgestellt. Diese muss bei der Konzeptionierung des Partikelfilters beachtet werden, da nicht jedes lokale Optimum zu einer korrekten Pose entwickelt werden kann. In Abbildung 5.1 sind die Ergebnisse eines durchgeführten Experiments zu sehen. Hierzu wurde ein initial korrekt ausgerichtetes Modell durch Manipulation zweier Rotationsangaben durch die Voxelwolke des rekonstruierten Menschen hindurch gedreht. Das Rotationszentrum liegt hierbei in der Hüfte des Menschen, so dass eine Drehung um die y-Achse etwa der Drehung um die Längsachse des Menschen entspricht. Wie der Abbildung zu entnehmen ist, liefert die Abstandsfunktion in beiden Fällen multimodale Ergebnisse mit zwei bis drei lokalen Optima. Die x-Achse entspricht einer Winkelangabe im Bogenmaß, die Periodizität des Graphen ist daher zu beachten. Der Versatz zwischen den beiden prägnantesten lokalen Optima beträgt jeweils ungefähr  $\pi$ , entspricht also einer Drehung um  $180^\circ$ . Dies würde bedeuten, dass

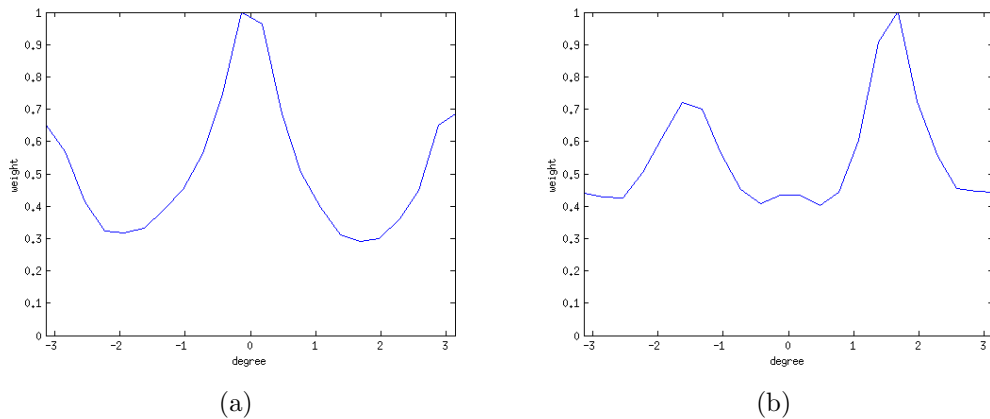


Abbildung 5.1: Darstellung des multimodalen Suchraums: Rotation des Modells um die (a)) x- und (b) y-Achse.

das Modell entweder mit entgegengesetzter Blickrichtung ausgerichtet wird oder einen ungewollten Kopfstand vollführt.

Um eine solche Fehlpositionierung zu korrigieren, muss eine Partikelbewegung stattfinden, welche die Distanz zwischen den beiden lokalen Maxima überwindet. Um also etwa während des Initialisierungsproblems (siehe C) durch diese  $180^\circ$ -Mehrdeutigkeit keine falsche Ausrichtung vorzunehmen, wird innerhalb des ersten Annealing-Layers eine Varianz von mehr als  $180^\circ$  gewählt, sofern die Gelenkansschläge dies zulassen.

### 5.1.1 Diffusionsstrategien

Die Partikelbewegung wird im Verlauf des Verfahrens durch die Diffusionstrategie immer weiter eingeschränkt. Dies stellt sicher, dass in den ersten Iterationsschritten ausreichend große Bereiche des Suchraumes untersucht werden und im weiteren Verlauf interessante Bereiche genauer abgetastet werden.

Reduziert ein Verfahren das additive Rauschen zu schnell, werden entweder keine Optima gefunden oder lokale Optima nicht mehr verlassen. Bleibt hingegen der Rauschterm auch in den letzten Iterationen zu hoch, werden Partikel nicht ausreichend optimiert. Eine bereits gefundene gute Pose lässt sich meist nur weiter optimieren, indem in ihrer unmittelbaren Nähe gesucht wird, was mit einem sehr geringen additiven Rauschen zu erreichen ist. In Abschnitt 4.4.5 werden drei Verfahren zur Steuerung der Partikelbewegung vorgestellt, deren Auswirkungen Abbildung 5.2 zu entnehmen sind.

Auffallend ist, dass mittels konstanter Reduzierung des Rauschens keine Verbesserung der Pose erreicht werden konnte. Das Rauschen sinkt zu schnell, das Verfahren verfängt sich in einem lokalen Maximum, welches nicht mehr verlassen werden kann. Die varianzbasierte Methode liefert bereits bessere Ergebnisse, jedoch wird hier die Korrelation der Suchraumdimensionen nicht berücksichtigt. Auch hier nimmt die Partikelstreuung zu schnell ab – das Verfahren stagniert etwa ab dem 10. Layer. Im Gegensatz hier-

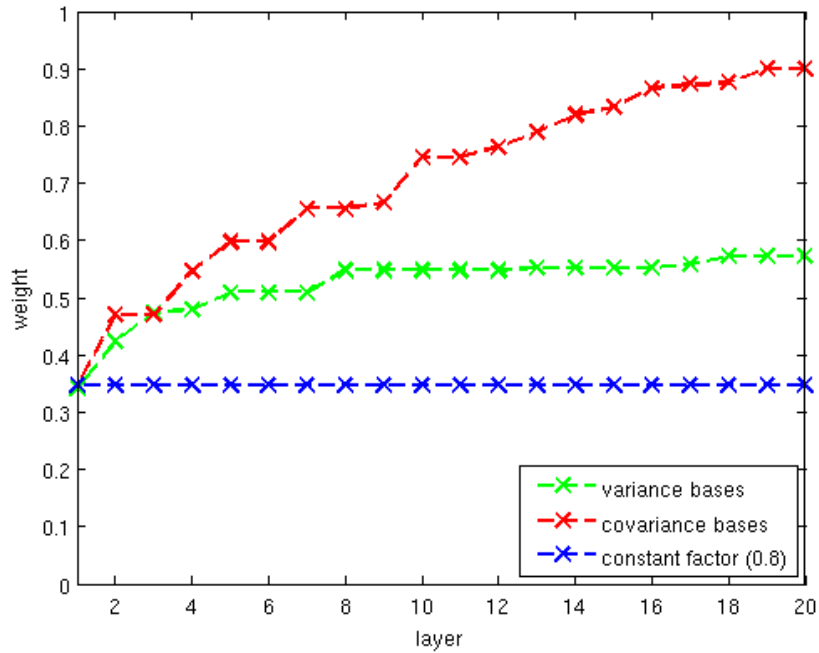


Abbildung 5.2: Gegenüberstellung der Diffusionsstrategien: kovarianzbasierte, varianzbasierte und konstant (abnehmende) Diffusion.

;

zu stehen die Ergebnisse der kovarianzbasierten Partikelstreuung, eine kontinuierliche Verbesserung findet hier über alle 20 Iterationen hinweg statt.

### 5.1.2 Einfluss der Abstandsfunktion

Das in Abschnitt 4.6.1 vorgestellte Messmodell liefert die Übereinstimmung zwischen einer Modellkonfiguration und der Volumen-Rekonstruktion des Menschen. Um den angeführten Vorteil einer dreidimensionalen Messung, welche Mehrfachzuordnungen eines Voxels vermeiden kann, zu belegen, wurde diese Prüfung von Körperteilüberlagerungen vorübergehend deaktiviert und die erzielten Ergebnisse dem ursprünglichen Verfahren gegenüber gestellt.



Die Ergebnisse eines durchgeführten Testlaufs sind Abbildung 5.3 zu sehen. Die obere Zeile (a-c) zeigt die Ergebnisse ohne Überlappungserkennung, die untere Zeile (d-f) stellt dem die Ergebnisse bei aktivierter Erkennung gegenüber. Das jeweils mittlere Bild zeigt die ermittelte Pose nach 20 Abkühl-Iterationen. Besonders die Fehlausrichtung des linken Armes ist deutlich zu erkennen. Während bis zur 13. Iteration vergleichbare Exponenten für die Gewichtsfunktion gewählt werden (rechte Abbildung), bildet sich danach bei einem Verzicht auf die Überlappungserkennung ein Plateau heraus. Die finale Partikelgewichtung fällt bei dem verwendeten Verfahren mit Überlappungserkennung etwa 25% höher aus. Dies ist sowohl auf die korrekte Zuordnung aller Körperteile als auch die – innerhalb der letzten Iterationen – effizientere Optimierung zurückzuführen. Ein optischer Vergleich der beiden Endposen (Abbildung 5.3 b und e) belegt die Vorzüge des verwendeten Verfahrens deutlich.

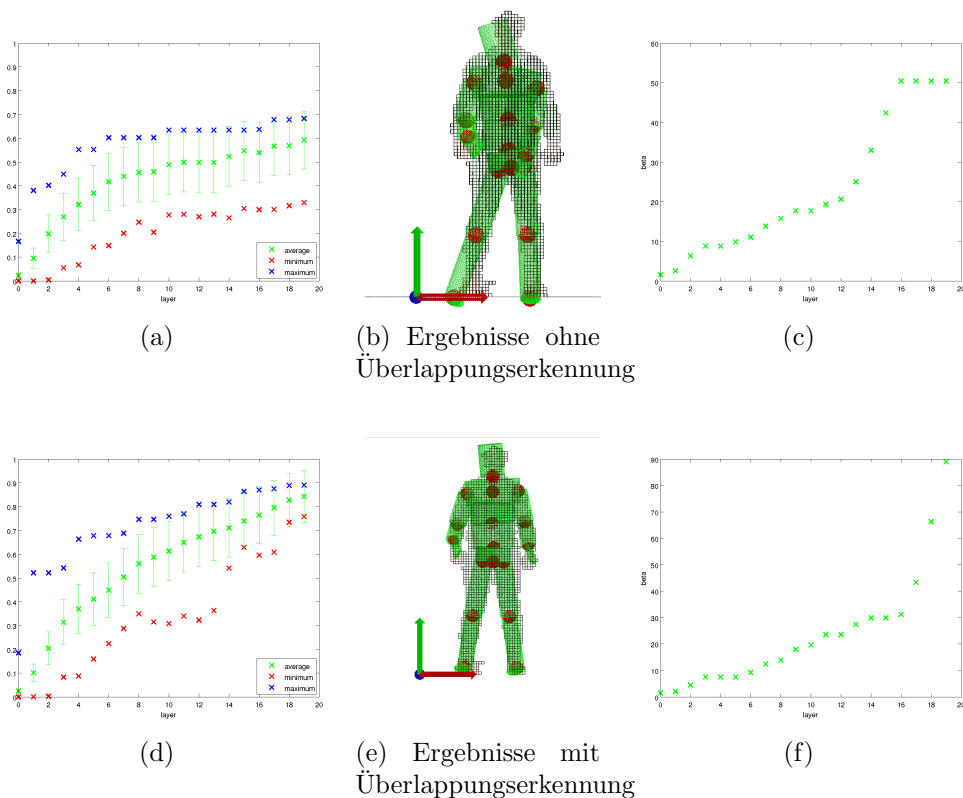


Abbildung 5.3: Abstandsmaße ohne (a-c) und mit Überlappungserkennung (d-f), von links nach rechts: Partikelgewichtung, finale Pose, Gewichtungsexponenten

### 5.1.3 Partikelselektion

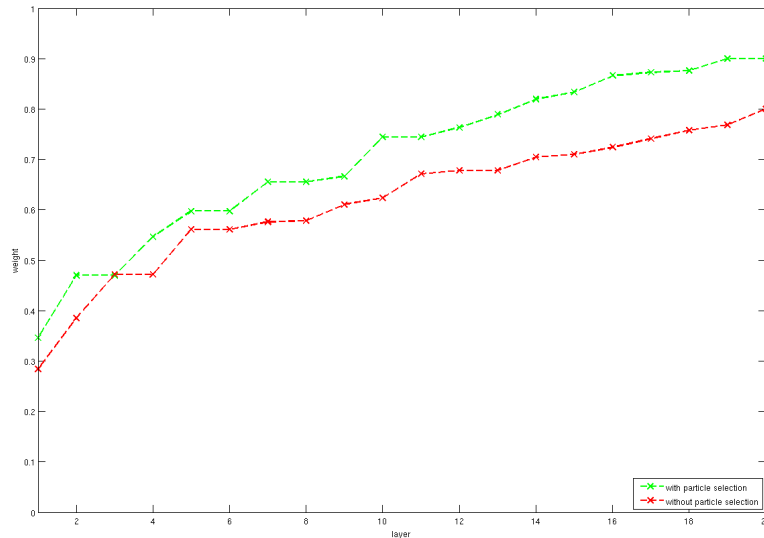


Abbildung 5.4: Gegenüberstellung zweier Strategien zur Partikelselektion: konventionelle Selektion und Selektion mit zusätzlichem Transfer der vorhergehenden Optima.

Die Übertragung der Partikel in die folgende Iteration (resampling-Schritt) erfolgt in der Originalfassung [DBR09] durch stochastisches Ziehen mit Zurücklegen, welches der Gewichtung der Partikel folgt. Auf die so gewonnene Partikelmenge wird direkt das Bewegungs- bzw. Rauschmodell angewandt, so dass die Optima der vorherigen Iteration verloren gehen. Um das Wissen vorangegangener Iterationen nutzen zu können, wird innerhalb dieser Arbeit zusätzlich ein Transfer der  $n$  besten Partikel vorgenommen. Diese werden nach dem Aktualisierungsschritt des Annealed Particle Filters ohne Anwendung eines Rauschmodells in die nächste Iteration übertragen. Der Effekt dieser Vorgehensweise ist in Abbildung 5.4 zu sehen: Besonders während der ersten Iterationen erfolgt aufgrund der hohen Varianz des Rauschens häufig eine Verschlechterung der Partikelgewichte. Die bereits gefundenen, höher gewichteten Partikel werden verworfen und stehen als Ausgangspunkt für weitere Partikelbewegungen nicht mehr zur Verfügung. Daher werden unnötig schlechte Startwerte für die Bewegung verwendet, was die Optimierung stark verlangsamt.

### 5.1.4 Suchraumpartitionierung

In Grafik 5.5 wird das vorgeschlagene Partitionierungsschema einer simultanen Optimierung aller Freiheitsgrade gegenüber gestellt. Die linke Abbildung zeigt das Ergebnis einer Suche innerhalb der ersten Partition, welche 20 Freiheitsgrade umfasst. Trotz dieser deutlichen Reduzierung des Suchaufwandes fallen die Abstandsbewertungen

schlechter aus. Dies dürfte auf die Mehrdeutigkeiten zurückzuführen sein, welche sich bei der Ausrichtung dieser stark verkürzten Starrkörperkette ergeben. Die Zahl der Partikel, welche zur Suche in der ersten Partition verwendet wurde, entsprach mit 1000 Partikeln der Hälfte dessen, was zur globalen Suche verwendet wurde. Eine weitere Reduzierung der Partikelzahl verschlechterte die Ergebnisse deutlich. Die verbleibenden 1000 Partikel wurden auf die 19 weiteren Freiheitsgrade des Modells verteilt, auch hier war keine deutliche Absenkung der Partikelzahl möglich. Vergleicht man unter diesen Gesichtspunkten die beiden Diagramme, so erscheint die gleichzeitige Optimierung aller Freiheitsgrade bei vergleichbarem Aufwand und geringerer Fehleranfälligkeit sinnvoller.

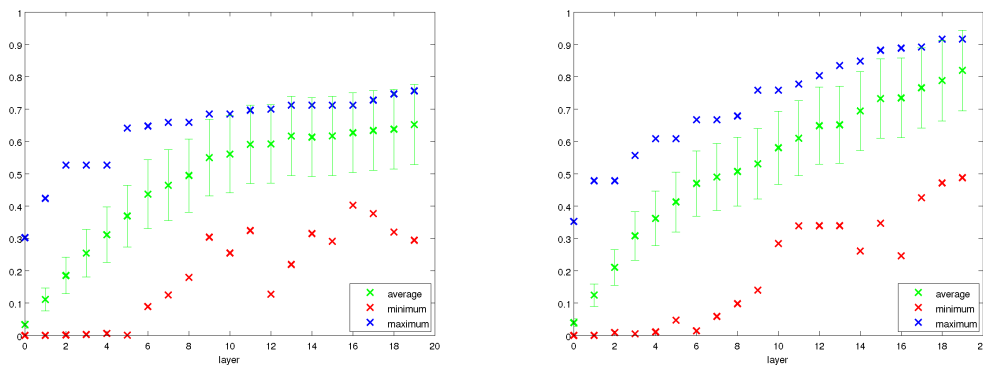


Abbildung 5.5: Vergleich von globaler Pose-Optimierung und partitionierter Suche.

## 5.2 Wahl der Laufzeit-Parameter

Die Effizienz des Annealed Particle Filters basiert maßgeblich auf einer möglichst optimalen Ausnutzung der vorhandenen Partikelzahl. Neben der Frage, auf wie viele Iterationen diese Partikel zu verteilen sind, ist zu klären, wie die Überlebensrate sinnvoll zu bestimmen ist.

### 5.2.1 Überlebensrate

Die in Abschnitt 4.4.3 beschriebene Überlebensrate der Partikel bestimmt die Wahl der Gewichtungsexponenten, welche zur Glättung oder Schärfung der Gewichtungsfunktion verwendet werden. Eine niedrige Überlebensrate sorgt dafür, dass nur Partikel mit einer vergleichsweise hohen Wahrscheinlichkeit den Resampling-Schritt überleben, während eine hohe Rate zu einer gleichmäßigeren Verteilung der gezogenen Partikel bewirkt. Zur Ermittlung geeigneter Werte wurden Experimente mit wechselnder Überlebensrate und ansonsten gleicher Filterparametrisierung durchgeführt. Die Ergebnisse für einen Zeitpunkt sind exemplarisch in Abbildung 5.6 dargestellt.

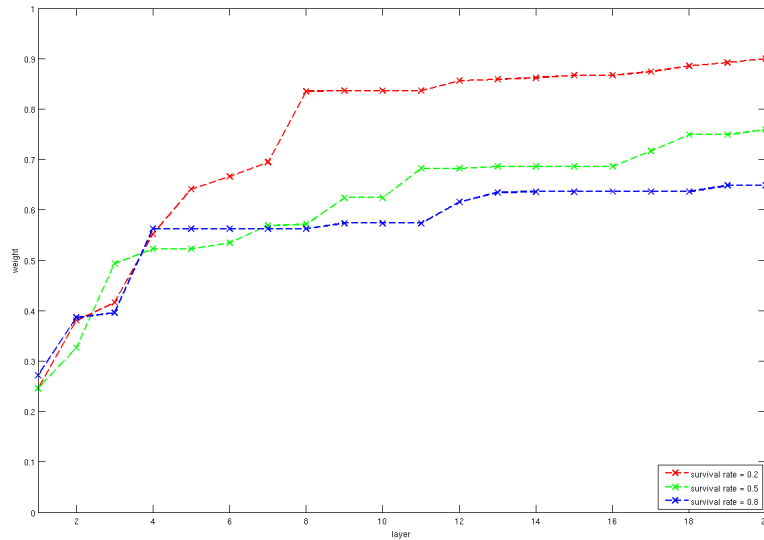


Abbildung 5.6: Gewichtsfunktion in Abhängigkeit von der gewählten Abkühlrate.

Wie der Grafik (Abbildung 5.6) zu entnehmen ist, führt eine niedrigere Überlebensrate besonders in den ersten Iterationen zu einem schnelleren Anstieg der Gewichte. Dies ist insbesondere deshalb von großem Vorteil, weil die Wahrscheinlichkeit weitere, neue Optima zu finden, mit steigender Iterationszahl und damit sinkender Partikelstreuung abnimmt.

### 5.2.2 Layer und Partikelzahl

Als Maß für die Effizienz eines Partikelfilters kann die Zahl der Partikel betrachtet werden, welche zum Erreichen eines gewissen Ergebnisses benötigt wird. Im Umkehrschluss soll also mit einer festen Partikelzahl ein möglichst gutes Ergebnis erreicht werden. Die effektive Partikelzahl ergibt sich für den Annealed Particle Filter als Produkt aus der Partikelzahl pro Iteration und der Zahl der durchzuführenden Iterationen. Die so ermittelte Größe entspricht der Zahl der auszuführenden Partikelgewichtungen und Bewegungen. Bei konstanter Partikelzahl sind verschiedene Aufteilungschemata denkbar. Abbildung 5.7 zeigt die gleichmäßige Verteilung von 10.000 Partikeln auf 10, 20, 40 und 80 Iterationen.

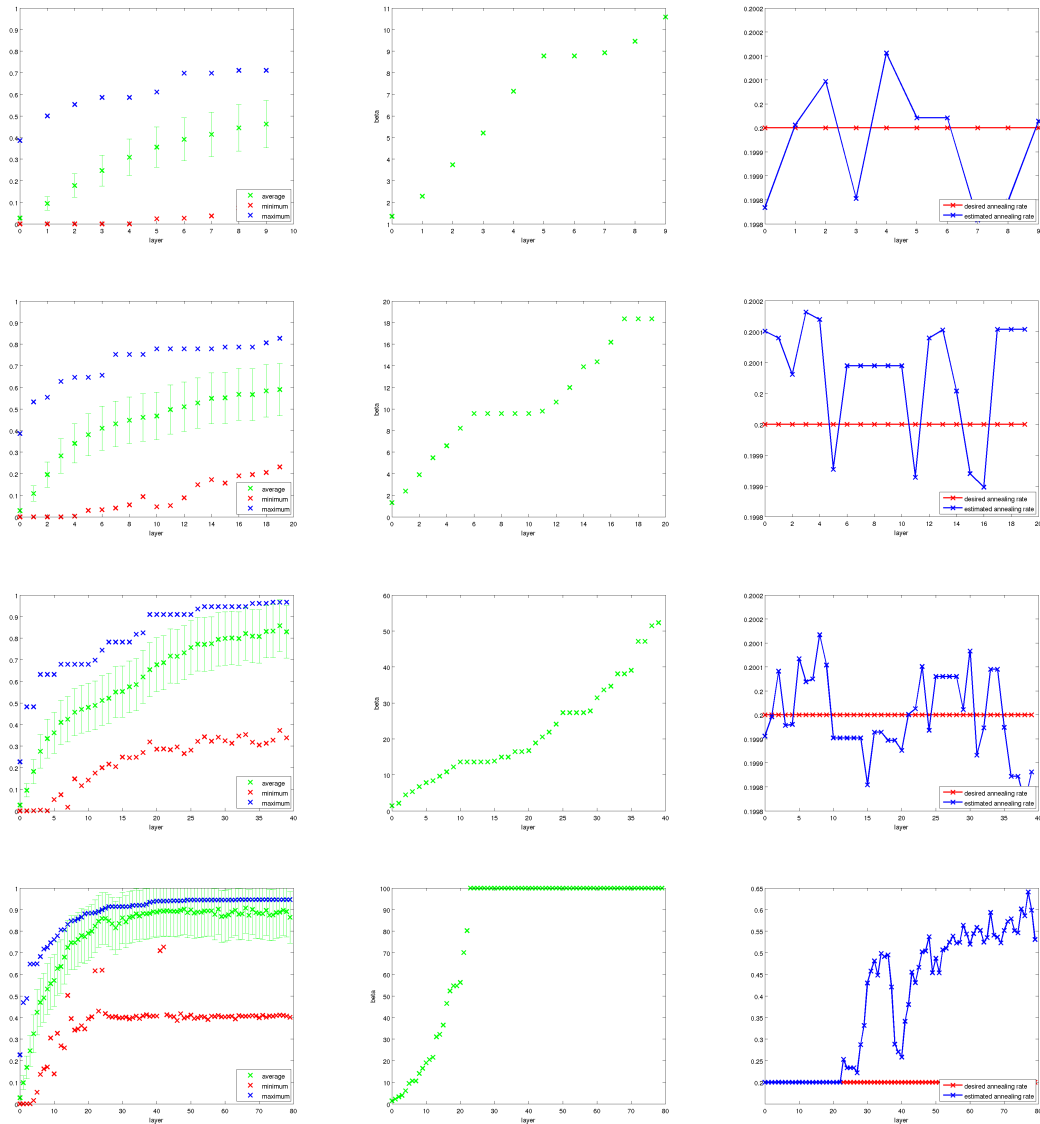


Abbildung 5.7: Gewichtung, Gewichtungsexponent und Überlebensrate für  $(p = 1000, l = 20)$ ,  $(p = 500, l = 20)$ ,  $(p = 250, l = 40)$  sowie  $(p = 125, l = 80)$ , erwartete Überlebensrate= 0.25.

Der Versuch wurde unter Verwendung eines Modells mit 39 Freiheitsgraden ohne Unterteilung des Suchraums durchgeführt. Die gewünschte Überlebensrate betrug 20%. Die initiale Partikelstreuung wurde so gewählt, dass nur sehr vages Vorwissen über die Modellausrichtung erforderlich war. Die Partitionierungsschemata mit 10 und 20 Iterationen weisen eine nicht ausreichende Iterationstiefe auf. Gefundene Maxima sind lokale Ausreißer, die Partikelmasse verteilt sich auf deutlich schlechter gewertete Partikel, wie die Darstellung der Standardabweichung (grüner Balken der jeweils linken Grafik) zeigt. Das Optimum des Durchlaufs mit 40 Iterationen liegt bereits knapp am Rand der Standardabweichung der 40. Iteration und ist das

höchste der vier Testserien. Auffällig ist das Verhalten der Gewichtungsexponenten im Testlauf mit 80 Abkühliterationen. Der Verlauf des Graphen ist durch zwei Aspekte zu erklären. Zum einen findet eine Begrenzung der Exponenten statt, da die Potenzierung der auf  $[0, \dots, 1]$  normierten Gewichte mit zu großen Exponenten zu numerischen Instabilitäten führt. Zum anderen ist der vergleichsweise schnelle Anstieg der Exponenten innerhalb der ersten 20 Iterationen auf eine zu geringe Partikelzahl pro Iteration zurückzuführen. Die verwendeten 125 Partikel sind daher für eine hinreichende Abtastung des Suchraumes nicht ausreichend. Die Überlebensrate der Partikel steigt bedingt durch die Limitierung der Exponenten auf bis zu 65% (+45%) an, während der Fehler in allen anderen Experimenten unter einem Prozent bleibt. Ebenfalls anzumerken ist, dass eine weitere Steigerung des Maximums innerhalb der letzten 40 Iterationen nicht stattfindet.

Abschließend lässt sich festhalten, dass die Wahl der Partikelzahl pro Iteration sowie die Gesamtzahl der Iterationen innerhalb problemspezifischer Grenzen erfolgen muss. Die minimale Partikelzahl hängt sowohl von der Suchraumdimensionalität als auch den betrachteten Wertebereichen ab. Die im folgenden Abschnitt präsentierten Ergebnisse wurden teils mit deutlich geringeren Partikelzahlen erreicht, da bei einer kontinuierlichen Poseschätzung die Veränderung zwischen aufeinander folgenden Zeitpunkten stark einzugrenzen ist.

### 5.3 Verfahren zur Messung und Bewertung der Übereinstimmung

Ziel der Pose-Rekonstruktion ist es, eine Modellkonfiguration zu finden, welche die Übereinstimmung zwischen Modell und Beobachtung maximiert. Diese Übereinstimmung kann auf unterschiedliche Weisen gemessen werden. Im Rahmen dieser Arbeit wird die Übereinstimmung mit einer optimalen Modellkonfiguration gemessen, welche durch visuelle Inspektion manuell generiert wurde. Da viele Aufnahmen zusätzlich mit einem VICON<sup>©</sup> Kamerasystem beobachtet wurden, wurde mittels *mkd-Tools* durch inverse Kinematik die sich aus den Markern ergebende Konfiguration einer identischen Starrkörperkette ermittelt. Eine genauere Beschreibung beider Vergleichsmethoden erfolgt in den beiden sich anschließenden Abschnitten. Gemeinsam ist beiden jedoch, dass keine binäre Klassifikation der Resultate in korrekt respektive fehlerhaft rekonstruierte Posen vorgenommen wird, da ein geeigneter objektiver Klassifikator kaum zu ermitteln ist. Auch ließe eine solche Klassifikation keine Rückschlüsse auf die Ursache einer Fehlklassifikation zu. Daher werden im weiteren Verlauf die relativen Fehler der volumetrischen bzw. absoluten Fehler der geometrischen Übereinstimmung angegeben.

### 5.3.1 Visuelle Übereinstimmung als Goldstandard

Die in Anhang A beschriebene, im Rahmen dieser Arbeit entwickelte, Evaluationsumgebung erlaubt eine exakte manuelle Ausrichtung des Modells und lässt dabei jederzeit Rückschlüsse über die erreichte Überlappung zwischen Modell und Volumenrekonstruktion zu. Das Modell kann hierbei frei im 3D-Raum positioniert werden, so dass eine visuelle Inspektion der Übereinstimmung aus allen Blickwinkeln möglich ist.

Die so gewonnenen Modellkonfigurationen werden als Goldstandard für die Bewertung der mittels Vorhersageverfahren errechneten Konfigurationen verwendet. Auf diese Weise wird erreicht, dass die Sollwerte direkt in passender Form vorliegen und nicht erst Markerpositionen und Kalibrierungsdaten errechnet werden müssen.

### 5.3.2 Vergleich von Gelenkpositionen

Aus der zylindrischen Volumenbeschreibung ergeben sich gewisse Ambiguitäten, insbesondere für die äußeren Extremitäten, da z.B. eine Rotation des (Unter-) Armes um seine Längsachse in den Volumendaten nicht zwangsläufig sichtbar wird. Prinzipiell lassen sich aus einer zeitlichen Glättung und etwa der Handstellung Rückschlüsse ziehen. Dennoch ist auch auf diese Weise eine Fehlparametrisierung nicht auszuschließen. Da auch im Falle solcher Mehrdeutigkeiten die volumetrische Übereinstimmung zwischen Modell und Volumendaten nicht beeinträchtigt wird, empfiehlt es sich, statt der Gelenkwinkel, die sich ergebenden Gelenkpositionen in absoluten Weltkoordinaten gegenüber zu stellen. Diese können aus den Parametervektoren wie folgt zurück gerechnet werden. Ein Vergleich mit diesen Daten lässt keine Rückschlüsse darüber zu, in welchem Teil der Prozesskette (Kamerakalibrierung, Volumenrekonstruktion, Modellausrichtung) ein etwaiger Fehler verursacht wurde. Dennoch erscheint der Vergleich sinnvoll, da markerbasierte Systeme sehr exakte Rekonstruktionen der Bewegungsabläufe liefern.

## 5.4 Ergebnisse der Pose-Rekonstruktion

Zur Evaluation der Erkennungsraten für einen kontinuierlichen Zeitraum wurde ein Volumendatensatz ausgewählt, welcher den Bewegungsablauf eines Karatekämpfers zeigt. Die Testsequenz besitzt eine Länge von 1000 Einzelvolumen. Eine Auswertung wurde nach jeweils 50 Volumen vorgenommen. Der Partikelfilter wurde mit 100 Partikeln und 20 Abkühliteration konfiguriert, als Überlebensrate wurden 20% erwartet. ( $\alpha = 0.2$ ). Eine Darstellung der erzielten Partikelgewichtungen ist in Abbildung 5.8 zu sehen. Die modellierte Optimalpose entspricht jeweils einer Gewichtung von eins, die beste Pose eines jeden Messzeitpunktes ist grün hervorgehoben. Die sich anschließende Grafik 5.9 zeigt rekonstruierte Posen ausgewählter Zeitpunkte in chronologischer Reihenfolge.

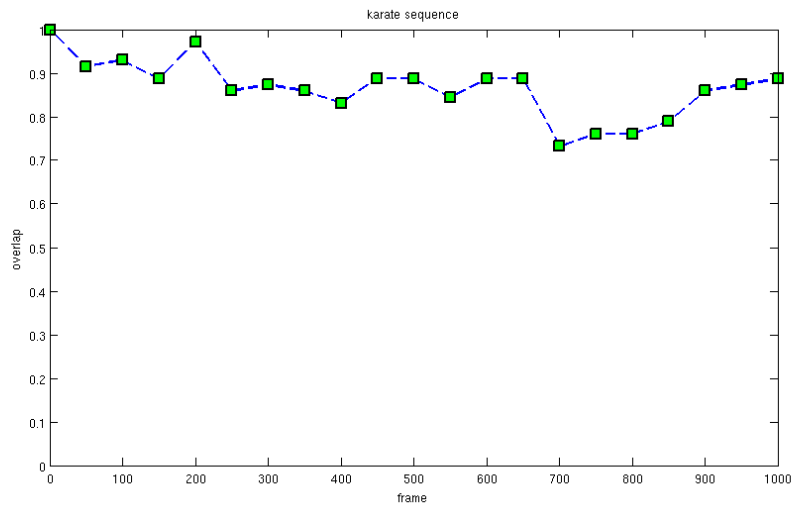


Abbildung 5.8: Detektionsergebnisse Karatekämpfer: Frame 0 – 1000

Die erreichten Überlappungsmaße entsprechen insgesamt einer hohen Übereinstimmung zwischen Modell und Volumendaten, lediglich der deutliche Einbruch der Gewichtung im 700. Volumen zeugt von einer nicht vollständig erfassten Bewegung. Der Tuner geht zu diesem Zeitpunkt schnell in die Hocke, die folgende starke Veränderung der Winkel im Kniegelenk wird nicht ausreichend erfasst. Innerhalb der folgenden Auswertungen findet eine kontinuierliche Korrektur der Fehlausrichtung statt. Da, wie bereits beschrieben, die Rotation eines Zylinders um seine Längsachse nicht sicher rekonstruiert werden kann, kann es zu einem zunehmenden Drift der Armdrehung kommen. Dieser Effekt verstärkt sich zunehmend und führt letztendlich zu einer massiven Fehlstellung des Schultergelenks. Ein weiteres Problem kann auftreten, wenn die Überlappung von Starrkörpern mit verhältnismäßig geringem Volumen zu keiner ausreichenden Abwertung der Pose führt. Kreuzen sich die Beine im Kniebereich, so ist die auftretende Überlappung schwer zu beseitigen, wird das Bein zur Seite bewegt, verschiebt sich lediglich der Schnittpunkt vom Knie in Richtung Unter- oder Oberschenkel. In Abbildung 5.10 ist eine von diesem Problem betroffene Konfiguration dargestellt.



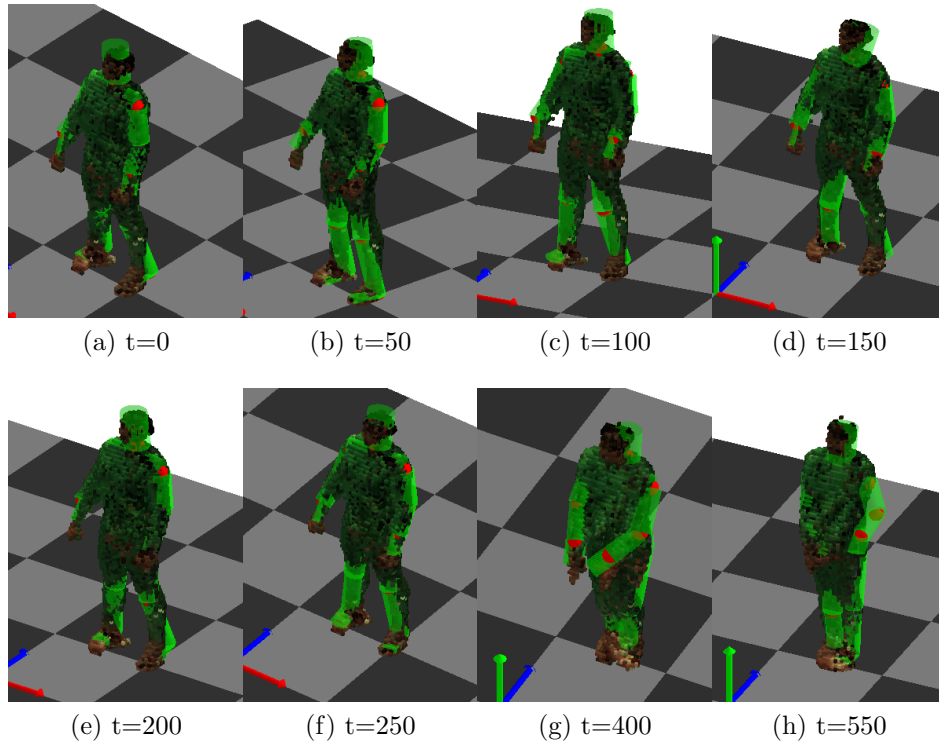


Abbildung 5.9: Chronologische Darstellung der bestimmten Modell-Konfigurationen.

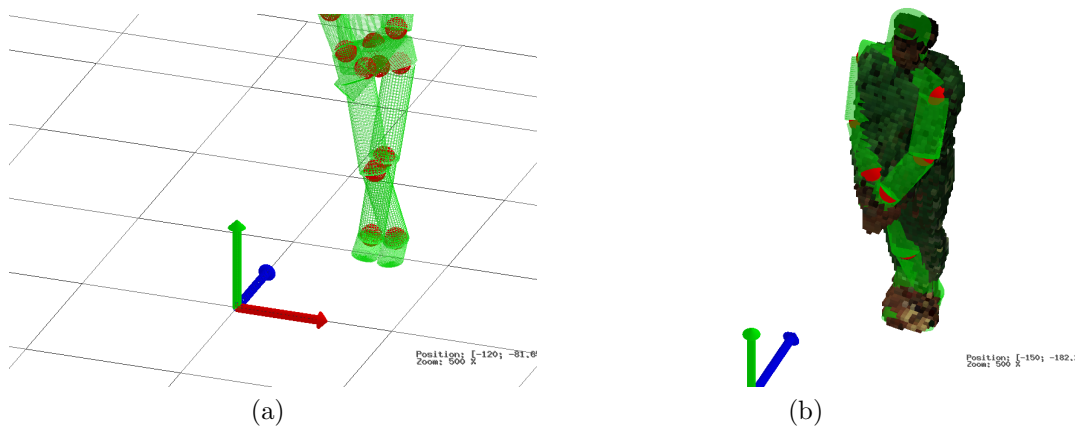


Abbildung 5.10: Frame 450: Fehluordnung und Überlappung der Beine.

Zusätzlich zu den volumetrischen Abständen wurde die Übereinstimmungen mit Markerdaten untersucht. Sowohl die Gelenkkordinaten aus der markerbasierten Rekonstruktion, als auch die durch den Zustandsschätzer ermittelten Koordinaten wurde hierzu errechnet und miteinander verglichen.

In Abbildung 5.11 sind die Ergebnisse eines Zeitpunktes stellvertretend dargestellt. Zum Vergleich ist in a) die volumetrische Übereinstimmung des Modells dargestellt. Eine Visualisierung der Gelenkpositionen beider Rekonstruktionen ist in Abbildung b)

zu sehen. Die blau eingefärbten Gelenkpunkte entsprechen der Modellkonfiguration aus der Volumenrekonstruktion.

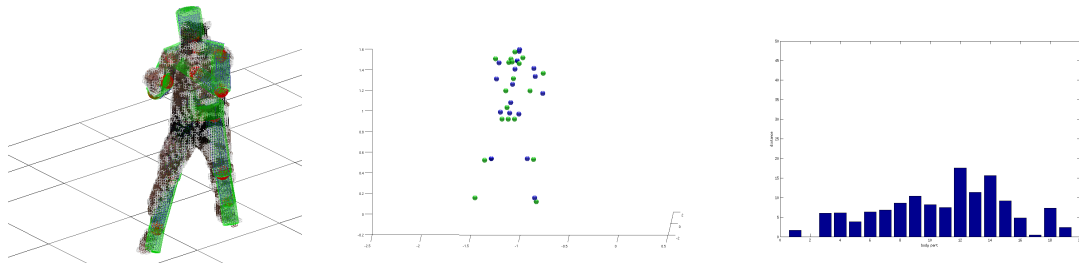


Abbildung 5.11: Ergebnisse des markerbasierten Vergleiches, Visualisierung der Gelenkpunkte (links) und Abweichungen pro Gelenkpunkt in Zentimetern (rechts).

Der unter c) aufgetragene Fehler bezieht sich auf die einzelnen Gelenkpositionen und ist in Zentimetern angegeben. Die durchschnittliche Abweichung beträgt 6.7cm und entspricht damit bei der gewählten Modellauflösung in etwa der Breite von drei Voxel. Im Rumpf und Beinbereich (Starrkörper 1- 11) ist der Fehler vergleichsweise niedrig. Die Ausrichtung des linken Armes bereitet die größten Probleme, da hier die Volumenrekonstruktion besonders starke Artefakte aufweist. Folglich fällt hier der Fehler hier besonders stark aus. Die Abweichung im linken Handgelenk beträgt etwa 12 Zentimeter.

## 5.5 Experimente zur Modellinitialisierung

Das Initialisierungsproblem der Pose-Rekonstruktion besteht in der Ermittlung geeigneter Startwerte, ohne dass zeitliches Vorwissen verfügbar ist. Häufig wird dieses Problem getrennt von der kontinuierlichen Verfolgung der Pose behandelt oder die Ausrichtung des Modells zum ersten Zeitpunkt manuell vorgenommen. Nach einer Erhöhung der Partikelzahl und Aufhebung der Winkelbeschränkungen konnte mit dem vorgestellten Filter die Position aus einer stark abweichenden Initialisierung ermittelt werden. Die Initialpose sowie die geschätzte Endposition sind in Abbildung 5.12 zu sehen.

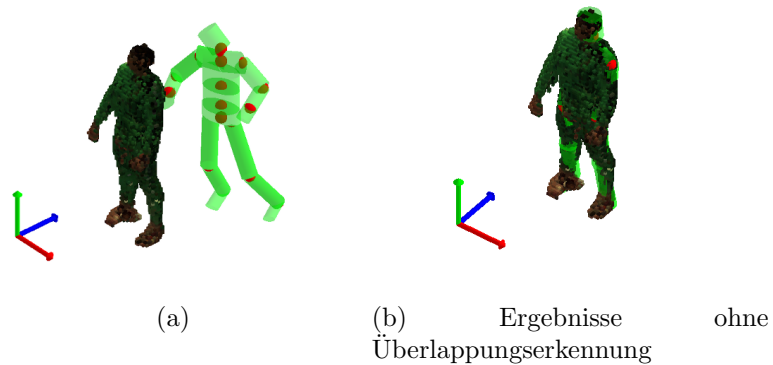


Abbildung 5.12: Erfolgreiche Modellausrichtung bei schlechter Initialpose

# Kapitel 6

## Schlussbetrachtung

Im folgenden Kapitel werden die Ergebnisse dieser Arbeit zusammengefasst und die entscheidenden Erkenntnisse herausgearbeitet. Ein Ausblick auf mögliche, ausstehende Erweiterungen und Verbesserungen des Verfahrens findet sich im Anschluss.

### 6.1 Fazit

Der Annealed Particle Filter ist ein stochastischer Zustandsschätzer in Kombination mit einem heuristischen Optimierungsverfahren. Seine Tauglichkeit zur Pose-Schätzung in hochdimensionalen Suchräumen konnte im Rahmen dieser Arbeit bestätigt werden. Die ermittelten und im vorherigen Kapitel präsentierten Ergebnisse veranschaulichen den Einfluss alternativer Filterkomponenten sowie der entscheidenden Parameter. Durch die vollständige Umsetzung des Verfahrens im dreidimensionalen Raum und den Verzicht auf eine den Informationsgehalt reduzierende Projektion konnte die Robustheit des Verfahrens deutlich gesteigert werden.

Die Fusion der Kameradaten zu einer volumetrischen Beschreibung der Szenen wird vor einem Modellvergleich ausgeführt. Dieser anschließende Vergleich zwischen Szenebeschreibung und Modell basiert auf einer Schnittmengenbestimmung der beteiligten Voxelräume.

Die Kombination dieser beiden Ansätze ermöglicht es, Mehrdeutigkeiten in der Zuordnung von Modell- zu Szenekoordinaten zu vermeiden und liefert daher exaktere Abstandsmaße.

Die Integration von *a priori* Modellwissen wurde durch Modellierung von anthropometrisch korrekten Gelenkfreiheitsgraden und Anschlägen vorgenommen. Eine Veränderung von Gelenkstellungen zwischen zwei Beobachtungszeitpunkten unterliegt weiterhin Grenzen, welche durch die vergangene Zeit bestimmt sind. Basierend auf diesen Angaben kann eine Suchraumeingrenzung vorgenommen werden, so dass nur ein geringer Bereich des eigentlichen Modellraumes untersucht werden muss. Trotz der hohen Dimensionalität des verwendeten Modells (39 Freiheitsgrade) ist eine Pose-Bestimmung mit moderaten Partikelzahlen möglich. Der vorgestellte Ansatz zur Par-

titionierung des Suchraums erwies sich in der bisherigen Form der Umsetzung als nicht ausreichend tragfähig, da die Ergebnisse unter einer fehlerhaften Parameterschätzung in übergeordneten Partitionen leiden. In [BB09] wird ein Vorgehen präsentiert, welches verschiedene Partitionierungsschemata zeitgleich gegeneinander antreten lässt und die Ergebnisse später miteinander fusioniert. Auf diesem Wege ließe sich das Risiko minimieren, jedoch stiege die Partikelzahl durch die parallel Betrachtung mehrerer Hypothesen linear an. Wurde der Partikeltransfer zwischen den Abkühlphasen mit einer ausreichend hohen Überlebensrate durchgeführt, so konnte bei simultaner Optimierung aller Freiheitsgrade die Pose erfolgreich geschätzt werden. Die Evaluation der Simulierten Abkühlung zeigt, dass die Gewichtung der Partikel im Lauf der Abkühlphasen deutlich verbessert werden kann. Ein konventioneller Partikelfilter würde Ergebnisse liefern, welche jenen nach der ersten Abkühlphase entsprechen. Selbst bei einer flachen Partitionierung mit 1000 Partikeln pro Iteration wurden zu diesem Zeitpunkt noch keine als hinreichend exakt einzustufenden Konfigurationen gefunden. Aus den Messreihen lässt sich weiterhin ableiten, dass eine sinnvolle Wahl der Zahl der Abkühlvorgänge und der sich daraus ergebenden Partikelzahlen pro Iteration die Genauigkeit der Ergebnisse entscheidend beeinflusst. Der Übergang der einzelnen Abkühlphasen sollte mit einer kovarianzbasierten Partikelstreuung erfolgen, bereits gefundene Optima sollten zusätzlich explizit übernommen werden.

## 6.2 Ausblick

Die bisherigen Ergebnisse unterstreichen die Tragfähigkeit des gewählten Ansatzes. Weitere Verbesserungen und Anpassungen sind innerhalb der bestehenden Entwicklungsumgebung mit verhältnismäßig wenig Aufwand umzusetzen. So ist etwa eine weitere Verfeinerung der Gewichtungsfunktion denkbar, welche den Abstand zwischen der Volumendarstellung des Modells und der Szene bestimmt. Belegen mehrere Starrkörper den selben Raum, könnte dies mit einer Abwertung der Partikelgewichtung sanktioniert werden. Die aktuelle Umsetzung basiert auf einer uniformen Gewichtung aller Übereinstimmungen. Eine gezielte Erhöhung bzw. Abschwächung der Bewertung bestimmter Starrkörperabschnitte könnte verwendet werden um eine exaktere Ausrichtung im Bereich der Gelenkpunkte zu erreichen.

Zusätzlich zu einer Beschreibung der Starrkörpersegmente durch geometrische Primitive könnte eine Segmentierung des Beobachtungsraumes eingesetzt werden, um passendere Voxeldarstellungen zu finden. Hierzu würde das Modell wie gehabt mit einer einfachen Starrkörperrepräsentation initial ausgerichtet. Ist eine passende Pose gefunden, wird die Volumenrekonstruktion im Bereich der Gelenkpunkte aufgetrennt und die Voxel der Rekonstruktion ersetzen die bisherige Starrkörperrepräsentation. In den folgenden Zeitschritten würde somit der Modellvergleich auf Basis einer deutlich genaueren Volumenbeschreibung erfolgen.

Die komponentenweise arbeitende Gewichtsfunktion, welche die Übereinstimmung jedes Starrkörpers mit der Volumenrekonstruktion separat ermittelt, könnte genutzt werden, um mittels Rücksetzverfahren (Backtracking) Hypothesen partiell auszuwerten und gegebenenfalls eine Verzweigung vorzunehmen oder diese zu verwerfen: Erfolgt etwa eine Fehlpositionierung des Rumpfes, braucht keine Auswertung der Gewichtsfunktionen für die Extremitäten mehr vorgenommen werden. Wurden jedoch Teile der Starrkörperkette bereits korrekt ausgerichtet, so wird der Zustandsvektor dupliziert und die noch unbekannt Parameter werden durch Rauschterme verändert. Dieser Ansatz würde der Baumtopologie des Modells gerecht und bietet sich daher als Alternative zur Suchraumpartitionierung an.

Die momentan verwendete Rotationsdarstellung durch Eulerwinkel wurde aus Gründen der Kompatibilität zu bestehenden Umsetzungen gewählt. Die sich hieraus ergebende kardanische Aufhängung unterliegt einer Blockade, wenn zwei Drehachsen in die gleiche Ebene rotiert werden (gimbal lock). Dies führt zum Verlust eines Freiheitsgades und kann die korrekte Modellausrichtung verhindern. Auch wenn dieser Effekt durch die gewählte Rotationsreihenfolge meist vermieden werden kann, könnten die Vorteile einer Winkeldarstellung durch Quaternionen untersucht werden, insbesondere da diese die Interpolation von Modellposen ermöglichen würden.



# Anhang A

## Implementierung

Das betrachtete Verfahren zur Poserekonstruktion besteht aus einer Vielzahl von Einzelkomponenten und Stellparametern. Die korrekte Wahl dieser Bestandteile erfordert eine ausführliche Evaluation der sich ergebenden Effekte und eine Gegenüberstellung zu alternativen Umsetzungen. Eine Recherche nach geeigneten, frei verfügbaren Testplattformen lieferte keine zufriedenstellenden Resultate, so dass im Rahmen dieser Arbeit eine eigene Umgebung entwickelt wurde. Mit *wombat*( **w**orking environment for **m**odel **b**ased **t**racking ) entstand eine Reihe von Werkzeugen, welche die Erstellung und Verwendung von Starrkörper-Modellen ermöglichen.

### A.1 Verwendete Vorarbeiten

In Form der *mkdTools* von Günther Stelzner [Ste09] existiert bereits ein in Matlab umgesetztes Starrkörpermodell, welches am Institut für Anthropomatik am KIT Karlsruhe zu Evaluationszwecken eingesetzt wird. Auch wenn geringe Laufzeiten nicht im Fokus dieser Arbeit lagen, sind die Rechenzeiten dieser Umsetzung für den Einsatz in einem Partikelfilter-Framework ungeeignet. Um die Vorteile dieser Werkzeuge (Modelleinpassung, Evaluation von Markerdaten, inverse Kinematik etc.) nutzen zu können, wurde ein Starrkörpermodell implementiert, welches bezüglich seiner Vorwärtskinematik kompatibel zu den vorgestellten Werkzeugen ist. Innerhalb der im Rahmen dieser Arbeit entstandenen Testumgebung können beide Starrkörpermodelle verwendet und jederzeit durch einander ersetzt werden. Für die native C++-Variante steht weiterhin ein Editor zur Verfügung, welcher die Modellierung weiterer, angepasster Menschmodelle erlaubt.

### A.2 Architektur

Die in Kapitel 5 präsentierten Ergebnisse erfordern ein modulares Framework, welches den schnellen Tausch von Einzelkomponenten ermöglicht. Diesem Anspruch wird mit



einem modular aufgebauten Zustandsschätzer und einer austauschbaren Starrkörper-Physik Rechnung getragen.

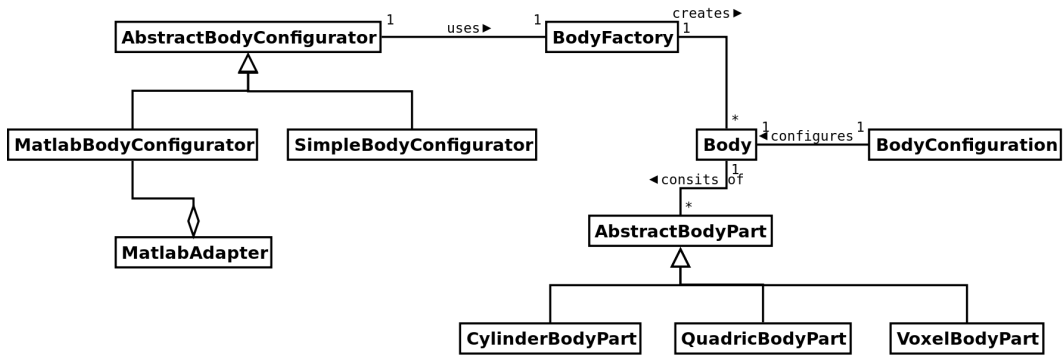


Abbildung A.1: UML-Klassendiagramm der Starrkörper-Komponenten

Abbildung A.1 zeigt den Aufbau der Starrkörperkette. Ein Körper besteht aus einer Menge potentiell verschiedenartiger Körperteile. Mittels eines Zustandsvektors (Body-Configuration) wird die gewünschte Modellausrichtung beschrieben, welche von zwei unterschiedlichen Konfiguratoren erzeugt werden kann.

## A.3 Arbeitsumgebung und Werkzeuge

### A.3.1 Die Arbeitsumgebung

Die *wombat*-Arbeitsumgebung erlaubt ein projektbasiertes Arbeiten mit nahezu beliebigen Kombinationen aus Starrkörperketten, Filterkonfigurationen und Voxelräumen. Sämtliche in dieser Arbeit beschriebenen Filter- und Modellparameter können manipuliert und in ihren Auswirkungen getestet werden. Das Programm (siehe Abbildung A.3) dient ebenfalls als Analysewerkzeug: Zum einen können hier die optimalen Modellkonfigurationen händisch erzeugt werden, zum anderen ist ein Export aller Systemzustände in Matlab-kompatibler Form möglich. Für jedes Projekt ist der Typ des Modell-Adapters festzulegen (Matlab oder nativ), sowie die zu analysierende Volumen-Datei. Alle weiteren Parameter können zur Laufzeit angepasst werden. Neben den volumen- und modellspezifischen Parametern sind dies insbesondere die Filterparameter, welche sich in 4 Kategorien aufteilen.

1. **Rasterisierung:** Hier kann die gewünschte Voxelraumdiskretisierung eingestellt werden. Eine grobe Einstellung reduziert die Berechnungskomplexität, jedoch auch die geometrische Genauigkeit des Verfahrens.
2. **Gewichtsfunktion:** Die Gewichtung der Voxelraumüberschneidungen kann unter Berücksichtigung von vorherigen Schnittbildungen erfolgen.
3. **Partitionierung:** Der Suchraum kann in Subräume unterteilt werden, indem einzelne Starrkörper Partitionen zugewiesen werden. Diese Zuweisung wird bei der Auswertung der Gewichtsfunktionen als auch der Partikelstreuung berücksichtigt.
4. **Simulierte Abkühlung:** Die unter 4.4 beschriebenen Parameter des Annealed Partikelfilters können hier eingestellt werden.

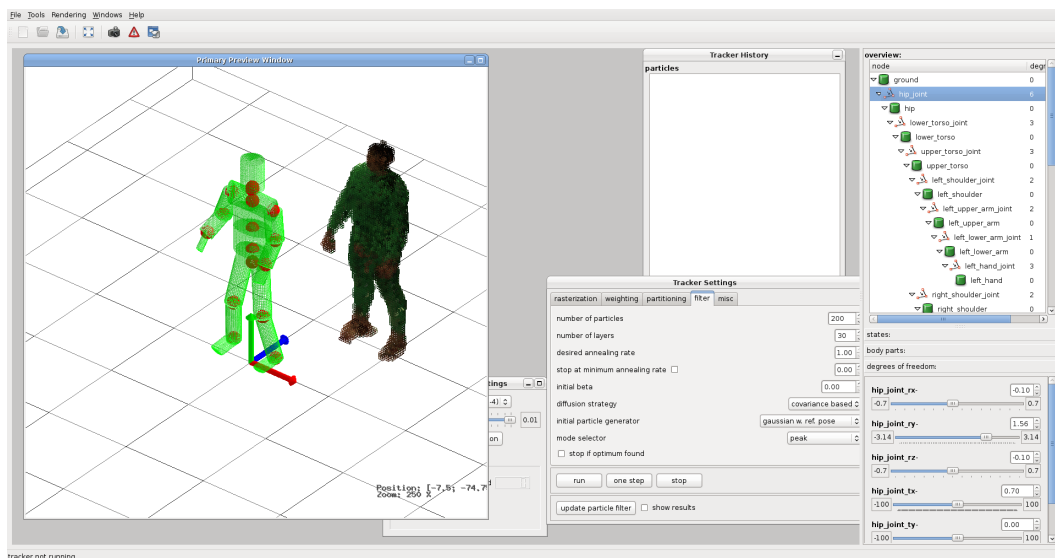


Abbildung A.2: Projektansicht der Wombat-Arbeitsumgebung

## A.3.2 Der Modell-Editor

Der Modelleditor erlaubt das intuitive Erstellen von Starrkörperketten. Hierzu kann das Modell als Szenegraph definiert und direkt visualisiert und getestet werden. Ein Zugriff auf alle Modellparameter, sowohl von Starrkörpern als auch Gelenkkoordinatensystemen, ist von hier aus möglich. So können etwa die Starrkörperformen und Größen angepasst werden, oder einzelne Freiheitsgrade eines Gelenks gesperrt werden. Die Modellbeschreibung kann als XML-Beschreibung gespeichert und zu einem späteren Zeitpunkt innerhalb der Arbeitsumgebung verwendet oder weiter bearbeitet werden.

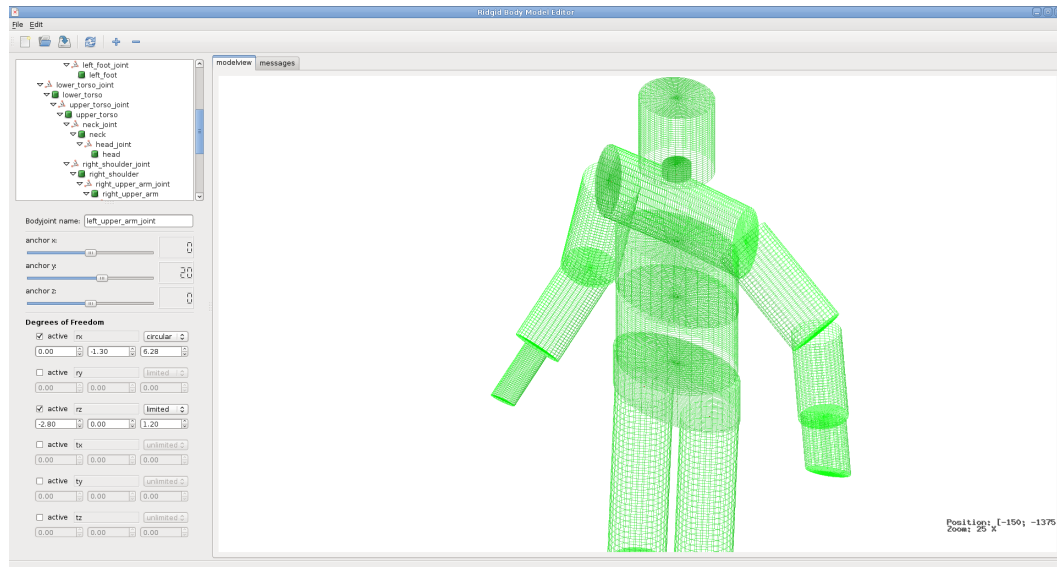


Abbildung A.3: Projektansicht der Wombat-Arbeitsumgebung

# Anhang B

## Mathematische Grundlagen

### B.1 Allgemein

#### B.1.1 Boltzmann-Statistik

Die Boltzmann-Statistik beschreibt die Wahrscheinlichkeit das ein System von Elementen sich in einem thermodynamischen Zustand befindet.

Die Zustände  $z$  sind mit  $j = 1, 2, \dots, N$  nummeriert und besitzen jeweils die Energie  $E_j$ . Die zugehörige Wahrscheinlichkeit, welche angibt ob man sich im Zustand  $z_j$  befindet, ist gegeben durch:

$$p_j = \frac{1}{Z} e^{-\beta E_j} \quad (\text{B.1})$$

$Z$ , die Zustandsumme wird hierbei als Normierungskoeffizient verwendet. Die Unbekannte  $\beta$  ergibt sich wie folgt aus der Boltzmann-Konstanten  $k_B$ :

$$\beta = \frac{1}{k_B T} \quad (\text{B.2})$$

#### B.1.2 Markov-Kette

Eine Markov-Kette ist ein stochastischer Prozess, bei dem ein begrenztes Vorwissen identische Aussagen über die Zukunft zulässt, wie die vollständige Historie. Für den Fall Markov-Ketten erster Ordnung bedeutet dies, das lediglich der aktuelle Zustand benötigt wird um Prognosen über den Folgezustand zu treffen.

### B.1.3 Metropolis-Algorithmus

Der Metropolis-Algorithmus erzeugt eine Markov-Kette entsprechend der Boltzmann-Verteilung. Sei  $\vec{x}$  ein mehrdimensionaler Zustand und  $\vec{x}_i$  der Zustand nach  $i$  Iterationen.

1. Neuer Zustand:  $\vec{y} = \vec{x}_i + (2r - 1) \times \delta$ , wobei  $r$  eine Zufallszahl  $\{0,1\}$  sowie  $\delta$  ein fest gewählter Abstand.
2. Als nächstes wird die Energiedifferenz zwischen Zustand und Folgezustand berechnet (Gütefunktion).  $\Delta E = E(\vec{y}) - E(\vec{x}_i)$ .  
Die Wahrscheinlichkeit

$$p_a = \min(1, e^{-\frac{\Delta E}{kT}}) \quad (\text{B.3})$$

gibt an, ob die neue Konfiguration akzeptiert wird. ( $T$  gibt die Temperatur an,  $k$  entspricht der Boltzmann-Konstanten)

- ist  $\Delta E \leq 0$  so wird  $\vec{y}$  als neuer Systemzustand akzeptiert:  $\vec{x}_{i+1} = \vec{y}$
- ist  $\Delta E \geq 0$  so wird  $\vec{y}$  mit der Wahrscheinlichkeit  $p_a$  akzeptiert. Eine aktuell generierter Zufallswert  $r \in \{0,1\}$  wird mit  $p_a$  verglichen und  $\vec{y}$  wird akzeptiert, falls  $r < p_a$  gilt.
- wird der Zustand nicht akzeptiert, so gilt  $\vec{x}_{i+1} = \vec{x}_i$

# Anhang C

## Glossar

**Initialisierungsproblem** Das Initialisierungsproblem besteht im Kontext der Poserekonstruktion in der Bestimmung einer Ausgangskonfiguration des Modells. Zum Startzeitpunkt der Rekonstruktion ist kein Vorwissen vorhanden, welches eine grobe Modellausrichtung ermöglicht.

**Kinetik** Die Kinematik beschreibt die Veränderung von Bewegungsgrößen unter dem Einfluss der auf einen Körper wirkenden Kräfte.

**Poseschätzung** Unter Poseschätzung wird jener Vorgang verstanden, welcher basierend auf verfügbaren Beobachtungen eine Schätze der Orientierung und Lage eines Objektes vornimmt.

**Prozessrauschen** Beobachtungsprozesse unterliegen häufig prozessspezifischen Meßfehlern, deren Ursache und Auswirkung nicht exakt modellierbar ist. Dieser Fehler wird als Prozessrauschen bezeichnet.

**Rückwärtskinematik** Die Rückwärtskinematik dient der Ermittlung aller Modellparameter basierend auf einer Modellpose. Gegeben ist also die Position aller Modellbestandteile in Absolutkoordinaten, gesucht der Zustandsvektor der diese Konfiguration erzeugt.

**Stickmodell** Ein Stickmodell ist Starrkörpermodell, welches Körperteile durch seine Hauptachsen darstellt. Es besitzt keine volumetrische Ausdehnung.

**Szenegraph** Ein Szenegraph ergibt sich durch die Anordnung von Grundelementen, welche zur Beschreibung einer Szenen notwendig sind. Dies sind im allgemeinsten Fall Geometrie- sowie Transformationsknoten.

**Tracking** Tracking lässt sich allgemein als das Verfolgen von Trajektorien über die Zeit im Bild- bzw. Beobachtungsraum beschreiben.

**Vorwärtskinematik** Die Vorwärtskinematik dient der Ausrichtung eines Modells anhand einer gegebenen Beschreibung. Diese kann etwa in Form eines Zustandsvektors vorliegen, welcher alle benötigten Translationen und Rotationen enthält.

**Volumendaten** Volumendaten sind Datensätze welche eine räumliche Beschreibung einer Beobachtung liefern.

**Voxel** Volumetrische Entsprechung eines Pixels, während bei einem Pixel häufig keine flächenmäßige Ausdehnung angenommen wird, ergibt sich im Rahmen dieser Arbeit das Volumen eines Voxels aus der gewählten → Voxelraumdiskretisierung

**Voxelraumdiskretisierung** Wie auch bei der Bildentstehung ist eine Diskretisierung notwendig, da der Raum nur an einer endlichen Anzahl von Stellen abgetastet werden kann.

# Weiterführende Literatur zu wissenschaftlichen Ausarbeitungen

- [BB09] BANDOUCHE, Jan ; BEETZ, Michael: Tracking Humans Interacting with the Environment Using Efficient Hierarchical Sampling and Layered Observation Models. In: *IEEE Int. Workshop on Human-Computer Interaction (HCI). In conjunction with ICCV2009*, 2009
- [BEB08] BANDOUCHE, Jan ; ENGSTLER, Florian ; BEETZ, Michael: Evaluation of Hierarchical Sampling Strategies in 3D Human Pose Estimation / Technische Universität München, Munich, Germany. 2008. – Forschungsbericht
- [CH04] CAILLETTE, Fabrice ; HOWARD, Toby: Real-Time Markerless Human Body Tracking Using Colored Voxels and 3-D Blobs / Advanced Interfaces Group, University of Manchester. 2004. – Forschungsbericht
- [Das03] DASZKOWSKI, Alexandra: *Das Körperbild bei Frauen und Männern*. Tectum-Verlag, 2003. – ISBN 3828884709
- [DBR09] DEUTSCHER, Jonathan ; BLAKE, Andrew ; REID, Ian: Articulated Body Motion Capture by Annealed Particle Filtering / University of Oxford, Microsoft Research. 2009. – Forschungsbericht
- [DNBB99] DEUTSCHER, J. ; NORTH, B. ; BASCLE, B. ; BLAKE, A.: Tracking Through Singularities and Discontinuities By Random Sampling. In: *In IEEE international conference on computer vision*, 1999, S. 1144–1149
- [DR03] DEUTSCHER, Jonathan ; REID, Ian: Articulated Body Motion Capture by Stochastic Search / University of Oxford. 2003. – Forschungsbericht
- [EDD03] ELGAMMAL, Ahmed M. ; DURAISWAMI, Ramani ; DAVIS, Larry S.: Efficient Kernel Density Estimation Using the Fast Gauss Transform with Applications to Color Modeling and Tracking. In: *IEEE Trans. Pattern Anal. Mach. Intell.* 25 (2003), Nr. 11, S. 1499–1504
- [FDW09] FELDMANN, Tobias ; DIESELBERG, Lars ; WÖRNER, Annika: Adaptive Foreground/Background Segmentation using Multiview Silhouette Fusion. In: DENZLER, Joachim (Hrsg.) ; NOTNI, Gunther (Hrsg.) ; SÜSSE, Herbert (Hrsg.): *Pattern Recognition, 31st DAGM Symposium, Jena, Germa-*



ny, September 9-11, 2009. *Proceedings* Bd. 5748, Springer, 9 2009 (Lecture Notes in Computer Science)

- [Gav99] GAVRILA, D. M.: The Visual Analysis of Human Movement: A Survey. In: *Computer Vision and Image Understanding* 73 (1999), S. 82–98
- [GD96] GAVRILA, D. M. ; DAVIS, L.S.: 3-D Model-based Tracking of Humans in Action: a Multi-View Approach. In: *In IEEE computer vision and pattern recognition*, 1996
- [Gra01] GRANUM, Thomas B. Moeslund A.: A Survey of Computer Vision-Based Human Motion Capture. In: *Computer Vision and Image Understanding* 81 (2001), S. 231–268. <http://dx.doi.org/10.1006/cviu.2000.0897>. – DOI 10.1006/cviu.2000.0897
- [HSS<sup>+</sup>09] HASLER, N. ; STOLL, C. ; SUNKEL, M. ; ROSENHAHN, B. ; SEIDEL, H.-P.: A Statistical Model of Human Pose and Body Shape. In: DUTR’E, P. (Hrsg.) ; STAMMINGER, M. (Hrsg.): *Computer Graphics Forum (Proc. Eurographics 2008)* Bd. 2. Munich, Germany, März 2009
- [IB98] ISARD, Michael ; BLAKE, Andrew: Condensation - conditional density propagation for visual tracking. In: *International Journal for Computer Vision* 29 (1998), Nr. 1, 5-28. <ftp://ftp.robots.ox.ac.uk/pub/ox.papers/VisualDynamics/ijcv98.ps.gz>. ISBN 0920–5691
- [JU04] JULIER, Simon J. (Hrsg.) ; UHLMANN, Jeffrey K. (Hrsg.): *Unscented Filtering and Nonlinear Estimation*. IEEE, 2004
- [Kal60] KALMAN, R. E.: A new approach to linear filtering and prediction problems. (1960)
- [KGV07] KIRKPATRICK, S. ; GELATT, C. D. ; VECCHI, M. P.: Optimization by Simulated Annealing. In: *Science, New Series* 220 (2007), S. 671–680
- [KVD07] KNOOP, Steffen ; VACEK, Stefan ; DILLMANN, Rüdiger: A Human Body Model for Articulated 3D Pose Tracking / University of Karlsruhe, Germany. 2007. – Forschungsbericht
- [LC03] LEE, Mun W. ; COHEN, Isaac: Human Body Tracking with Auxiliary Measurements, 2003
- [LC06] LEE, Mun W. ; COHEN, Isaac: A Model-Based Approach for Estimating Human 3D Poses in Static Images. In: *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE* 28 (2006), S. 905–917
- [MI00] MACCORMICK, John ; ISARD, Michael: Partitioned sampling, articulated objects, and interface-quality hand tracking / University of Oxford AND Compaq System Research Centre. 2000. – Forschungsbericht

- [SG99] STAUFFER, Chris ; GRIMSON, W.: Adaptive background mixture models for real-time tracking. In: *International Conference on Computer Vision and Pattern Recognition* Bd. 2, 1999, S. 252–259
- [SMC01] STENGER, B. ; MENDONCA, P. R. S. ; CIPOLLA, R.: Model-Based Hand Tracking Using an Unscented Kalman Filter / University of Cambridge. 2001. – Forschungsbericht
- [ST03] SMINCHISESCU, Cristian ; TRIGGS, Bill: Estimating Articulated Human Motion With Covariance Scaled Sampling. In: *International Journal of Robotics Research* 22 (2003), S. 371–393
- [Ste09] STELZNER, Günther: Zur Modellierung und Simulation biomechanischer Mehrkörpersysteme. In: *Schriftenreihe des Instituts für Technische Mechanik* 11 (2009)
- [UF03] URTASUN, Raquel ; FUA, Pascal: 3D Human Body Tracking using Deterministic Temporal Motion Models / Computer Vision Laboratory Lausanne, Switzerland. 2003. – Forschungsbericht
- [WB06] WELCH, Greg ; BISHOP, Gary: An Introduction to the Kalman Filter / University of North Carolina at Chapel Hill. 2006. – Forschungsbericht
- [WR05] WANG, Ping ; REHG, James M.: A Modular Approach to the Analysis and Evaluation of Particle Filters for Figure Tracking / Georgia Institute of Technology. 2005. – Forschungsbericht
- [YJS06] YILMAZ, Alper ; JAVED, Omar ; SHAH, Mubarak: Object Tracking: A Survey. In: *ACM Computing Surveys* 38 (2006). <http://dx.doi.org/10.1145/1177352.1177355>. – DOI 10.1145/1177352.1177355
- [ZH04] ZHOU, Huiyu ; HU, Huosheng: A Survey - Human Movement Tracking and Stroke Rehabilitation / University of Essex. 2004. – Forschungsbericht. – ISSN 1744 - 8050

