

**Lehrstuhl für Ergonomie  
der Technischen Universität München  
Univ.-Prof. Dr.rer.nat. H. Bubb**

**Videobasierte Messung menschlicher  
Bewegungen konform zum Menschmodell RAMSIS**

**Thomas Seitz**

Vollständiger Abdruck der von der Fakultät für Maschinenwesen der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr.rer.nat.)

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr.-Ing. Klaus Bender

Prüfer der Dissertation:

1. Univ.-Prof. Dr.rer.nat. Heiner Bubb

2. Univ.-Prof. Dr.rer.nat. Bernd Radig

Die Dissertation wurde am 04.06.2003 bei der Technischen Universität München eingereicht und durch die Fakultät für Maschinenwesen am 17.11.2003 angenommen.



## Danksagung

An dieser Stelle möchte ich nicht versäumen mich bei dem einen oder anderen zu bedanken, der mich bei der Durchführung dieser Arbeit unterstützte und so seinen Anteil am Gelingen trägt.

Mein ganz besonderer Dank gilt Herrn Prof. Bubb, Ordinarius des Lehrstuhls für Ergonomie der TUM, für die Initiierung dieser Arbeit und für die vielen lehrreichen Stunden, interessanten (Fach-)Gespräche und Erlebnisse, sowie die freundliche und motivierende Unterstützung und die vielen Ratschläge.

Mein weiterer Dank gilt Herrn Prof. Radig vom Lehrstuhl Informatik IX Bildverstehen und wissensbasierte Systeme der TUM als meinem Zweitgutachter, der mir die Sicht der Informatik vermitteln konnte. Ferner möchte ich Herrn Prof. Bender vom Lehrstuhl für Informationstechnik im Maschinenwesen, der den Prüfungsvorsitz übernahm, danken.

An dieser Stelle möchte ich mich auch bei Herrn Prof. Krueger von der ETH Zürich bedanken, der mir nicht nur die Möglichkeit gab durch einen Forschungsaufenthalt an seinem Institut eine „Außensicht“ der Münchener Ergonomie zu gewinnen, sondern der auch durch intensive Gespräche meine Arbeit und auch meine weitere Zukunftsplanung vorantrieb.

Weiterhin danke ich allen Kollegen und Mitarbeitern des Lehrstuhls für Ergonomie, insbesondere der „Mittagsclique“, meinen Kollegen an der ETH Zürich, sowie allen meinen Freunden und Bekannten für die tatkräftige Unterstützung, die vielen Anregungen und Fachgespräche.

Garching, im November 2003

Thomas Seitz



## Zusammenfassung

Für die Simulation von Mensch-Maschine-Interaktionen, wie sie für die Produkt- und Produktionsentwicklung nötig sind, werden dreidimensionale Menschmodelle benutzt. Diese zunächst statischen Modelle müssen mit Anthropometrie-, Haltungs- und Bewegungsmodellen ausgestattet werden, so dass menschliches Verhalten simuliert werden kann. Die Vielfalt menschlicher Anthropometrien, Haltungen und Bewegungen macht es hierbei unerlässlich zahlreiche Messungen durchzuführen, um detaillierte Theorien bilden zu können. Doch bis heute stellt die Messung der menschlichen Parameter, besonders der Bewegung, ein Problem dar. Zum einen ist die Kopplung zwischen Mensch und Modell schwierig und zum anderen ist mit heutigen Systemen meist eine zeitaufwändige Präparierung des Probanden mit Markern nötig.

Im Rahmen dieser Arbeit wurde ein Ansatz unter Verwendung des Menschmodells RAMSIS entwickelt, um auf Basis von Videobeobachtungen, Bewegungen dreidimensional zu messen. Ausgehend von biologischen Prinzipien der Bewegungserkennung wurde ein modellbasiertes Verfahren erarbeitet, das ohne eine Präparierung mit Markern Bewegungen erfassen kann. Das System arbeitet zur Bildgewinnung mit Standardvideotechnik. In die digitalisierten Bildsequenzen wird nach einer Kalibrierung ein individuelles und dreidimensionales Modell des Probanden projiziert, wobei am Anfang der Bildsequenz die Haltung des Modells exakt mit der Haltung des Probanden zur Deckung gebracht wird. Damit können Bildbereiche mit Modellbereichen assoziiert werden und mit Hilfe digitaler Bildverarbeitung können die Bilder analysiert werden, um die Bewegung in den Folgebildern zu ermitteln. Hierzu wird zum einen untersucht, wohin sich eine Bildstruktur bewegt hat und zum anderen werden Bildkonturen bzw. -kanten benutzt, um das Menschmodell auszurichten. Menschmodell- und Bildinformation werden wechselseitig in die Interpretation integriert.

Mit Hilfe dieses Ansatzes können konform zum verwendeten Menschmodell Haltungen und Bewegungen ohne eine vorherige Präparierung der Versuchsperson ermittelt werden. Damit können Untersuchungen ohne physische Beeinflussung des Probanden erfolgen.



# Inhalt

<b>DANKSAGUNG.....</b>	<b>III</b>
<b>ZUSAMMENFASSUNG.....</b>	<b>V</b>
<b>INHALT .....</b>	<b>VII</b>
<b>1 EINLEITUNG.....</b>	<b>1</b>
1.1 Allgemeines .....	1
1.2 Stand der Forschung und Technik .....	2
1.2.1 Bewegungsmessung in der Ergonomie.....	2
1.2.2 Bewegungsmessung in der Informatik .....	6
1.2.3 Zielsetzungen der Bewegungsmessung.....	11
1.2.4 Systematik exakter Bewegungsmessverfahren .....	13
1.3 „Ideales“ Bewegungsmesssystem .....	20
<b>2 MENSCHMODELL RAMSIS UND DAS MESSSYSTEM PCMAN .....</b>	<b>23</b>
2.1 Entwicklungsgeschichte ergonomischer Menschmodelle .....	23
2.2 Aufbau des Modells RAMSIS .....	26
2.2.1 Struktur des inneren Modells .....	27
2.2.2 Struktur des äußeren Modells .....	28
2.3 PCMAN – das RAMSIS-Messsystem für Anthropometrie- und Haltungsmessung....	29
2.3.1 Prinzip.....	30
2.3.2 Kalibrierung .....	31
2.3.3 Anthropometrische Modelladaption .....	33
2.3.4 Haltungsanpassung .....	36
2.4 PCMAN als Basis für die Bewegungsmessung.....	38
<b>3 METHODIK .....</b>	<b>42</b>
3.1 Psychologisch-physiologische Mechanismen der Bewegungserkennung.....	42
3.1.1 Neuronale Bildinterpretation .....	42
3.1.2 Interpretation von Bewegung.....	45
3.1.3 Kognitive Verarbeitung .....	49
3.1.4 Zusammenfassung.....	51

3.2	Bildverarbeitung.....	51
3.2.1	Bildvorverarbeitung .....	52
3.2.2	Bewegungsmessung.....	55
3.2.3	Kantenbilder .....	60
3.2.4	Zusammenfassung.....	61
<b>4</b>	<b>REALISIERUNG.....</b>	<b>62</b>
4.1	Zweidimensionale Bewegungsmessung ohne Modell.....	62
4.2	Dreidimensionale Bewegungsmessung mit Modell.....	66
4.2.1	Präpositionierung des Modells .....	66
4.2.2	Gelenkpunktverschiebung detektieren .....	70
4.2.3	Resultierende Haltung im Raum bestimmen.....	72
4.2.4	Integration von Kanteninformation .....	73
4.3	Probleme .....	75
4.3.1	Drift beim Verfolgen unprägnanter Strukturen.....	76
4.3.2	Nicht ausgeprägte Kanten .....	77
4.3.3	Verdeckung .....	78
4.3.4	Rotation .....	78
4.3.5	Genauigkeit .....	80
4.4	Zusammenfassung.....	80
<b>5</b>	<b>AUSBLICK.....</b>	<b>82</b>
<b>6</b>	<b>LITERATUR.....</b>	<b>83</b>
<b>7</b>	<b>ANHANG.....</b>	<b>89</b>

# 1 Einleitung

## 1.1 Allgemeines

Schon lange stellt man sich in der Arbeitswissenschaft die Frage, wie man menschliche Bewegungen oder ganz allgemein, wie man Personen erkennen und ihre Position, ihre Haltung und Bewegung messen und interpretieren kann. Bereits Leonardo da Vinci (1452 – 1519) beobachtete die Bewegungen schaufelnder Arbeiter und untergliederte den Arbeitsverlauf in Teilarbeiten und Teilzeiten [29]. Wie Leonardo da Vinci, so versuchten viele Wissenschaftler im Bereich Arbeitswissenschaft, Methoden der Messung und Interpretation von Bewegungen zu entwickeln.

Die Untersuchung menschlicher Haltungen und Bewegungen erfolgte aus unterschiedlichen Gründen. Früher wurden zahlreiche Untersuchungen in Fabriken durchgeführt, um die Leistungsfähigkeit der Arbeiter zu erforschen. Ferner versuchte man Arbeitsabläufe effizienter zu gestalten, um optimale Produktivität zu erreichen. In dem hier im Vordergrund stehenden Bereich *Ergonomie* dient die Haltungs- und Bewegungsanalyse der Gewinnung von empirischen Daten, um Arbeitsabläufe im Sinne des Arbeitsschutzes zu verbessern oder um allgemeine Modelle physischen menschlichen Verhaltens für Simulationszwecke aufzustellen.

Gerade der letzte Punkt ist eng verknüpft mit der in den 80er und 90er Jahren aufgekommen digitalen Menschmodellierung. Die digitale Menschmodellierung wird meist im Umfeld des CAD<sup>1</sup> eingesetzt bzw. wurde durch Ansprüche an das CAD initiiert. Moderne Produktentwicklung, geprägt vom Begriff *Rapid Prototyping*, spielt sich heutzutage fast ausschließlich im Computer ab. So erscheint es nahe liegend für Produkte, die mit Menschen wechselwirken - wie beispielsweise Computertastaturen oder Bügeleisen aber auch so komplexe Systeme wie das Cockpit von Flugzeugen oder Autos - diese Mensch-Maschine-Interaktion im Computer zu simulieren. Derartige Menschmodelle können anthropometrisch korrekt Personen spezifischer Nationalitäten, Kräfte und Momente bestimmter Gelenke, Ermüdung, Komfort, Sicht und Haltungen jeweils in Abhängigkeit äußerer Randbedingungen, wie beispielsweise der Geometrie eines Arbeitsplatzes, simulieren. Weitverbreitete Menschmodelle sind z.B. JACK, SAFEWORK oder das hier im Vordergrund stehende Menschmodell RAMSIS.

Um jedoch menschliches Verhalten mit Hilfe geeigneter Programme simulieren zu können, müssen zunächst Daten gesammelt werden, deren mathematische Analyse das Ableiten

---

<sup>1</sup> Computer Aided Design

verallgemeinerter Aussagen erlaubt. Dies gilt für einzelne statische Haltungen, wie Sitzen, genauso wie für Bewegungen, wie das Einsteigen ins Auto, das Hinsetzen oder Anschnallen. Gerade die Simulation von Bewegungen ist jedoch bis heute nur unzureichend und sehr rudimentär realisiert. Meist handelt es sich nicht um Simulationen, als vielmehr um das Abspielen von Animationen, also Trickfilmen, die zwar die Bewegung eines Menschen zeigen, sich aber nicht individuell mit Hilfe eines Bewegungsmodells an die Umwelt anpassen können. Eine Bewegungssimulation hingegen nutzt Modellinformationen und errechnet, wie eine Bewegung durch geometrische, anthropometrische und psychische Einflüsse beeinflusst wird. Die realistische Dynamik eines Menschen könnte betrachtet werden und das in Abhängigkeit von Umwelteinflüssen. Damit wären die Möglichkeiten der virtuellen Entwicklung von Produkten und ergonomischen Bewertung derselben, wesentlich verbessert.

Jedoch fehlt heute die Datengrundlage, um solch komplexes menschliche Verhalten ableiten zu können; letztlich, weil die Datenerhebung schwierig, langwierig und kostspielig ist. Es stellt sich also die Frage, ob es Möglichkeiten gibt, Bewegungen einfach und schnell zu messen.

## **1.2 Stand der Forschung und Technik**

Die Bewegungsmessung des Menschen hat heute nicht nur in der Ergonomie Bedeutung, sondern auch im Bereich der Informatik. Deshalb folgt zunächst ein genereller Überblick über die Entwicklung verschiedener Ansätze, die im Bereich Ergonomie und Informatik entwickelt wurden. In den anschließenden Kapiteln werden diese Methoden strukturell aufgearbeitet und die Anforderung an ein ideales Bewegungsmessverfahren für die Ergonomie definiert.

### **1.2.1 Bewegungsmessung in der Ergonomie**

Das Beobachten menschlicher Bewegungen hat in der Ergonomie eine lange Tradition. Motiviert, auf der einen Seite, durch den Wunsch dynamische Arbeitsabläufe im Sinne der Arbeitswissenschaft optimieren zu können, auf der anderen Seite die Dynamik des Menschen an sich zu verstehen, entstanden im Laufe der Zeit eine Reihe von Konzepten, Methoden und Untersuchungswerkzeugen. Im Folgenden soll ein kurzer historischer Einblick in dieses Teilgebiet der Ergonomie gegeben werden.

Die ersten Regeln zur Arbeitsorganisation und Arbeitszeitmessung wurden unter S. de Vauban und besonders unter B. F. de Bélidor (1729) im Zuge französischer Festungsbauten

erstellt. Bald erkannte man einfache Zusammenhänge zwischen Ernährung und Leistung sowie zwischen Ermüdung und Leistungsabfall [29].

1895 präsentierten Braune und Fischer [8] eine Beobachtung des menschlichen Ganges, wobei sie ein Spurverfahren einsetzten, um die Bewegungen auf Filmplatten festzuhalten. Dies war eine der ersten Arbeiten, bei der mit zeitgemäßen Messmethoden Daten gewonnen wurden. In diesem Zusammenhang sind auch die Arbeiten von Taylor [68] zu erwähnen, der in der Midvale Steele Company umfangreiche Experimente mit Arbeitern machte. Er versuchte aus den Beobachtungen der „besten Arbeiter“ Verallgemeinerungen zu treffen, um Arbeitsvorgänge optimieren zu können.

1917 veröffentlichte das Ehepaar F.B. und L.M. Gilbreth [25] Ergebnisse, die mit Hilfe der Zyklografie erstellt wurden. Zyklografie ist eine Methode, bei der die menschliche Bewegung mit Hilfe von Glühlämpchen, die am Körper, vorwiegend an Gelenken, befestigt werden, fotografisch in Form von Leuchtspuren festgehalten wird. Diese zweidimensionale Methode erlaubt zyklische Arbeitsvorgänge im Hinblick auf Ermüdungserscheinungen zu beurteilen (Abbildung 1.1).

Parallel zur Zyklografie, fanden auch gewöhnliche Filmaufnahmen, wie Barnes [5] darstellt, Einzug in die Arbeitswissenschaft. Die Motografie, die vor allem von Baum [6] geprägt wurde, wurde bis hinein in die Mitte der 80er Jahre verwendet und weiterentwickelt. Baum befestigte Glühlämpchen an Probanden und erhielt, durch entsprechend lange Belichtungszeiten, Leuchtspuren dieser Glühlämpchen. Durch einen kurzen Blitz in die Szenerie, konnte er auch ein Bild des Probanden in die Leuchtspuraufnahme blenden. Durch Anwendung einer Spiegeltechnik konnte er ferner monokular dreidimensionale Informationen über Trajektorien gewinnen (siehe Abbildung 1.2).

Heutzutage wird statt der Motografie ein ganz ähnliches Verfahren eingesetzt, dass statt der Glühlämpchen passive Markierungen (sog. Marker) am Körper benutzt (vgl. SIMI MOTION [76] oder VICON [77]). So benutzt VICON kleine reflektierende Kugeln, die am Probanden mit Hilfe von Klettbindern befestigt werden (siehe Abbildung 1.3). Während eines Experiments wird die Versuchsperson mit 2 und mehr Videokameras überwacht. An den Kameras sind infrarot Emitter angebracht, die kurze Blitze aussenden. Das Infrarotlicht wird an den Markern des Probanden reflektiert und durch die Kameras detektiert. Eine on-board realisierte Bildverarbeitung extrahiert die Position der Marker. Durch Kombination der Signale aller Kameras wird die Lage im Raum eines jeden sichtbaren Markers berechnet und die Trajektorien der Bewegung ermittelt. Ganz ähnlich arbeitet das System von SIMI wobei hier im rein visuellen Wellenlängenbereich gearbeitet wird. Diese Bewegungsanalysesysteme sind kommerziell erhältlich und äußerst weit entwickelt. Neben ausgereifter Messtechnik bieten

diese Systeme auch eine komfortable Software für die Weiterverarbeitung der Bewegungsdaten. Einen aufschlussreichen Artikel zum Einsatz, Gebrauch und Funktionsweise markerbasierter Messsysteme in der Ergonomie publizierte Quesada [48].

Einen neuartigen Ansatz beschreibt Speyer et al. [67]. Speyer benutzt als Marker aktive Elektronikelemente, die mittels eines enganliegenden Spezialanzugs direkt am Probanden fixiert sind. Bei diesen Elektronikelementen handelt es sich um Mini-Gyroskope, die ihre aktuelle Orientierung und Beschleunigung an einen Sender funken. Durch eine geeignete Software werden durch zweimaliges Integrieren aus den Messdaten die Raumlage eines Körperelements und durch Kopplung an ein dreidimensionales Menschmodell die Haltung ermittelt. Großer Vorteil dieser Methode ist, dass die Signale der Sensoren per Funk übertragen werden. Damit kann es zu keinerlei Verdeckung von Körperelementen kommen. Ferner ist der Empfänger klein genug, um das System mobil einzusetzen. Leider ist bis heute die Messgenauigkeit nicht ausreichend, um genaue Haltungs- und Bewegungsmessungen durchzuführen.

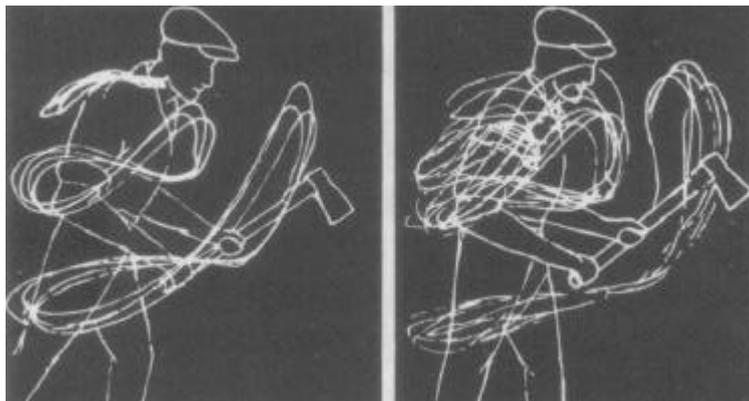


Abbildung 1.1: Zyklografische Aufnahmen von Arbeitsbewegungen mit eingezeichneter Arbeitsperson. Das linke Bild zeigt den Arbeitsvorgang im nicht ermüdeten Zustand, während rechts anhand der Aufweitung der Bewegungsspuren eine deutliche Ermüdung zu erkennen ist (Abb. aus Schmidtke [58], Seite 138).



Abbildung 1.2: Motografische Aufnahme des Wählens einer Nummer mit Tasten-telefon. Die eingespiegelte zweite Ansicht erlaubt eine dreidimensionale Analyse der Bewegungspur (Abb. aus [6]).



Abbildung 1.3: Klassische Methode zur Bewegungsmessung: Die Lage der Marker, wird mit Hilfe von Bildverarbeitung dreidimensional ermittelt. Von der Raum-koordinate der Marker wird auf die Bewegung des Körpers geschlossen (Foto: Vicon).

## 1.2.2 Bewegungsmessung in der Informatik

Verfahren zur Bewegungsmessung des Menschen sind heute in der Informatik (speziell der Bereich Computer Vision), aber auch im Bereich Elektrotechnik und Physik ein Thema. Während sich die Ergonomie - bzw. verwandte Bereiche wie Biomechanik u.ä. - von der anwendungsbezogenen Seite der Bewegungsmessung nähert, kommt die Computerwissenschaft von der prinzipiellen, algorithmischen Seite. Heute wie früher wird in diesem Zusammenhang meist Videotechnologie in Verbindung mit Bildverarbeitung zur Analyse von Bewegungen eingesetzt. Frühere Arbeiten beschäftigten sich daher insbesondere mit der Segmentierung von Menschen, also der Unterscheidung zwischen Hintergrund und Person, dem Erkennen von Haltungen, dem Nachführen von Bewegung oder der Interpretation von Bewegungen.

Die Zielsetzungen der visuellen Menscherkennung sind sehr weit gestreut. So finden sich Untersuchungen zu Gesichts- und Mimikerkennung, die zum Ziel haben, Sprache an den Lippen und Gesichtszügen „abzulesen“ oder im Sinne einer Überwachung Personen zu erkennen (z.B. Inst. f. Phonetik [43] oder Radig [49]). Die Gestikinterpretation wird oft in Verbindung mit der Mensch-Maschine-Kommunikation eingesetzt, wobei häufig versucht wird, online die Bewegung von Hand oder Arm zu interpretieren (Bakic und Stockman [4], Davis und Bobick [18], Pentland [45], Morguet und Lang [42]). In diesem Zusammenhang sei vor allem das Projekt „HAL“ am M.I.T. erwähnt, in dessen Rahmen es um die Realisierung eines intelligenten Raums geht, der auf die Bedürfnisse des Menschen reagiert, indem die Gestik und Sprache des Menschen interpretiert wird. Bewegungen – jedoch nicht nur die des Menschen – werden auch detektiert, um Filme optimal, d.h. in der Regel platzsparend, speichern zu können (z.B. Dateiformat MPEG). In der Automobiltechnik, setzt man die Menscherkennung - angetrieben vom Gedanken „Autonomer Fahrzeuge“ ein -, um beispielsweise Fußgänger automatisch erkennen zu können. Eine andere Anwendung im Automobilbau ist die Messung von Position und Haltung der Passagiere, um den Airbag im Falle eines Unfalls optimal auslösen zu können [21]. Ausführliche Überblicke vermitteln die Publikationen von Gavrilu [22], Aggarwal und Cai [1], Wachter [71] und von Kinzel [36].

Prinzipiell müssen zwei Methoden unterschieden werden, die Positionsinformation von Personen in ihrer Umwelt liefern. Das sind zum einen jene Methoden, die am Menschen Markierungen anbringen, mit deren Hilfe indirekt die Person detektiert wird, und die Methoden, die versuchen die Person direkt zu detektieren. Im Folgenden werden nur letztere Verfahren vorgestellt und gezeigt, wo der Stand der Forschung heute ist.

Bei der Messung menschlicher Haltung und Bewegung tauchen zwei Probleme auf. Zum einen stellt sich die prinzipielle Frage, wie man Versuchsperson und Umgebung trennt. Dies

führt zum Problem der Segmentierung. Zum anderen stellt sich die Frage, wie man die Haltung des Menschen (der Mensch stellt ein mehrgliedriges, bewegliches Objekt dar) beschreibt.

Das Problem der Segmentierung kann durch mehrere Ansätze z.T. gelöst werden. Sollen Messungen auf Basis handelsüblicher Videokameras erfolgen, so bietet sich beispielsweise die „Blue-Box-Methode“ an [17]. Bei dieser, aus Film und Fernsehen bekannten Methode, wird der Hintergrund monochromatisch gehalten. Durch entsprechende Filter wird die Farbe des Hintergrundes herausgefiltert, so dass eine vollständige Trennung zwischen Vordergrundobjekt (i. d. R. eine Person) und Hintergrund erfolgt, sofern das Vordergrundobjekt nicht die Farbe des Hintergrundes enthält. Diese Methode kann jedoch nur angewandt werden, wenn man den Hintergrund wirklich monochromatisch gestalten kann, was dazu führt, dass im wesentlichen eine planare Hintergrundwand und ggf. der Fußboden herausgefiltert werden kann. Bei komplexeren Hintergründen, kommt es sehr leicht zu Reflexionen und Abschattungen. Dies führt zu unbefriedigender Extraktion des Hintergrundes.

Verwandt mit dieser Methode ist die Schwellwertmethode, bei welcher der Hintergrund dunkler oder heller als das Objekt gehalten wird und somit eine Grauwertschwelle erzeugt wird, mit deren Hilfe das Objekt segmentiert wird. Rehg und Kanade [51] oder Mochimaru [41] setzten diese Methode zum Detektieren von Hand- und Fingerhaltungen ein. Durch geeignete Wahl der Beleuchtung kann diese Art der Segmentierung sehr stabil erfolgen. So verwenden Davis und Bobick [18] einen Infrarot-Ansatz, der eine bessere Trennung von Hintergrund und Person erlaubt, als die traditionelle Methode im sichtbaren Wellenlängenbereich [47].

Eine andere Methode stellt die Differenzmethode dar, die Hogg [30] beschreibt. Dabei wird ein Bild der Umgebung, ohne das zu beobachtende Objekt, verglichen, mit einem Bild, welches das Objekt erhält. Durch einfache Subtraktion kann der Hintergrund eliminiert und das Objekt segmentiert werden – ein klassisches Segmentierungsverfahren. Durch dieses Verfahren ist es möglich Objekte auch in einer komplexen Umgebung zu segmentieren. Dieses Verfahren ist allerdings sehr empfindlich gegenüber Beleuchtungsänderungen, wie sie insbesondere durch Schattenwurf des Objekts auftreten (vgl. Abbildung 1.4).

Eine weitere Methode stellt die farbbasierte Segmentierung dar. Sie wird beispielsweise verwendet, um die Position des Gesichtes mit Hilfe des Merkmals „Gesichtsfarbe“ zu detektieren. Diese Methode ist ohne weitere Präparierung der Versuchsperson nicht geeignet um einen gesamten Menschen zu segmentieren. Wird jedoch mit Hilfe geeigneter Messkleidung die Person farblich markiert, so kann der Ansatz auf den gesamten Menschen ausgedehnt werden. Dies wurde beispielsweise in der Arbeit von Gavrilu [23] z.T. ausgenutzt, wobei sich die farbliche Kodierung der Person(en) auf einzelne Körperteile beschränkte, um rechtes und

linkes Bein oder rechten und linken Arm zu unterscheiden. Auch bei Ridder [53] wurde Farbinformation nicht nur für die Gesichtserkennung sondern auch im Sinne von Markern benutzt, die punktförmig die Position bestimmter Gelenkpunkte an der Person beschrieben. Dieser Ansatz ist vergleichbar mit klassischen Markertechniken.

Mit Hilfe der Segmentierung (entweder in kanten- oder flächenbasierter Form) lassen sich beispielsweise Schwerpunkt, Orientierung oder Form des Objekts bestimmen und analysieren. Durch sequentielle Betrachtung einer zeitlichen Abfolge von Segmentierungen kommt man zur Analyse und Interpretation von Bewegung, Haltung und Gestik des Menschen.

Bewegung stellt die räumliche Veränderung eines Objekts über der Zeit dar. Die Herausforderung besteht darin, mit robusten Verfahren ein Objekt zu *tracken*. Im einfachsten Fall, kann der Mensch als ein starres Objekt beschrieben werden, dessen Bewegung, sich aus dem Aneinanderreihen einzelner Segmentierungen ergibt. Beim Projekt „Intelligenter Bahnsteig“ von FORWISS wird durch Gesichtssegmentierung die Position des Kopfes ermittelt und in Raum und Zeit verfolgt. Diese Art der Bewegungs- bzw. Positionsbestimmung kann in verfeinerter Weise zur Bestimmung von Blickrichtungen eingesetzt werden. So bestimmen Bakic und Stockman [4] neben der Position des Kopfes auch die Position der Augen und damit die Blickrichtung.

Im 2-D-Ansatz von Haritaoglu, Harwood und Davis [27] werden die getrackten Menschen in einzelne Gebiete unterteilt, die Kopf, Torso, Arme und Beinen entsprechen. Mit Hilfe der Differenzmethode werden Personen erkannt und durch Formanalyse der segmentierten Menschen die Gebiete differenziert, d.h. den Bildmerkmalen wurden Labels zugeordnet. Die einzelnen Körperteile werden durch unverbundene Ellipsen beschrieben. Das weitere Tracking erfolgt durch *template matching*, d.h. in den Folgebildern wird nach gelabelten Bildmustern gesucht, die denen im Vorbild entsprechen. Ein differenzierteres Labeling betrieb Mochimaru und Yamazaki [40]. Sie untergliederten in ihrem 2-D-Ansatz den Menschen in Segmente wie Kopf, Torso, Oberarm, Unterarm, Hand usw., die jedoch ebenfalls nicht verbunden waren und versuchten durch *template matching* die Bewegung zu ermitteln.

Wren und Pentland [75] detektierten die Position von Kopf und der Hände durch Farbinterpretation. Die Ergebnisse des Trackings konnten mit einem einfachen gekoppelten Strichmenschmodell des menschlichen Oberkörpers bestehend aus Kopf, Torso und Armen sowie Händen gekoppelt werden. Das Hinzuziehen eines Menschmodells erlaubt die Interpretation der gemessenen Bewegungen und auch die Ableitung weiterer Daten, wie Armhaltung, die sich aus den geometrischen Einschränkungen des Modells abschätzen ließen.

Schließlich fanden die ersten Volumenmodelle des Menschen Einzug in die bildverarbeitungs-gestützte Interpretation menschlicher Bewegungen, wie beispielsweise in den Arbeiten von Rohr [54], Wachter und Nagel [70], Gavrilu [20] oder Yamamoto [78] zu sehen. Rohr analysierte monokulare Aufnahmen von Fußgängern in normaler Umgebung, wobei die Bewegung des Gehens parallel zur Bildebene ablaufen musste. Sein Menschmodell war aus Zylindern aufgebaut und repräsentierte nicht die tatsächliche äußere Form der Versuchsperson. Durch Analyse der Veränderungen in Folgebildern ermittelte Rohr die Bewegungsrichtung der Körperelemente. Im weiteren analysierte er Kanten in der Umgebung der Person und richtete die Projektion seiner zylindrischen Körperelemente an diesen Kanten, die zuvor linearisiert wurden, aus. Ganz ähnlich arbeitet das monokulare Verfahren von Wachter und Nagel, deren Menschmodell aus Kegelstümpfen besteht. Gavrilu erweiterte diesen Ansatz auf mehrere Kameraansichten und orientierte sein 3-D-Modell ebenfalls am gradientengefilterten Bild (chamfer matching) um so die Haltung zu erhalten. Die Versuchspersonen mussten bei Gavrilu einen speziellen Messanzug tragen, der die Kantendetektion ermöglichte. Auch Chung, Nagata und Ohnishi [16] nutzen Gradientenbilder, um Haltungen zu bestimmen, verzichteten jedoch auf ein detailliertes Volumenmodell. Yamamoto [78, 79] bestimmte durch Detektion von Markern die neue Lage seines kastenförmigen Volumenmodells. Sein binokularer Ansatz benötigte die Präparierung der Versuchsperson in der Form, dass diese einen weißen Anzug trägt, auf dem schwarze Markierungen angebracht sind (Abbildung 1.6).

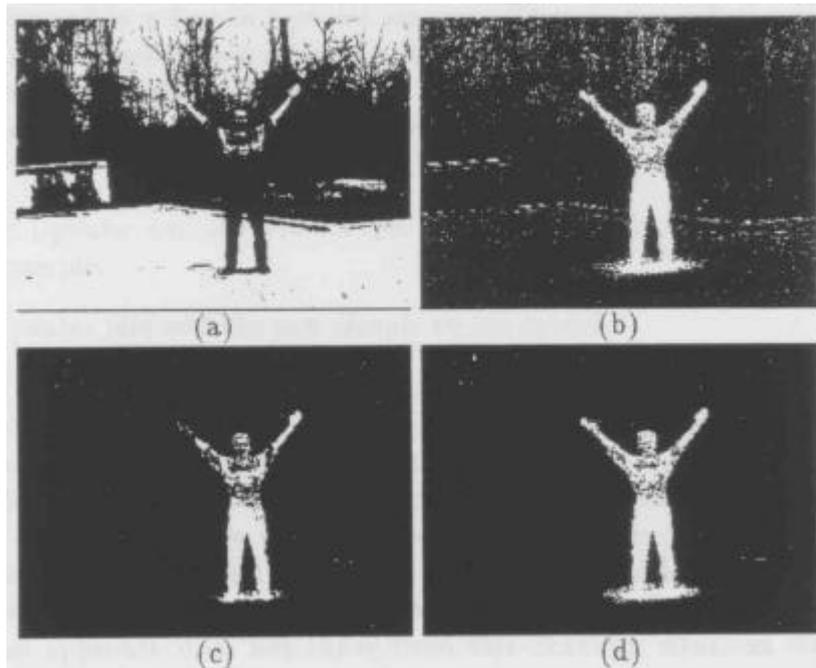


Abbildung 1.4: Darstellung der Differenzmethode zur Segmentierung einer Versuchsperson. Zu erkennen ist der Einfluss des durch die Versuchsperson erzeugten Schattens. In b) bis d) sind die Resultate unterschiedlicher Methoden dargestellt (aus Elgammal, Harwood und Davis [19]).



Abbildung 1.5: Detektion der Kopfposition durch Merkmal „Gesichtsfarbe“. Diese Aufnahme entstammt dem Projekt „Intelligenter Bahnsteig“ von FORWISS, TU München.

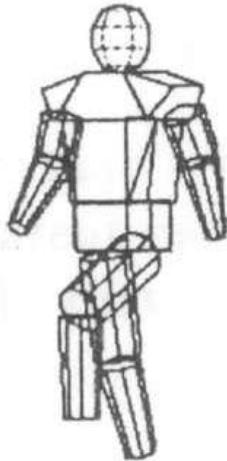


Abbildung 1.6: Yamamoto [79] ermittelt Bewegungen in Bildern mit Hilfe spezieller Kleidung und projiziert ein Modell darüber.

### 1.2.3 Zielsetzungen der Bewegungsmessung

Bei der Betrachtung des Themas Bewegungsmessung des Menschen kann festgestellt werden, dass heute prinzipiell drei Zielrichtungen bzw. Anwendungsfelder unterschieden werden müssen, nämlich

- das Erkennen von Personen
- das Erkennen von Gesten und
- das exakte Messen individueller Bewegungen.

Je nach Zielsetzung unterscheiden sich die entwickelten Ansätze und Methoden. Die Punkte „Erkennen von Personen“ und „Erkennen von Gesten“ sind Bereiche, die besonders in den Computerwissenschaften vertreten sind. Das exakte Messen individueller Bewegungen ist eine Domäne der Ergonomie und Biomechanik und somit die Zielsetzung, die hier im Vordergrund steht.

#### 1.2.3.1 Erkennen von Personen

Unter Erkennen von Personen ist die Beantwortung der Frage „Ist in der zu untersuchenden Szenerie eine Person vorhanden?“ gemeint. Anwendungsfelder sind z.B. die Überwachung von Bahnsteigen, Banken oder Straßen (Inst. f. Phonetik [43], Radig [49]). Bei diesen Anwendungen wird aufgrund von Bildern einer Überwachungskamera detektiert, wo Personen sind (siehe Abbildung 1.5). In der Regel kann auch ermittelt werden, in welche Richtungen

sich die Personen bewegen. Damit lässt sich beispielsweise, wie Heinrich [28] zeigt, ein Fußgängerdetektor realisieren. Ein Kamerasystem im fahrenden Auto analysiert hierbei fortlaufend die Umgebung und warnt, sobald sich eine Person der Fahrbahn kritisch nähert. Bei den Systemen zur Personenerkennung ist es unerheblich wie sich einzelne Körperelemente einer Person bewegen. Auch die Genauigkeit der Ortsbestimmung im Raum ist sekundär. Diese Systeme zielen vielmehr darauf ab, die Person als ein starres Objekt zu detektieren und zu interpretieren.

### *1.2.3.2 Erkennen von Gesten*

Beim Erkennen von Gesten geht es i.d.R. darum, Bewegungen der Hände und Arme oder des Gesichts zu interpretieren, um z.B. einem Computer Information zu übertragen. Ziel ist eine einfachere, natürlichere Mensch-Maschine-Kommunikation zu ermöglichen (Bakic und Stockman [4], Davis und Bobick [18], Pentland [45], Morguet und Lang [42]). Bei der Gestikererkennung müssen einzelnen Haltungen der Hände oder Arme gegenüber anderen getrennt werden. Bei der Handgestikererkennung wird beispielsweise über die Form der Hand ermittelt welche Fingerhaltung eine Person einnimmt (Davis und Bobick [18], Rehg und Kanade [51], Mochimaru [41]). D.h. die Forminterpretation ist der Prozess, um Handgesten zu unterscheiden, nicht zwingend die exakte Position der einzelnen Finger. Dieses Vorgehen findet sich bei fast allen Gestikinterpretationen.

### *1.2.3.3 Exaktes Messen individueller Bewegungen*

Während bei den oben genannten Punkten die exakte Position des Körpers bzw. eines Körperelements nicht relevant war, ist dies bei der Messung individueller Bewegungen ein entscheidender Punkt. Denn hier sollen die Haltungswinkel einer Person genau erfasst werden, um Bewegungen vergleichen zu können, um statistische Modelle über Bewegungsmuster erzeugen zu können oder weitergehende Modellierung dynamischer menschlicher Eigenschaften zu ermöglichen. Für diese Zwecke werden heute ausschließlich Bewegungsmesssysteme verwendet, die mit so genannten Markern arbeiten. Dabei wird eine Versuchsperson mit Markern, meist in der Form von Kugeln präpariert, die ihr auf Körperelemente geklebt werden. Diese Marker sind technisch leicht detektierbar und die Position im Raum dieser Marker ist exakt bestimmbar. Somit kann aufgrund der Lage der Marker auf die Lage von Körperelementen wie Kopf, Arme und Beine geschlossen werden (vgl. SIMI MOTION [76] und VICON [77]).

## 1.2.4 Systematik exakter Bewegungsmessverfahren

Die Verfahren, mit denen exakt Bewegungen gemessen werden können, lassen sich heute in drei Kategorien einteilen. Es sind dies die Verfahren, die zweidimensional, die dreidimensional und die dreidimensional unter Verwendung eines Menschmodells arbeiten.

### 1.2.4.1 2-D-Methoden

Mit zweidimensionalen Methoden lassen sich Bewegungen innerhalb einer Bewegungsebene messen. In der Regel erfolgt die Erfassung der Bewegung mit einer Kamera, d.h. die Bewegungsebene ist parallel zur Kameraebene (vgl. Abbildung 1.7). Letztlich wurde so schon durch Braune, 1895, Bewegung interpretiert und auch die Zyklografie ist ein 2-D-Verfahren. Mit diesen älteren Techniken lassen sich nur qualitative Aussagen zur beobachteten Bewegung machen. Heute, liegen i.d.R. die Daten digital vor, so dass sich quantitative Analysen durchführen lassen. Mit einer einfachen Videokamera, deren Bilder digitalisiert werden, können so einfache Bewegungsanalysen durchgeführt werden. In Abbildung 1.8 ist ein Bild aus einem Videofilm gezeigt. Die aufgeklebten Markierungen ermöglichen dabei die Positionsbestimmung derselben. Die Positionsbestimmung kann manuell erfolgen. Durch heutige Möglichkeiten der Bildverarbeitung ist es jedoch relativ einfach, die Markerpositionen automatisch zu messen. Die resultierenden Orts-Zeit-Daten ermöglichen weitere Analysen wie z.B. Fourieranalyse, um Bewegungsmuster genauer zu betrachten.

Derartige 2-D-Verfahren sind gut geeignet um schnell und unkompliziert qualitative Bewegungsaussagen zu treffen. Es ist allerdings schwierig, mit diesen Daten detailliertere Modellierungen vorzunehmen. So ist z.B. nicht gewährleistet, dass die Bewegung wirklich nur in der Bildebene abläuft, d.h. die Komponente senkrecht zur Bildebene bleibt verborgen. Aus diesem Grund können auch nur bedingt Haltungswinkel bestimmt werden.

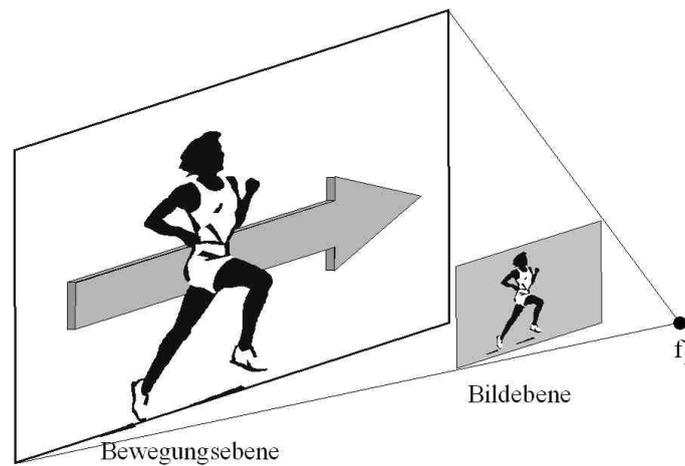


Abbildung 1.7: Beim zweidimensionalen Ansatz muss die Bewegungsrichtung in der Bildebene liegen.



Abbildung 1.8: Seitenansicht eines Kleinkindes beim Zähneputzen. Mit Hilfe der Markierungen an Kopf, Oberarm, Unterarm und Hand kann die Zahnputzbewegung zweidimensional analysiert werden (vgl. Seitz et al. [65]).

#### 1.2.4.2 3-D-Methoden ohne Modell

Wird der 2-D-Ansatz um eine Kamera erweitert, so ist es möglich durch Triangulation von Punkten, die in beiden Kameraansichten zu sehen sind - sog. korrespondierende Punkte -, die Position dreidimensional zu erfassen. Somit ist die Bewegungsmessung unabhängig von der Bildebene möglich und es können alle Bewegungen frei im Raum gemessen werden

(vgl. Abbildung 1.9). Da der menschliche Körper keine markanten Punkte besitzt, die ausgemessen werden könnten, werden externe Marker angebracht, wie dies Abbildung 1.3 zeigt. Über die Lage der Marker im Raum wird die Bewegung beschrieben. Dazu ist es jedoch nötig, dass die Marker an definierten Stellen des Probanden angebracht sind, so dass klar definiert ist, welcher Marker mit welchen knöchernen Strukturen korrespondiert; d.h., um z.B. die Bewegung des Unterarms beschreiben zu können, wird mindestens ein Marker am Ellbogengelenk und einer am Handgelenk benötigt. Mit diesen zwei Markern kann die Lage des Unterarms im Raum beschrieben werden. Zur Vergleichbarkeit der Daten zu anderen Versuchspersonen ist es wichtig, dass die Marker immer an den gleichen Körperpunkten befestigt werden, wie es z.B. in der Benutzeranleitung zu den Systemen SIMI MOTION oder VICON beschrieben wird. In der Realität ist dies allerdings schwierig.

Die richtige Interpretation der Markerkoordinaten setzt voraus, dass die Markerposition bzgl. des korrespondierenden Körperelements konstant bleibt. D.h. Verschiebungen der Marker durch Weichteilverformung oder durch Kollision mit der Umgebung werden nicht erfasst, treten aber in der Realität auf und stellen demnach ein Problem dar.

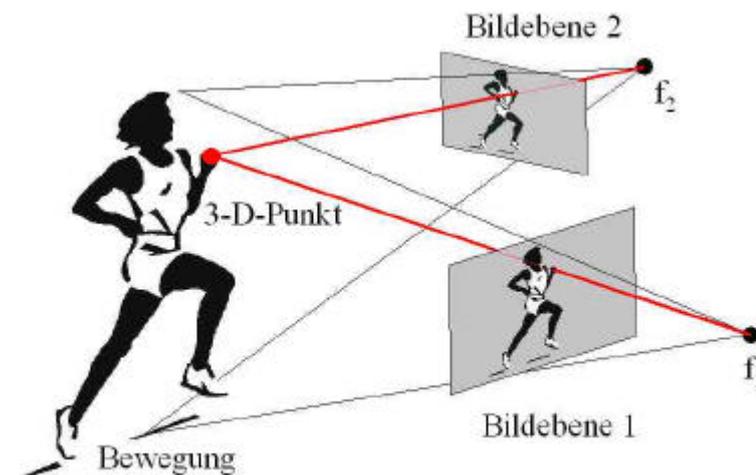


Abbildung 1.9: Mit Hilfe von zwei Kameras (oder mehr) können Punkte im Raum per Triangulation dreidimensional vermessen werden.

#### 1.2.4.3 3-D-Methoden mit Modell

Die bisherige Klassifizierung der generellen Methoden stützte sich auf Markierungen geeigneter Art, die am Probanden befestigt werden. Die 2-D- und 3-D-Methoden liefern dabei die Position und Zeitpunkt der Messung einer Markierung. Stellt sich die Frage, wie der Übergang von den aus messtechnischen Gründen benötigten Markern und körpereigenen Elementen realisiert werden soll, um z.B. eine Körperhaltung durch Gelenkwinkel zu beschreiben. Deshalb ist im Bereich Menschmodellierung nicht die Frage, wo ist ein Marker, sondern die Frage, wo ist ein Gelenkpunkt oder Körperelement, zu beantworten. Durch die wahre Kenntnis der Körperposition eines Probanden im Raum, kann die eigentliche Modellierung menschlichen physischen Verhaltens erst beginnen. Diese Lücke zwischen Proband und Markerpositionen wird mit Hilfe von Menschmodellen versucht zu schließen.

Die einfachste Version stellt ein einfaches Strichmodell dar, das die Verbindung z.B. eines Ellenbogenmarkers zum Handgelenkmarker und weiteren Gelenken zueinander definiert, d.h. die Markerpositionen werden durch Linien miteinander verbunden, so dass die Kinematik des Menschen abgebildet wird. Ridder [53] beispielsweise verwendete ein solches Strichmodell (Abbildung 1.10). Mit dessen Hilfe ist es z.B. durch Hypothesenbildung möglich, aus all den detektierten Markern die richtigen miteinander zu verbinden. Ferner ist es möglich zeitweise verdeckte Marker wieder zu finden, da, aufgrund der Strichkopplung, nur ein eingeschränkter Bewegungsraum für einzelne Marker zur Verfügung steht. Die Einführung der Kopplung von Markern durch Linien hat hier also weniger etwas damit zu tun, dass es sich hierbei um ein Menschmodell handeln soll, als vielmehr damit, dass dadurch die technische Handhabung (Wiederfinden von Markern, Verbinden der richtigen Marker, Einschränkung und Vorausberechnung von Aufenthaltsmöglichkeiten für Marker) verbessert wird.

Oben erwähnte Strichmodelle sind an der Oberfläche des Probanden angebracht. Die Markerposition ist zugleich ein Gelenkpunkt des Strichmodells. Abbildung 1.11 zeigt, dass die anatomische Gelenkposition nicht detektiert wird, wenn die Marker nicht an selbiger angebracht sind. Es resultiert hieraus eine große Differenz zwischen Modellgelenk und anatomischen Gelenk. Ferner sind die Gelenkpunkte an der Oberfläche des Körpers angebracht. Anatomische Gelenke liegen jedoch im Körper und sind nicht sichtbar. D.h. ein Strichmodell müsste erst noch in geeigneter Form transformiert werden, so dass es die wahren Gelenkörper repräsentiert. Das wird mit etwas komplexeren Strichmodellen realisiert, z.B. durch die Kopplung eines Strichmodells an die Markerpositionen, in der Art, dass die Linien in etwa die Lage von Knochen und die Gelenke zwischen den Linien, die Lage der tatsächlichen Gelenke repräsentieren. Hierzu muss das Strichmodell durch Annahmen über Körper- und Skelettbau an die richtige Position gebracht werden. Im VICON-Messsystem wird dieses Verfahren ange-

wandt. Zur Bestimmung der Hüftgelenklage, beispielsweise, müssen Marker an Becken, Oberschenkel und Knie nach genau definierten Regeln angebracht werden. Durch die geometrische Anordnung der Marker kann das System auf die wahre Gelenkposition der Hüfte und des Knies schließen. Damit sind die Möglichkeiten mit reinen Strichmenschmodellen Gelenkpositionen und reale Haltungswinkel weitgehend erschöpft.

Die logische Erweiterung stellen Volumenmenschmodelle dar, die neben dem Skelett des Menschen auch dessen Oberfläche repräsentieren. In der Informatik sind hierzu einige Arbeiten durchgeführt worden. Besonders hervorzuheben sind dabei die in Kapitel 1.2.2 zitierten Ansätze von Yamamoto [78, 79] und Gavrilin [20]. Beide entwickelten ein Volumenmenschmodell, das sie in Videoaufnahmen projizierten. Yamamoto bestimmt dabei die Position des Menschmodells mit Hilfe von Markern an der Versuchsperson, Gavrilin hingegen analysiert Körperumrisse und richtet das Modell an diesen Umrissen aus. Vorteil beider Methoden ist es, dass das an die Marker oder Konturen angepasste Menschmodell die Lage von Gelenkpartnern und die Haltung konform zum verwendeten Modell beschreibt. Die Menschmodelle müssen an die Anthropometrie (die Maße des Menschen) der Probanden angepasst werden, d.h. das Modell soll die wahre Lage von Gelenkpunkten und die reale Oberfläche eines Menschen beschreiben.

Wie eigene Arbeiten zeigen (Abbildung 1.12) ist es mit der Überlagerung eines exakten Menschmodells möglich, auf eine definierte Anbringung von Markern zu verzichten. Die in Abbildung 1.13 gezeigte Kopplung eines Markers an das Menschmodell kann völlig frei erfolgen. Die neue Idee ist dabei, dass eine Versuchsperson mit Markern ausgestattet wird, so dass jedes Körperelement durch einen Marker detektiert werden kann. Der Unterschied zu heutigen Ansätzen ist dabei, dass der Operateur die Marker an *beliebigen* Stellen der betreffenden Körperelemente anbringen kann. Durch die Überlagerung eines Menschmodells erfolgt die Zuordnung zu Punkten des Menschmodells flexibel. Am sinnvollsten ist die Zuordnung zum nächstgelegenen Modellpunkt, in der Regel ein Punkt der Oberfläche des Menschmodells.

Durch diese Initialisierung von Bewegungsmessungen kann erstens darauf verzichtet werden, Marker immer an der selben Stelle anbringen zu müssen und zweitens können die Markerpositionen zu jeder Zeit in eine Haltung des benutzten Menschmodells überführt werden. Bei unterschiedlicher Platzierung der Marker bei verschiedenen Versuchspersonen oder bei derselben Versuchsperson im Falle von Folgemessungen (vgl. Abbildung 1.13, rechts), können, durch oben genannte Initialisierung, vergleichbare Daten erhoben werden, unter der Voraussetzung, dass das verwendete Modell möglichst exakt die Anthropometrie des Probanden widerspiegelt. Die bei gängigen 3-D-Methoden notwendige penible Präparierung der

Versuchsperson entfällt. Die Verwendung eines detaillierten dreidimensionalen Menschmodells kann also die Handhabung von Bewegungsmessungen vereinfachen und die Aussagekraft und Vergleichbarkeit entscheidend verbessern.

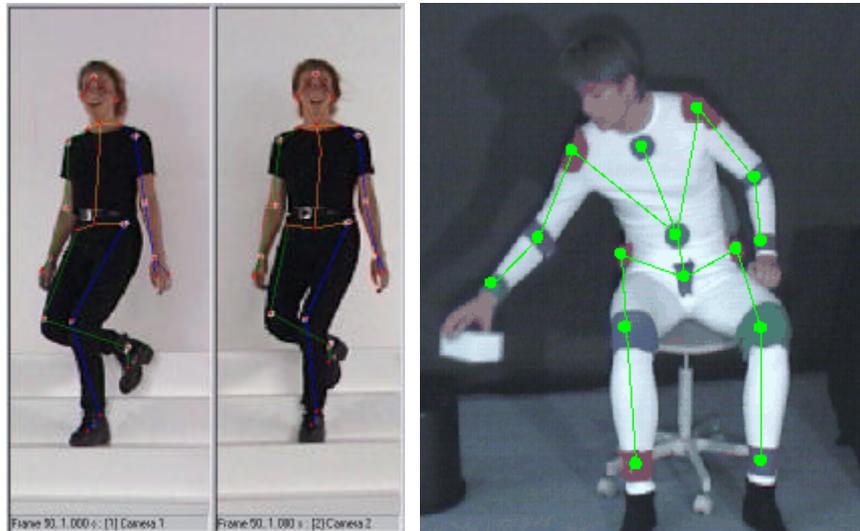


Abbildung 1.10: Einfache Strichmodelle. Links: Binokularen Messungen überlagertes Strichmodell (Foto: Fa. SIMI, Oberschleißheim). Rechts: Ebenfalls überlagertes Strichmodell (entnommen aus: Ridder [53]).

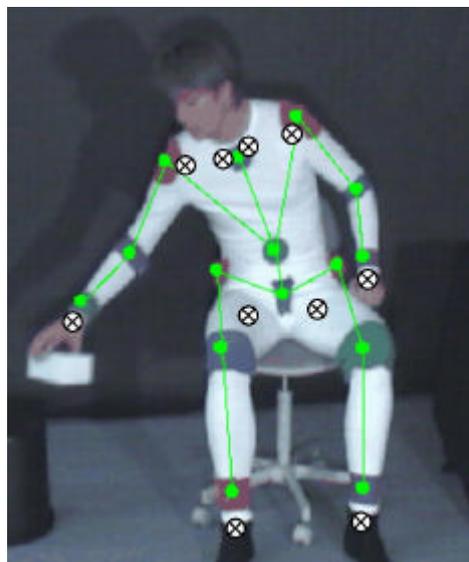


Abbildung 1.11: Diese Abbildung zeigt die Diskrepanz zwischen anatomisch korrekten Gelenkpositionen (Kreuze) und denen, die durch Marker bzw. ein Strichmenschmodell ermittelt werden, wenn die Marker nicht exakt an der Position der anatomischen Gelenkpositionen sind.

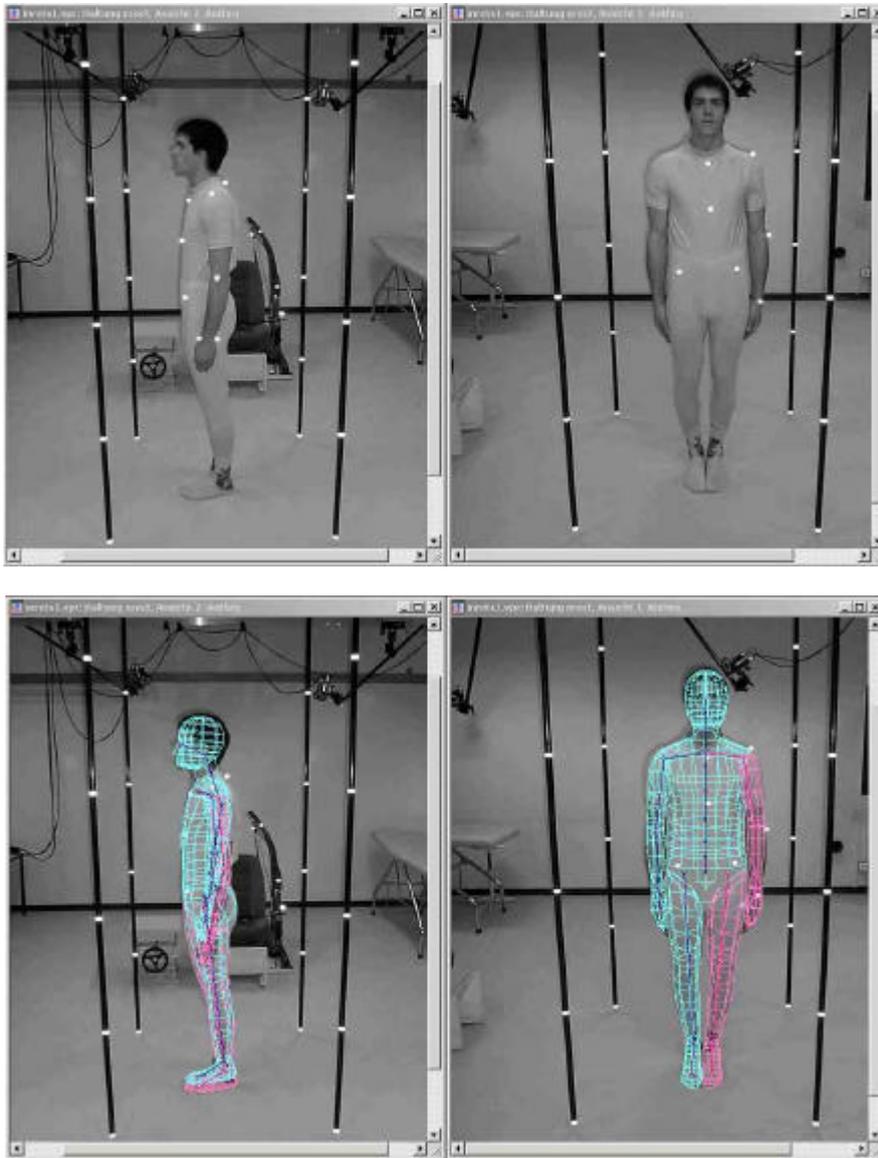


Abbildung 1.12: Oben: Versuchsperson, mit Markern präpariert. Unten: Überlages Menschmodell. Jetzt kann räumlich definiert werden, wie die Marker mit dem Modell verbunden werden.

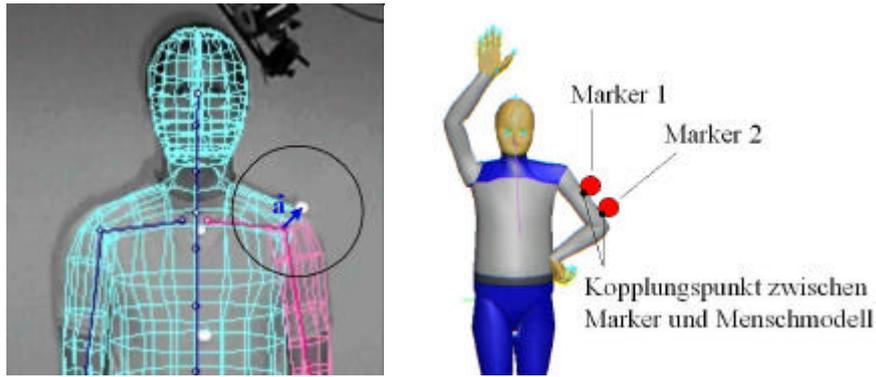


Abbildung 1.13: Links: Durch die exakte Überlagerung von Menschmodell und Versuchsperson, kann die Kopplung zwischen den Markern und dem Modell erfolgen. Hier ist die Kopplung eines Schultermarkers mit dem Schultergelenk dargestellt. Der Marker kann aber mit beliebigen anderen sinnvollen Modellpunkten gekoppelt werden (z.B. dem nächstgelegenen Hautpunkt). Rechts: Durch den Kopplungspunkt bzw. -vektor kann durch Marker 1 oder Marker 2 die Haltung des Oberarms gleichermaßen definiert werden, denn die Kopplung stellt eine eindeutige Verbindung zum Modell her. Die Markerlage ist also unerheblich.

### 1.3 „Ideales“ Bewegungsmesssystem

Wie gesehen, existieren heute verschiedene Verfahren zur Messung der menschlichen Bewegung, die Marker verwenden, welche an den Probanden angebracht werden. Doch das Anbringen von Markern ist zeitaufwändig. Außerdem stellt sich die Frage, wie von Markerpositionen auf tatsächliche reale Gelenkwinkel oder Gelenkpositionen geschlossen werden soll. Hierzu ist die Einführung eines Menschmodells, das den Probanden anthropometrisch repräsentiert, notwendig. Nur so lassen sich Markerinformationen modellbasiert interpretieren. Dennoch, die Marker können verrutschen, ohne dass ein Messsystem dies erfassen kann oder es kann zu Kollisionen zwischen Markern und der Umwelt kommen, so dass Bewegungsabläufe gestört werden.

Technische Systeme brauchen Marker, um exakt Positionen im Raum messen zu können. Doch die Natur führt uns täglich vor, dass höhere Lebewesen insbesondere der Mensch, täglich im eigenen Alltag die Bewegung und Position von anderen Objekten (wie Autos, Straßenbahnen oder anderen Menschen) so zuverlässig „messen“ und interpretieren kann, dass Kollisionen zwischen anderen Objekten relativ selten sind oder ein Vorgang, wie das Fangen eines Balls mit ein wenig Übung sehr gut vom Menschen bewältigt wird. Es drängt sich die Frage auf, ob das Verfahren, wie Lebewesen ihre Umwelt „messen“ nicht ein mögli-

cher Lösungsansatz für ein Bewegungsmesssystem ist. Solch ein System könnte die Bewegung des Menschen ad hoc messen. Jegliche Präparierung von Versuchspersonen würde entfallen und die Versuchsperson würde selbst kaum merken, dass sie gerade gemessen wird – man könnte von einem idealen Bewegungsmesssystem sprechen.

Das ideale Messsystem beeinflusst also in keiner Weise den Probanden (Verzicht auf Marker o.ä.) und ist ferner schnell und einfach zu handhaben. Darüber hinaus ermöglicht es Messungen in gewöhnlicher Alltagskleidung, so dass Probanden sich nicht umziehen müssen bzw. Messkleidung tragen müssen, die deren Wohlbefinden stört. Hinzu kommt, dass das ideale Messsystem alle relevanten Daten über Positionen und Winkel dreidimensional und konform zu einem dreidimensionalen Modell des Probanden generiert. Messwerte verschiedener Versuchspersonen sind dann über das Menschmodell direkt vergleichbar und alle Daten, welche die Anthropometrie beschreiben, jederzeit verfügbar (vgl. Tabelle 1.1).

<b>Forderung ans System</b>	<b>Folgerung aus Forderung</b>
Versuchsperson nicht beeinflussen	Bewegungsmessung ohne Marker o.ä.
Anatomische Gelenkörper messen	Technik, um „in die Tiefe zu blicken“
Einfache Handhabung und mobil	Geringer Hardwareeinsatz

Tabelle 1.1: Forderungen an das „Ideale Bewegungsmesssystem“ und resultierende Folgerungen für den zu wählenden Ansatz.

Damit sind der Ansatz und die Zielsetzung dieser Arbeit definiert. Es soll untersucht werden inwieweit und mit welchen Methoden und Mitteln ein ideales Bewegungsmesssystem realisiert werden kann, dass, erstens, die Versuchsperson nicht beeinflusst und, zweitens, modellbasierte Daten liefert, so dass jeder Zeitpunkt einer Bewegung im Computer komplett reproduzierbar ist.

Die Vorgehensweise gliedert sich nach dieser Definition im Wesentlichen in vier Hauptbereiche.

1. Menschmodelle: Menschmodelle spielen bei der Bewegungsmessung und Interpretation eine entscheidende Rolle. Menschmodelle stellen das Kopplungsglied zwischen Messpunkten der Bewegungsmessung und der Anthropometrie einer Versuchsperson dar. Sie ermöglichen eine computergestützte Darstellung. Es wird gezeigt, wie individuelle Menschmodelle erzeugt werden, welche die Anthropometrie eines Probanden repräsentieren und einen virtuellen Zwilling einer Versuchsperson darstellen. Des Weiteren wird gezeigt, wie mit Hilfe eines Modells Haltungen eines Probanden ermittelt werden.

2. Physiologisches Konzept des Bewegungssehens: Die Natur entwickelte ein Konzept der Bewegungsmessung und –interpretation, das es Mensch und Tier erlaubt, die Dynamik seiner Umwelt zu erfassen. Aus den Konzepten, welche die Natur hierbei einsetzt, werden die Ansätze für die Entwicklung eines Bewegungsmesssystems entwickelt, das ohne Marker auskommen kann und Bewegungen direkt in das verwendete Modell transformiert.
3. Bildverarbeitungsverfahren: Die physiologischen Konzepte des Bewegungssehens werden mit Hilfe der digitalen Bildverarbeitung umgesetzt. Hierbei wird im Wesentlichen die Korrelations- und Kantenganalyse verwendet, um strukturelle Bildänderungen zu detektieren – analog zur Muster- und Kantenganalyse, wie sie von der Natur verwendet wird.
4. Realisierung: Bei der Realisierung wurden die Erkenntnisse der Punkte 1, 2 und 3 zusammengeführt und ein Messsystem entwickelt und getestet das die heutigen Möglichkeiten und Grenzen der Entwicklung eines „idealen“ Messsystems“ aufzeigt.

## 2 Menschmodell RAMSIS und das Messsystem PCMAN

Das Menschmodell RAMSIS ist ein digitales Menschmodell zur Simulation menschlicher Anthropometrie und Haltung im CAD. Es wird verwendet, um Fahrzeuge, hauptsächlich Pkws, ergonomisch auszulegen. Neben RAMSIS existieren in der Ergonomie noch andere digitale Menschmodelle. Die Einführung von Menschmodellen generierte in der Vergangenheit ein neues Arbeitsfeld – die sog. Menschmodellierung.

Die Menschmodellierung (im internationalen Sprachgebrauch *digital human modeling* genannt) ist ein Teilgebiet der Arbeitswissenschaft bzw. der Ergonomie, das vor allem in den letzten Jahren in Industrie und Forschung etabliert wurde. Man versteht darunter die Nachbildung menschlicher Eigenschaften im weitesten Sinne, also physische und psychische, durch geeignete Computerprogramme und die Anwendung dieser virtuellen Menschen zur *Konstruktion* und *Auslegung* jeglicher *Mensch-Maschine-Systeme*, wie z.B. dem Cockpit im Automobil (Package-Analyse), dem Büroarbeitsplatz oder der Produktionsplanung an Fließbändern (Bubb [10]). Die Menschmodellierung ist in der Regel integriert in CAD-Systeme.

Zur Menschmodellierung gehören zum einen Menschmodelle, welche die geometrischen und kinematischen Eigenschaften des Menschen simulieren und Messtechniken, welche die Datengrundlagen für die eigentliche Modellierung liefern. Ausgezeichnete Übersichten der verschiedenen Entwicklungen auf diesem Gebiet finden sich bei Peacock und Karwowski [44], bei Karwowski, Gendaidy und Asfour [34], bei Badler, Phillips und Webber [3] sowie bei Raschke, Schutte und Chaffin [50]. Darüber hinaus finden sich Überblicke zur Menschmodellierung zum einen aus Sicht der amerikanischen Luftwaffe und zum anderen aus Sicht des Automobilbaus in den Artikeln von McDaniel [38] und Bubb [11].

### 2.1 Entwicklungsgeschichte ergonomischer Menschmodelle

Heute existieren im Wesentlichen drei Menschmodelle, RAMSIS, SAFEWORK und JACK, die kommerziell erwerbbar und im ingenieurwissenschaftlichen Bereich einsetzbar sind. Neben diesen drei Modellen wurde aber im Laufe der Zeit eine große Anzahl verschiedener Entwicklungen verfolgt. Die folgenden Abschnitte sollen einen historischen Überblick hierzu geben.

Die Entwicklung der Menschmodellierung kann mit der Entwicklung von BOEMAN in den 60er Jahren des 20. Jahrhunderts begonnen werden. Ryan und Springer des amerikanischen Flugzeugherstellers Boeing entwickelten ein Programm zur Simulation von Erreich-

barkeiten für einen Durchschnittsmenschen in Jagdflugzeugen (Ryan und Springer [55]). Die nichtlineare Optimierung der Haltungen war damals derart zeitaufwendig, so dass BOEMAN nicht in Alltagsfragestellungen angewendet werden konnte.

In den 70er Jahren wurde BOEMAN dann vom medizinischen Forschungslabor für Luft- und Raumfahrt (ARML) der amerikanischen Luftwaffe übernommen und das Model zur Haltungssimulation vereinfacht sowie die Möglichkeit implementiert, eine größere Anzahl männlicher und weiblicher Anthropometrien simulieren zu können, die in verschiedene Militärflugzeugtypen integriert werden konnten. Das aus den Forschungen des ARML resultierende Menschmodell wurde unter dem Namen COMBIMAN bekannt.

In den 80er Jahren wurde erkannt, dass COMBIMAN nicht nur zum Zwecke der Cockpitauslegung, sondern auch zur Simulation von Wartungs- und Servicearbeiten an Flugzeugen benutzt werden könnte. COMBIMAN wurde somit für die Simulation von stehenden, hockenden oder knienden Arbeiten mit Benutzung von Werkzeugen erweitert. Das resultierende Menschmodell bekam den Namen CrewChief. Die durch CrewChief aufzubringenden Kräfte bei den verschiedenen zu simulierenden Arbeiten, basierten auf einer Datenbank, die innerhalb des ARML durch Versuche ermittelt wurden.

Ein allgemeineres biomechanisches Modell zur Vorhersage populationsabhängiger Kraftvorhersagen wurde von Chaffin et al. [15] in den späten 60ern entwickelt. Hauptziel dieser Arbeit war die Simulation statischer Kräfte welche von Astronauten aufzubringen sind, die während eines Aufenthalts im All oder auf dem Mond, Objekte heben, ziehen oder schieben müssen. Hierzu wurden Kraftmessungen an über 2000 Menschen gemacht. Das hieraus resultierende Programm 3DSSPP (3D Static Strength Prediction Program) ist vor allem in den USA verbreitet und wird von Behörden und Organisationen eingesetzt.

Parallel zu den Entwicklungen in den USA, wurde in England das System SAMMIE durch Case, Porter und Bonnes [13] entwickelt, das schon perzentilierte Anthropometrien verwendete und z.B. Spiegelsicht bei Kraftfahrzeugen bewerten konnte. Daneben sind in den 80ern viele andere Systeme programmiert worden, wie ERGOMAN vom Laboratoire d'Anthropologie Appliquée et d'Ecole Hamaine in Paris, WERNER vom Institut für Arbeitsphysiologie an der Universität Dortmund, Tommy der TU Dresden [33], HEINER von der TU Darmstadt [57], sowie ANYBODY und ANTHROPOS der deutschen Firma IST.

In Kanada wurde ebenfalls während der 80er Jahre das Modell SAFEWORK an der Ecole Polytechnique Montreal entwickelt. Neben der Simulation verschiedener anthropometrischer Dimensionen bietet SAFEWORK auch die Möglichkeit mit Hilfe inverser Kinematik einfache Bewegungen zu simulieren.

Das Modell JACK wurde unter Federführung der NASA zusammen mit der Universität von Pennsylvania Mitte der 80er entwickelt [3]. Ursprünglich war dieses Modell, das zunächst den Namen TEMPUS trug, für die Arbeitsplanung beim Zusammenbau der heutigen Raumstation ISS konzipiert worden. Es sollten Fragen geklärt werden, wie z.B. ein Astronaut am besten Gegenstände erreichen und befestigen kann, aber auch, wie die Sicht des Astronauten mit Raumanzug ist. Im Laufe der Jahre wurde JACK auch für Analysen an Militärflugzeugen und anderen Fahrzeugen benutzt. JACK wurde mit einer anthropometrischen Datenbank ausgestattet sowie mit einer beweglichen Wirbelsäule und Gelenken, so dass JACK mit inverser Kinematik positioniert werden konnte. Ferner ermöglicht kraftgeführte Haltungs- und Bewegungssimulation die Berechnung komplexerer Aufgaben.

Speziell zugeschnitten für die Bedürfnisse in der Luftfahrtindustrie ist das Menschmodell BMD-HMS, das Boeing-McDonell-Douglas-Human-Modeling-System, welches 1990 eingeführt wurde. BMD-HMS wird benutzt, um manuelle Tätigkeiten die für Wartung und Inspektion, aber auch für die Bedienung von Flugzeugen, nötig sind, zu visualisieren und zu analysieren. Einen detaillierteren Überblick, leider fast nur über die Entwicklungen auf angelsächsischer Seite, bietet Chaffin [14].

Seit Ende der 80er Jahre wurde in Deutschland in Zusammenarbeit mit der deutschen Automobilindustrie, der Katholischen Universität Eichstätt, der Fa. tecmath und dem Lehrstuhl für Ergonomie der Technischen Universität München, das Menschmodell RAMSIS (Rechnergestütztes Anthropologisch-Mathematisches System zur Insassen Simulation) entwickelt. RAMSIS ist speziell für die Bedürfnisse der Automobilindustrie zugeschnitten und ermöglicht

- die perzentilierte und korrekte Darstellung des Menschen unter Zugrundelegung geeigneter Populationen
- die Simulation von realistischen und statistisch abgesicherten Körperhaltungen in Abhängigkeit geometrischer Einschränkungen
- die Bewertung des Komfortempfindens bei gegebener Haltung
- ergonomische Analysefunktionen z.B. für die Sicht oder den Gurtverlauf.

RAMSIS wird heute fast weltweit in der Automobilindustrie zur Packagebewertung und – konstruktion (vgl. Abbildung 2.1) eingesetzt, da nahezu alle Populationen anthropometrisch repräsentiert werden können (wie z.B. Deutsche, Japaner, Koreaner oder Amerikaner). Neben Erwachsenen können auch Kinder simuliert werden (Arlt und Marach [2]). RAMSIS findet auch Einsatz bei der Auslegung von Motorrädern (Kolling [37]), Flurförderfahrzeugen (Seitz [64]) und anderen Arbeitsplätzen.



Abbildung 2.1: Das Menschmodell RAMSIS, mit dessen Hilfe die Fahrzeug-Ergonomie ausgelegt wird, in der CAD-Umgebung eines KFZ (Foto: Human Solutions GmbH).

## 2.2 Aufbau des Modells RAMSIS

In dieser Arbeit wird die Struktur des Menschmodells RAMSIS verwendet. Deshalb soll an dieser Stelle das verwendete Modell dargestellt werden.

Leitgedanke bei der Entwicklung des Modells RAMSIS war, ein anthropometrisch und kinematisch korrektes 3-D-Computermenschmodell zu entwickeln, quasi einen virtuellen Zwilling des Menschen, wie Bubb [12] ausführt. Folglich ist RAMSIS analog zur menschlichen Anatomie konstruiert. Wie der Mensch, so hat RAMSIS ebenfalls ein Skelett, das sog. **innere Modell**, und eine Haut, das sog. **äußere Modell** (Abbildung 2.2).

Das innere Modell dient der Repräsentation der Kinematik und Längenmaße, wie Bein- oder Armlängen. Es ist dafür verantwortlich, Haltungen und Bewegungen exakt nachstellen zu können. Einzige Stützpunkte des Skeletts sind dabei die realen Gelenkdrehpunkte. Dies gilt allerdings nicht für die Wirbelsäule; aufgrund der Komplexität der Wirbelsäule wurde das Wirbelsäulenmodell vereinfacht und enthält weniger Wirbel als die echte Wirbelsäule. Die Beweglichkeit ist jedoch trotzdem vergleichbar.

An das innere Skelettmodell ist das Hautmodell angehängt und repräsentiert die äußere Erscheinung des Menschen. Die Haut ist jener Modellteil, der die Wechselwirkung mit der Umwelt übernimmt, z.B. den Kontakt beim Greifen eines Gegenstandes oder der Kontakt zwischen Stuhl und Gesäß beim Sitzen. Das Hautmodell definiert ferner durch das Volumen, das es umschließt, physikalische Eigenschaften, wie die Masse des virtuellen Menschen und dessen Trägheitseigenschaften.

Eine detaillierte Beschreibung des Aufbaus von innerem und äußerem RAMSIS-Modell findet sich bei Seidl [60] oder Geuß [24]. In den zwei folgenden Abschnitten sind die wichtigsten Merkmale zur Verdeutlichung dargestellt.

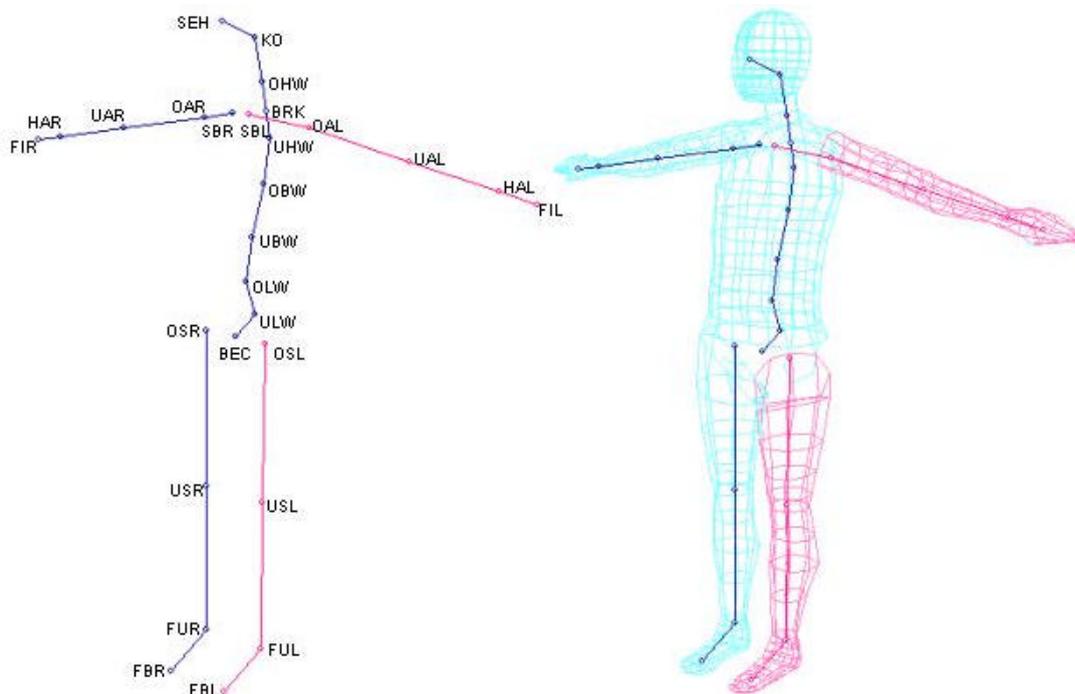


Abbildung 2.2: Inneres und äußeres Modell von RAMSIS. Am inneren Modell sind zusätzlich die Kurzbezeichnungen der Gelenke zu sehen, die hier Verwendung finden.

### 2.2.1 Struktur des inneren Modells

Das innere Modell besteht aus Elementkoordinatensystemen, die jeweils in einem Gelenkpunkt liegen. Die Elementkoordinatensysteme sind kartesische Koordinatensysteme. Liegt der Ursprung in einem Gelenkpunkt, so zeigt die Achse *Tangente* (T) in Richtung des nächsten Gelenkpunktes, der zugleich Startpunkt eines neuen Körperelements ist (Abbildung 2.3). Generell ist jeder Gelenkpunkt auch einem Körperelement zugeordnet. So ist beispielsweise

das linke Hüft- bzw. Oberschenkelgelenk, kurz OSL (vgl. Abbildung 2.2), auch diesem Körperelement (Oberschenkel) zugeordnet. Die Achse *Normale* (N) zeigt in Richtung Vorderseite des Körperelements. Die dritte Achse, die *Binormale* (B) ergibt sich dann durch das Kreuzprodukt aus Tangente und Normale. Jeder Gelenkpunkt ist relativ zu seinem Vorgängerpunkt durch einen Satz Elementkoordinaten definiert. Um das Modell zu bewegen dient das Gelenkkoordinatensystem. Es ist ebenfalls kartesisch und hat seinen Ursprung, wie das Elementkoordinatensystem, in Gelenkpunkten. Durch Anwendung geeigneter Drehmatrizen kann so ein Körperelement um ein Gelenk gedreht werden.

Hieraus ergibt sich der hierarchische Aufbau des Modells RAMSIS. Anfangspunkt oder Ursprung des Modells bildet hierbei das Hüftzentrum, die Mitte zwischen der Verbindungslinie vom rechten Oberschenkelgelenk zum linken, Kurzbezeichnung BEC (vergleiche Abkürzungen der Gelenkpunkte in Abbildung 2.2). Ans BEC sind zunächst die Gelenke der Wirbelsäule bis zum Augenpunkt in ihrer logischen Reihenfolge gekoppelt, es schließen sich rechter und linker Arm bis zu den Elementen FIL und FIR und schließlich die Elemente der Beine bis zu FBL und FBR an.

## **2.2.2 Struktur des äußeren Modells**

Das äußere Modell, das die Haut des Menschmodells repräsentiert, hat seinen Ursprung an jedem Körperelement und ist in dessen Koordinatensystem definiert. Die Haut wird durch Punkte repräsentiert, die untereinander verbunden sind und so einzelne übereinander gestapelte Scheiben bilden. Die Anzahl der Eckpunkte einer Scheibe eines Körperelements kann unterschiedlich sein. Die Arme haben 8 Eckpunkte, während die Füße oder der Kopf, die einen höheren Detaillierungsgrad besitzen, um die gesamte Vielfalt möglicher Oberflächen befriedigend abbilden zu können, 16 – 24 Eck- oder Stützpunkte haben. Die meisten Scheiben sind eben, d.h. sie liegen senkrecht auf der Verbindungslinie zwischen zwei Gelenkpunkten. Definiert wird die Lage der Eckpunkte im T-B-N-System.

Da der Abstand eines Hautpunkts zum inneren Modell innerhalb der zugehörigen Scheibenebene, frei skalierbar ist, kann jede äußere Anthropometrie also die Oberfläche einer Person nachgebildet werden. Durch vorherige Skalierung des inneren Modells ist somit eine sehr detaillierte Repräsentierung einer realen Person möglich. Ferner lassen sich auch beliebige perzentilierte Anthropometrien erzeugen.

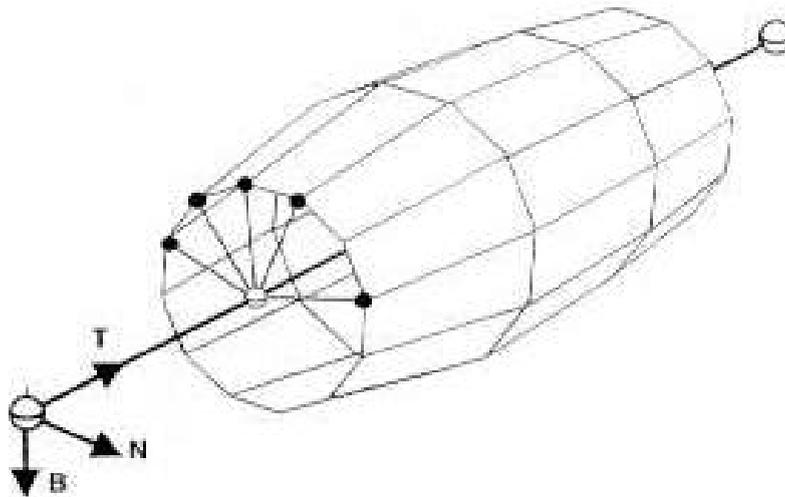


Abbildung 2.3: Äußeres Modell eines Körperelements (entnommen aus Seidl [60]).

## 2.3 PCMAN – das RAMSIS-Messsystem für Anthropometrie- und Haltungsmessung

PCMAN ist ein Anthropometrie- und Haltungsmesssystem, das die Modellstruktur des Menschmodells RAMSIS benutzt. Es wird seit Anfang der 90er Jahre am Lehrstuhl für Ergonomie entwickelt bzw. weiterentwickelt (Geuß [24], Seidl [60], Seitz [61], Seitz, Balzulat und Bubb [62]) und basiert auf Ansätzen von Weiß [72]. PCMAN bietet die Möglichkeit, dreidimensionale Anthropometrien und Haltungen auf Basis des Menschmodells RAMSIS anhand digitalisierter Fotos einer Versuchsperson zu messen. Hierzu wird unter Berücksichtigung perspektivischer Bedingungen das Menschmodell den Bildern überlagert und die Kontur des Modells mit der Kontur der Versuchsperson zur Deckung gebracht. Zur Erzeugung der Bilder werden mehrere handelsübliche CCD- oder Videokameras eingesetzt, welche den Probanden aus unterschiedlichen Perspektiven zeigen. Dadurch ist das System sehr mobil und flexibel einsetzbar (z.B. für Studien im fahrenden Auto oder in Fabrikhallen).

Das Vorgehen mit PCMAN untergliedert sich in drei Stufen:

1. Kalibrieren
2. Anthropometrische Modelladaption und
3. Haltungsanpassung

Mit Hilfe des Systems PCMAN kann ein Proband anthropometrisch beschrieben werden, sowie individuelle Gelenkwinkel gemessen werden. Auf Grundlage dieser Daten erfolgen weitere Analysen oder Modellierungen, z.B. Erzeugung anthropometrischer Typologien oder Erzeugung von Haltungsmodellen.

### 2.3.1 Prinzip

Grundprinzip von PCMAN ist der Überlagerungsgedanke. Dabei werden mit Video- oder Digitalkameras Fotos eines Probanden erstellt und in diese Fotos das dreidimensionale Menschmodell projiziert. Wenn das Menschmodell genau die Anthropometrie des Probanden trägt und die perspektivischen Verhältnisse der Fotos bekannt sind, so lässt sich das Modell exakt mit der Versuchsperson überlagern. Durch Benutzung von zwei oder mehr Kameras kann eine Szene mit Hilfe des Triangulationsprinzips dreidimensional erfasst werden. In Abbildung 2.4 ist dies dargestellt.

Das PCMAN-Prinzip ermöglicht zum einen durch anpeilen eines Punktes in den Bildern, der in allen Bildern eindeutig zu erkennen sein muss, dessen Lage im Raum zu ermitteln. PCMAN kann also als optisches 3-D-Messprogramm verwendet werden. Zum anderen erlaubt die Software das perspektivisch korrekte Projizieren des Menschmodells in die Bilder, so dass das Modell virtuell am Ort des Probanden zu stehen scheint.

Das zugrunde liegende Triangulationsprinzip, basiert auf der geometrischen Grundlage, dass zur exakten Vermessung eines Punktes im Raum mindestens zwei Sichtlinien aus unterschiedlichen Perspektiven benötigt werden. Der Schnittpunkt der Geraden definiert einen Punkt im Raum. Jeder Bildpunkt definiert dabei eine Sichtlinie vom Brennpunkt der Kamera durch ein Pixel der Bildebene zum realen Punkt. Wird eine Szene von zwei Kameras aus unterschiedlichen Perspektiven gefilmt, und ist  $g_1$  die Gerade durch die Bildkoordinate  $(x_1, y_1)$  von Bild  $B_1$  und  $g_2$  die Gerade durch die Bildkoordinate  $(x_2, y_2)$  von Bild  $B_2$ , dann schneiden sich  $g_1$  und  $g_2$  im Punkt  $P$ , wenn  $P$  ein so genannter korrespondierender Punkt ist, also ein Punkt, der in allen Bildern eindeutig zu sehen ist. Durch Auswahl korrespondierender Bildpunkte  $(x_1, y_1)$  und  $(x_2, y_2)$  können am Probanden anthropometrische Maße abgenommen werden.

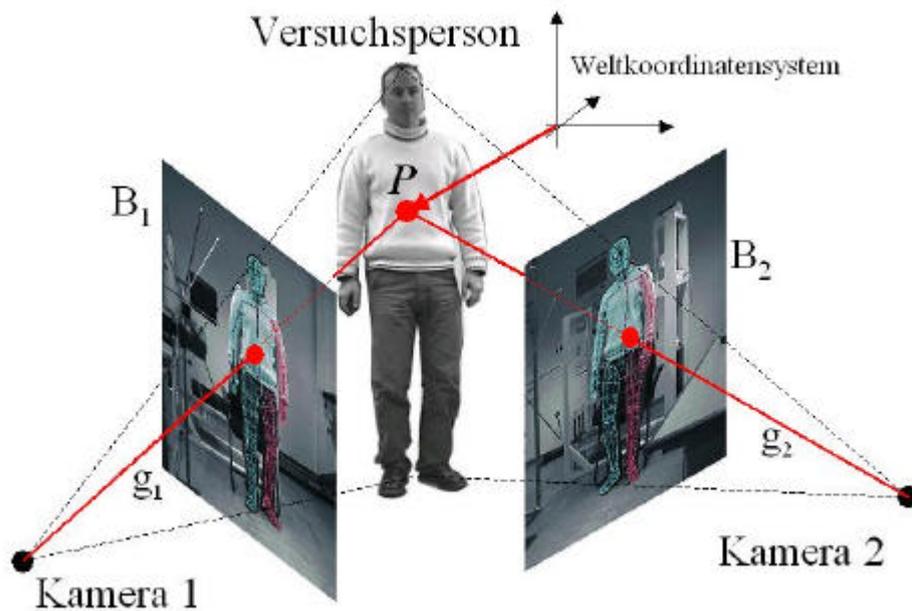


Abbildung 2.4: Das PCMAN-Prinzip. Mit Hilfe von Bildern einer Versuchsperson aus unterschiedlichen Perspektiven, lassen sich per Triangulation dreidimensionale Informationen gewinnen. In die Bilder wird das Menschmodell RAMSIS projiziert, das virtuell an die gleiche Stelle im Raum gerückt wird, wie die Versuchsperson und deshalb exakt mit dieser zur Deckung kommt. Es resultiert eine modellbezogene dreidimensionale Beschreibung von Ort und Haltung der Versuchsperson.

### 2.3.2 Kalibrierung

Bevor Messungen mit PCMAN erfolgen können, ist es notwendig das System zu kalibrieren. Mit Hilfe der Kalibrierung kann die Software die Kamerastandorte bzgl. eines Koordinatensystems, das durch den verwendeten Kalibrierkörper definiert wird, bestimmen und – zusammen mit Kameraparametern wie Brennweite und CCD-Größe – die Perspektive ermitteln. Um die Relation der Kameras zueinander mathematisch erfassen zu können, wird ein Kalibrierkörper verwendet, dessen Abmessungen genau bekannt sind. Prinzipiell können alle möglichen geometrischen Formen als Kalibrierkörper verwendet werden, doch wird hier in der Regel ein spezielles Messgestänge verwendet, wie es in Abbildung 2.5 gezeigt wird.

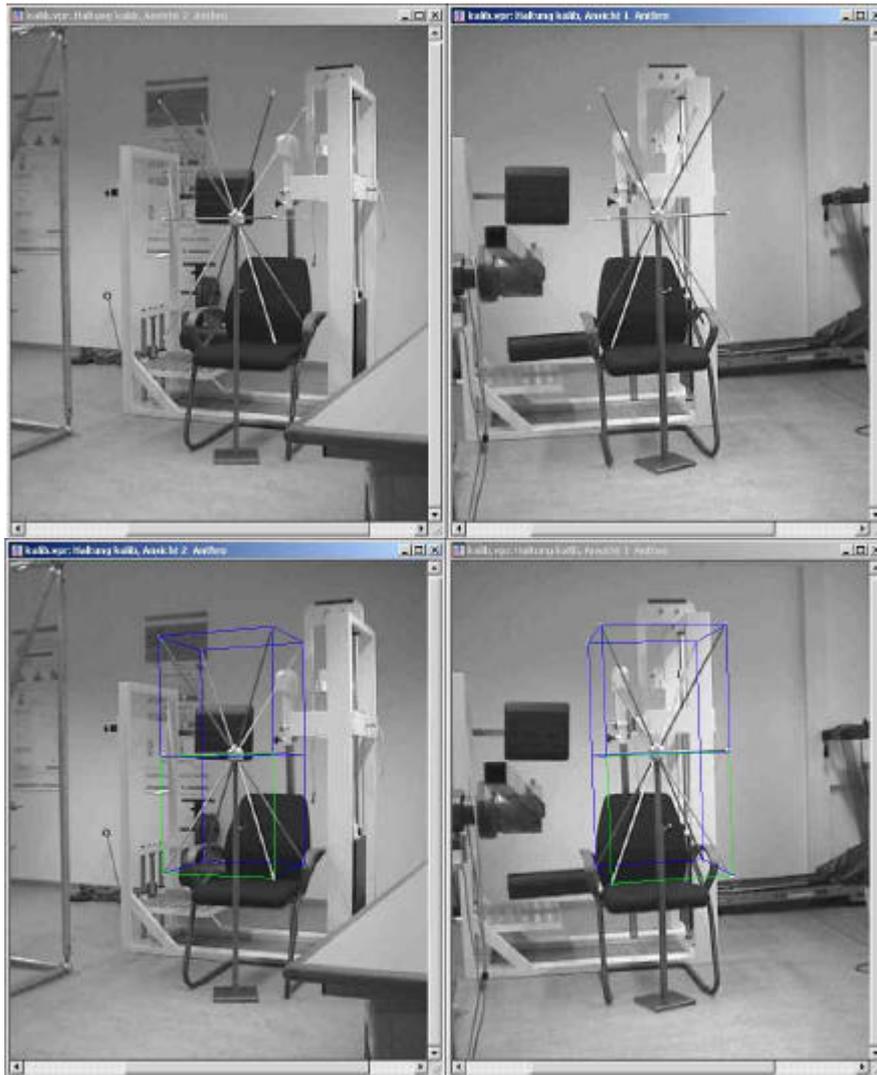


Abbildung 2.5: Oben: Der Standardkalibrierkörper. Unten: Nach dem manuellen anklicken der Enden der Stangen des Kalibrierkörpers wird ein Modell des Körpers überlagert, so dass der Benutzer ein Feedback über die Güte der Kalibrierung erhält. Die Ecken des Modells müssen sich dabei exakt mit den Enden der Stangen decken.

Der Benutzer selektiert zur Kalibrierung manuell die Endpunkte der Stangen des Kalibrierkörpers. Es wird mit Hilfe eines Algorithmus auf Basis des Levenberg-Marquard-Verfahrens [46] in die Bilder projiziert, so dass der Benutzer zum einen sieht, ob die Kalibrierung erfolgreich war - das Modell muss sich exakt mit dem Kalibrierkörper decken - und zum andern, ist nun definiert unter welcher Perspektive der Körper von der Kamera gesehen wird. Damit ist der Kamerastandort definiert. Alle weiteren geometrischen Berechnungen können nun auf Basis dieser Kalibrierung abgeleitet werden.

Da das System jede eigene Definition von Kalibrierobjekten erlaubt, können auch individuelle und an Experimente angepasste Kalibrierkörper entwickelt und benutzt werden. In Abbildung 2.6 ist eine solche individuelle Kalibrierung gezeigt. Hierbei wurde die Geometrie des Kalibrierkörpers so definiert, dass sie sich exakt mit prominenten Punkten (z.B. gut sichtbare Ecken oder Schraubenköpfe) des gezeigten Flugsimulator-Cockpits deckt. Bei Experimenten in beengten Verhältnissen ist dies eine geeignete Methode, um eine gute Kalibrierung zu erreichen (Kaibel et al. [35]).

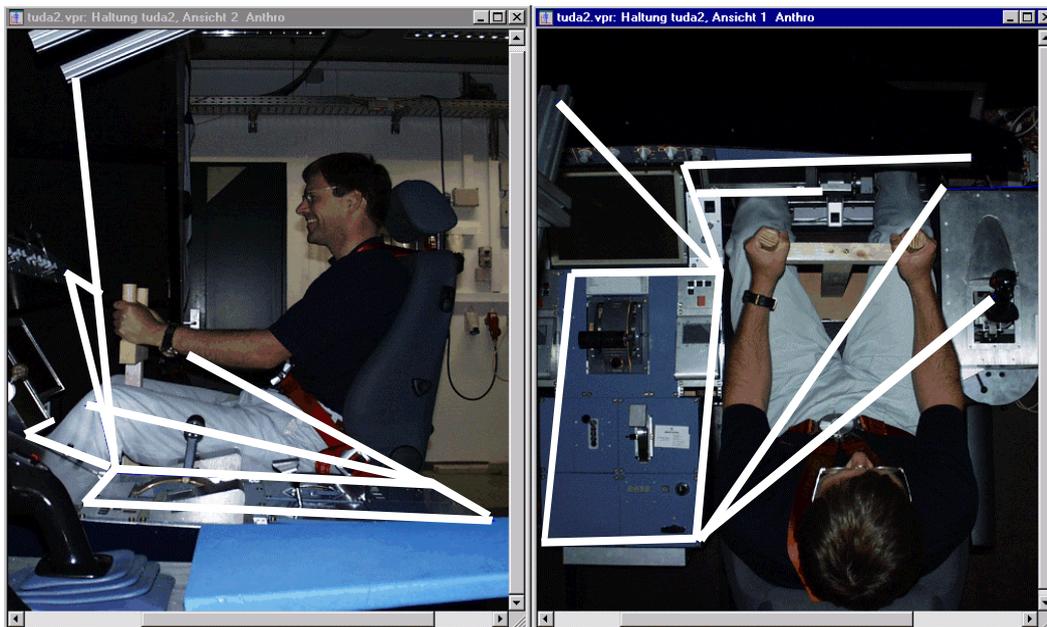


Abbildung 2.6: Kalibrierung eines Experiments durch markante Geometriepunkte eines Flugsimulator-Cockpits. Die überlagerten Linien stellen den Kalibrierkörper dar, dessen Relationen genau bekannt sind.

### 2.3.3 Anthropometrische Modelladaption

Das verwendete Menschmodell ist frei skalierbar, d.h. Knochenlängen und die Hautoberfläche lassen sich verändern (Abbildung 2.7). Die Vorgehensweise zur Adaption des Menschmodells an einen Probanden ist dabei fest vorgeschrieben. Da die Software nicht weiß, wo die Versuchsperson steht (Abbildung 2.10), schiebt der Benutzer das Modell zunächst über das Becken des Probanden. Ausgehend von dieser Position wird sukzessive die Beckenbreite, die Beinlänge, die Torsogröße, die Hals- und Kopfposition und die Armlänge manuell an den Probanden angepasst, so dass sich die Gelenkpunkte des Modells mit realen Gelenkorten decken. Da anatomische Gelenkörter nicht direkt sichtbar sind, werden hierfür verschie-

dene Haltungen benutzt. So kann das Hüftgelenk, das Kniegelenk und der Knöchel mit zwei verschiedenen Haltungen ermittelt werden. Dabei ist in einer Haltung das Bein gestreckt, in einer anderen angewinkelt (Abbildung 2.9). Aus geometrischen Gründen kann eine exakte Überlagerung beider Haltungen nur dann erfolgen, wenn die Unterschenkel- und Oberschenkellänge richtig justiert wird. Somit werden mit unterschiedlichen Haltungen zum einen Körperelementlängen richtig justiert und zum anderen die Gelenkdrehpunkte bestimmt. Mit Hilfe dieser beiden Messhaltungen, die Stehhaltung und Pharaohaltung genannt werden, werden die Gelenkkörper der Hüfte, des Knies, des Knöchels, der Schulter, des Ellbogens und des Handgelenks gefunden.

Nachdem durch obige Prozedur das Skelett des Menschmodells auf den Probanden skaliert wurde, muss die Oberfläche an die Versuchsperson adaptiert werden. Die Aufnahmen zur anthropometrischen Adaption des Menschmodells erfolgen in der Sagittal- und Frontalebene der Versuchsperson. Damit ist die Anpassung der Haut auch auf diese beiden Ebenen festgelegt, d.h. die Hautanpassung erfolgt auch im Modell nur in der Sagittal- und Frontalebene. Zur Adaption werden die Hautpunkte in der entsprechenden Ebene verschoben, bis sie sich mit der Kontur des Probanden deckt. Abbildung 2.8 zeigt die Adaption der Bauchoberfläche an eine Versuchsperson. Abbildung 2.11 zeigt eine fertig angepasste Anthropometrie.

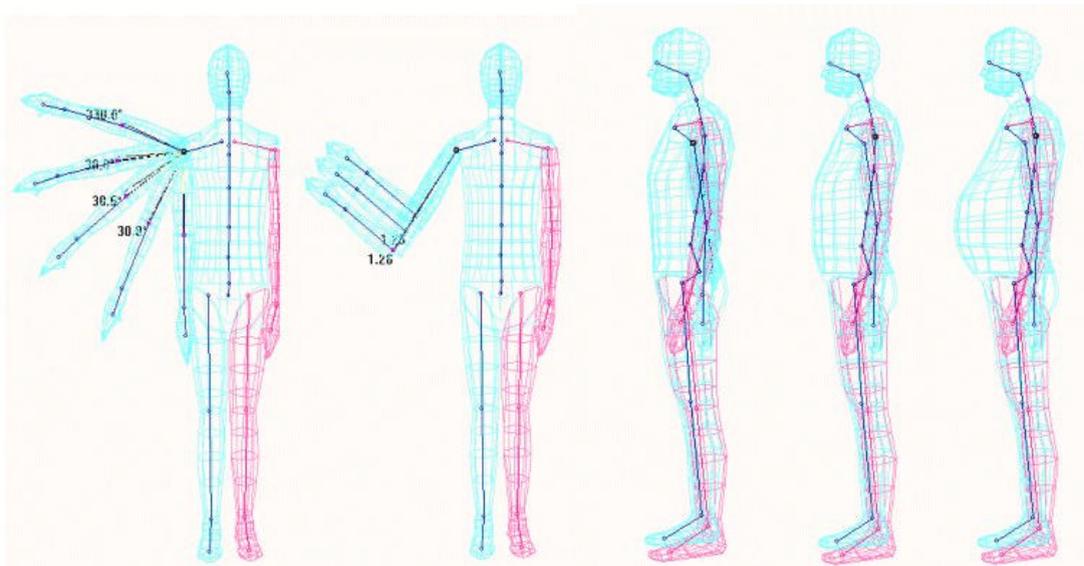


Abbildung 2.7: Verwandlung des Menschmodells. Alle Gelenke des Modells sind beweglich, doch darüber hinaus sind auch Knochenlängen und Oberflächenpunkte per Mausclick frei veränderbar und damit z.B. an ein Hintergrundbild, das einen Probanden zeigt, anpassbar.

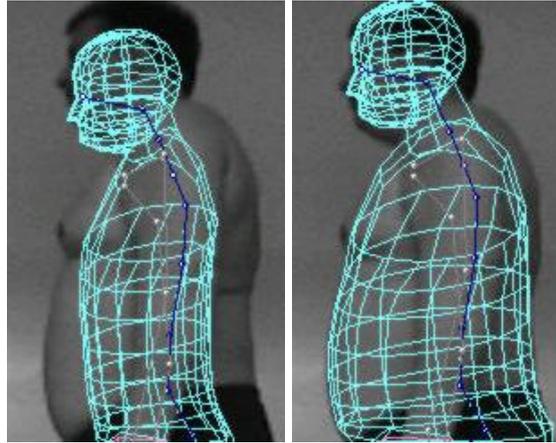


Abbildung 2.8: „Aufblasen“ des Hautmodells. Links ist der Anfangszustand gezeigt. Rechts der Endzustand, nachdem ein Operateur mit Hilfe von PCMAN manuell Hautpunkte in der Sagittal- und Frontalebene mit der Körperkontur zur Deckung brachte.



Abbildung 2.9: Zwei Messhaltungen mit jeweils zwei Bildpaaren die mit PCMAN bearbeitet werden.

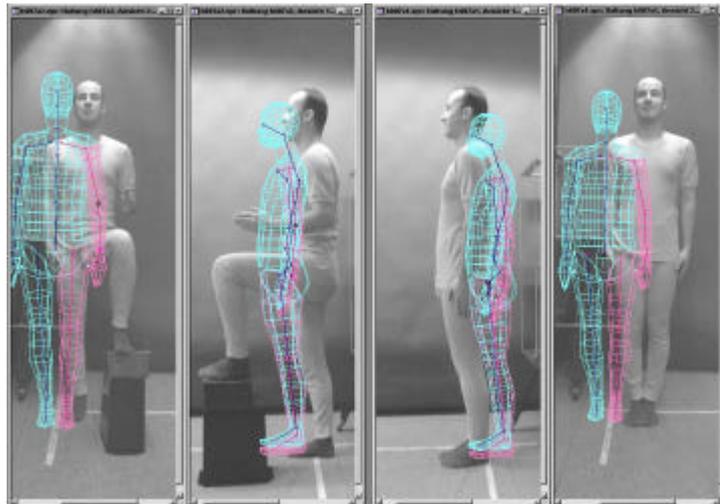


Abbildung 2.10: In die Bildpaare überlagertes Modell, das sich aber noch nicht mit dem Probanden deckt.

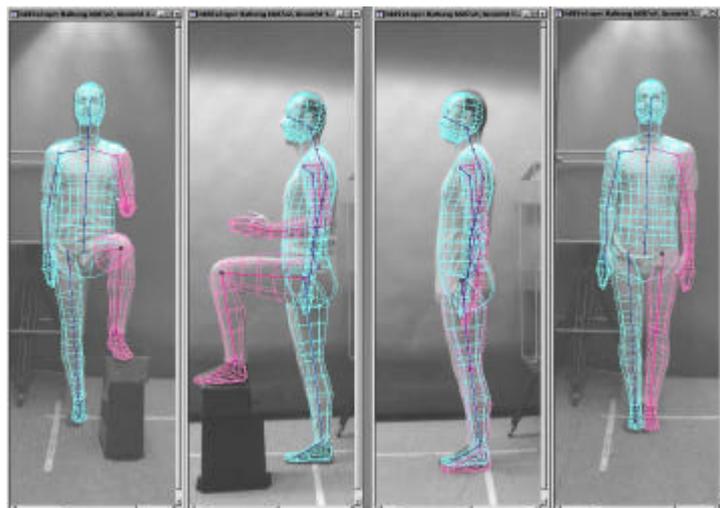


Abbildung 2.11: Das Modell wurde mit dem Probanden zur Deckung gebracht. Das Modell repräsentiert nun die Anthropometrie der Versuchsperson – ein virtueller Zwilling ist entstanden.

### 2.3.4 Haltungsanpassung

Das individuelle Modell, das bei der Anthropometrieanpassung entstand, kann in beliebige Bilder, welche die Versuchsperson zeigt, projiziert werden. Durch Veränderung der Gelenkwinkel nimmt das Modell die gleiche Haltung ein, wie die Versuchsperson und deckt sich optimal mit dieser. Dadurch wird die Haltung der Versuchsperson dreidimensional erfasst.

Die Haltungsdaten liegen konform zum verwendeten Modell vor. In Abbildung 2.12 ist das Prinzip der Haltungsmessung durch Überlagerung dargestellt.

Die Veränderung der Gelenkwinkel erfolgt, wie bei der Anthropometrieanpassung, manuell nach einem definierten Schema. So wird zunächst das Becken positioniert. Ausgehend von einer korrekten Beckenposition werden die Gelenkwinkel der Beine, des Torsos, des Kopfs und der Arme justiert. Diese Prozedur ist relativ einfach und schnell mit Hilfe einer ergonomischen Bedienung möglich, bei der der Benutzer mit der Maus Körperelemente anwählt und diese dann mit Hilfe der Maus verschieben kann, bis sich das entsprechende Körperelement mit der Fotografie deckt.

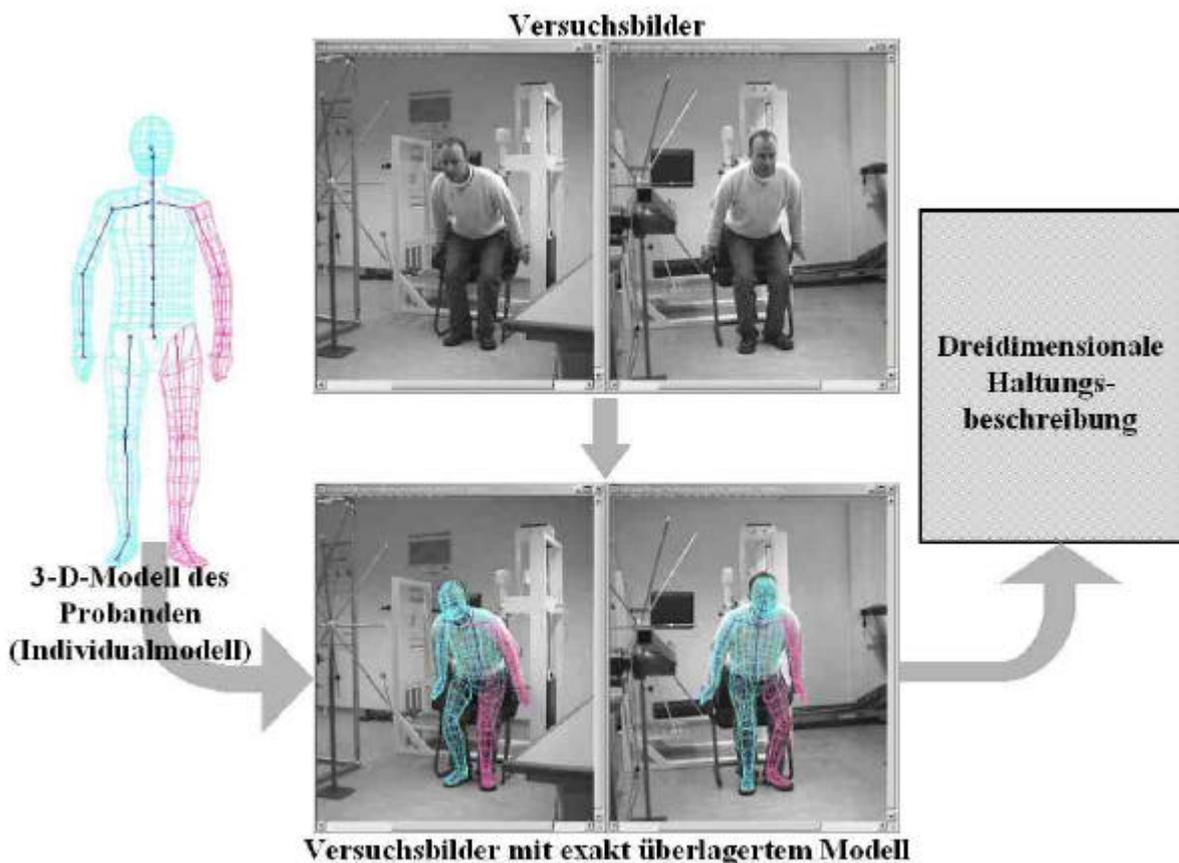


Abbildung 2.12: Prinzip der Haltungsmessung. Das individuelle Modell des Probanden repräsentiert dessen Anthropometrie und wurde zuvor erzeugt. Durch Überlagerung und entsprechende Veränderung der Gelenkwinkel wird es mit beliebigen Haltungen der Versuchsperson zur Deckung gebracht.

## 2.4 PCMAN als Basis für die Bewegungsmessung

Im letzten Kapitel wurde die Methodik vorgestellt, wie mit Hilfe eines exakten Menschmodells, die Haltung eines Probanden ermittelt wird. Das Verfahren lief manuell ab und ist für eine einzelne Haltung nicht sehr zeitintensiv. Für die Anpassung einer ganzen Serie von Bildern, wie sie durch einen Bewegungsfilm anfallen, ist jedoch die manuelle Anpassung nicht mehr vertretbar. Kann die Haltungsanpassung also automatisiert werden?

In Kapitel 1.2.4.3 wurde gezeigt, dass mit Hilfe von Markern automatisch die Haltung und durch zeitlich aufeinander folgende Haltungen auch die Bewegung eines Probanden gemessen und auf ein Modell übertragen werden kann. Bei den Überlegungen zum „Idealen Bewegungsmesssystem“ wurde festgestellt, dass Marker jedoch nicht optimal sind, da sie verrutschen und die Versuchsperson behindern können und ferner, der Aufwand des Anbringens hoch ist. Daraus resultierte die Überlegung auf Marker zu verzichten. Doch blieb die Frage offen, wie sonst eine Korrespondenz zwischen Modell und Versuchsperson hergestellt werden kann.

Das Messsystem PCMAN liefert mit Hilfe des Überlagerungsprinzips einen Lösungsansatz hierfür. Denn das überlagerte Menschmodell, definiert sehr genau jene Bildbereiche, die quasi unter dem Modell liegen und deshalb der Versuchsperson zuzuordnen sind und jene Bildbereiche die zum irrelevanten Hintergrund gehören. Die Überlagerung ermöglicht also, Bildbereichen Wissen zuzuordnen, wie dies in Abbildung 2.13 schematisch gezeigt ist. Ein Bild, das zuvor nur aus aneinandergereihten mehr oder weniger farbigen Bildpunkten bestand, wird dadurch bedeutungsvoll. Bildbereiche und deren strukturellen Merkmale können anatomisch interpretiert werden. Der Bildbereich „Hand“ kann vom Bereich „Kopf“ unterschieden werden, die Arme von den Beinen usw., da die Bildmuster und die Größe der Bereiche unterschiedlich sind. Die weiß umrandeten Bereiche in Abbildung 2.13 sollen dies verdeutlichen. Die Übertragung des Modellwissens auf Bildbereiche bedeutet auch, dass Bereiche in Folgebildern wieder gefunden werden und die Verschiebung gegenüber einem Bild vorher bestimmt werden könnte, sofern sich die strukturellen Merkmale der Bildbereiche nicht sehr verändert haben

Der Mensch hat keine Probleme oben geschilderte Aufgabe des Wiederfindens von Bildbereichen zu erfüllen. So erkennt ein Betrachter einer Bildsequenz sehr schnell, *welche* Objekte sich *wohin* bewegt haben. Auch die Interpretation der Objekte fällt leicht. So erkennt man in der Regel ohne weiteres wohin sich z.B. der Kopf, die Arme oder Beine eines Probanden bewegt haben. Für diese Bildinterpretation benötigt der Mensch keinerlei Marker an der Ver-

suchsperson. Die Natur hat uns mit einem *visuelles System* ausgestattet, mit dessen Hilfe wir derartige und viele andere Bewegungen ad hoc „messen“ können.

Die Natur nutzt im Wesentlichen zwei Informationsquellen zur Bewegungsdetektion: Bildkorrelation und Modellwissen. Bei der Korrelation werden Bildbereiche miteinander verglichen und Verschiebungen bestimmt. Die Verknüpfung mit Modellwissen stützt die Korrelation und ermöglicht Falschinformation zu unterdrücken. Im folgenden Kapitel werden diese Mechanismen, die letztlich bei der Realisierung des hier vorzustellenden Bewegungsmesssystems relevant sind, näher beschrieben, sowie, im übernächsten Kapitel, die technische Umsetzung mit Bildverarbeitungsalgorithmen vorgestellt.

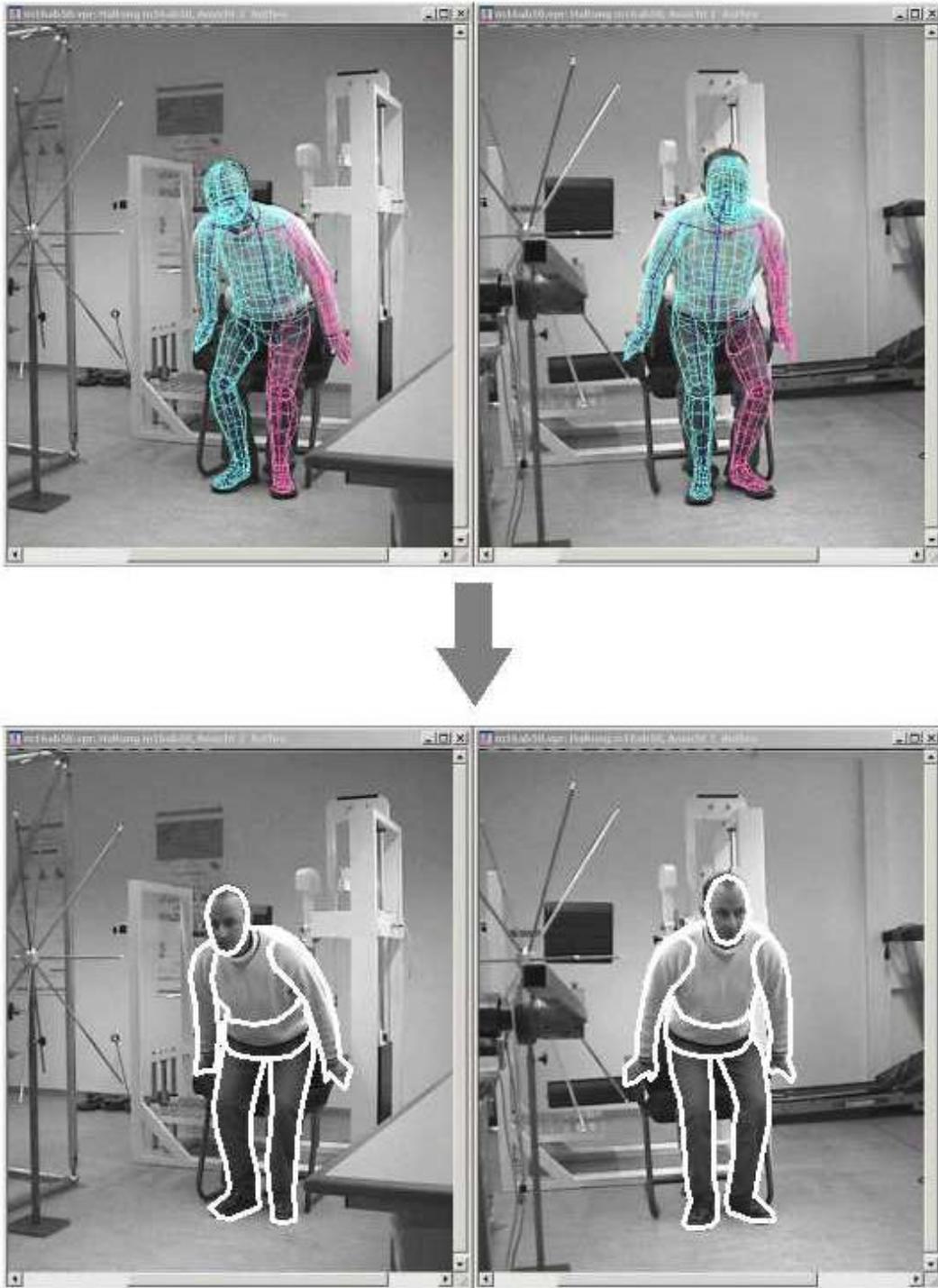


Abbildung 2.13: Durch die Überlagerung des Menschmodells in die Ansichten eines Experiments kann das Modellwissen in die Bildinformation übertragen werden. Klar umrissene Bildbereiche können Körperteilen, wie Kopf, Arme, Beine, usw. zugeordnet werden.

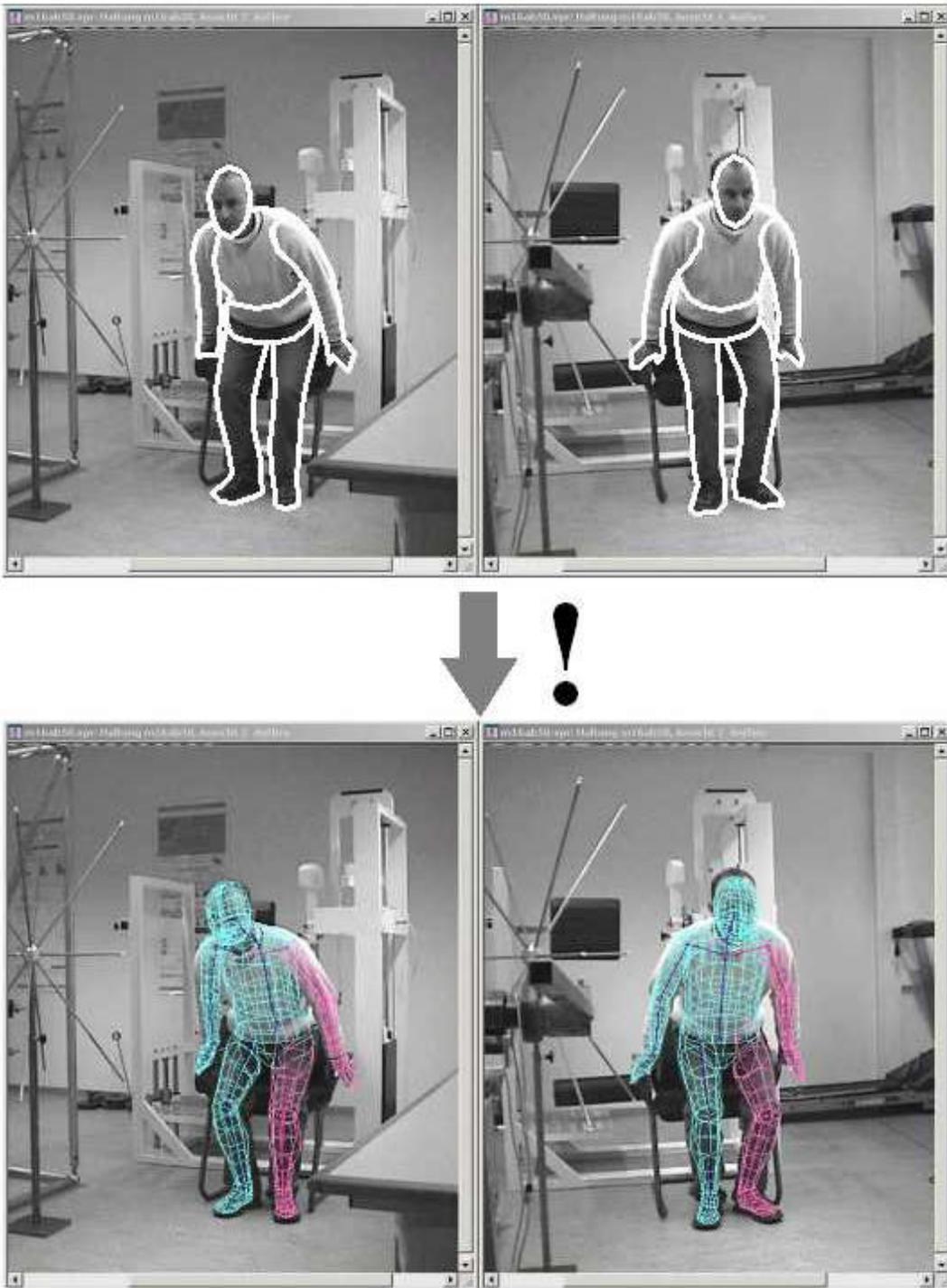


Abbildung 2.14: Bei der Bewegungsmessung stellt sich das Problem, dass zum einen Bildbereiche detektiert werden (oben) und diese dann korrespondierenden Körperelementen und Haltungswinkeln zugeordnet werden müssen.

## **3 Methodik**

### **3.1 Psychologisch-physiologische Mechanismen der Bewegungs-erkennung**

Der Mensch ist mit einem ausgezeichneten visuellen System ausgestattet, das ihm erlaubt, Objekte und deren Bewegung im Raum schnell und sicher zu erkennen. Dabei kann ein Objekt Änderungen in Ort und Zeit unterliegen. Dies bedeutet, dass der Beobachter ein Objekt aus unterschiedlichen Perspektiven während einer Bewegung sieht oder das Objekt aufgrund der Bewegung kleiner oder größer wird oder Farbänderungen unterliegt. Trotz solcher Veränderungen der Erscheinungsform interpretiert das menschliche visuelle System dies als Bewegung ein und desselben Objekts.

Die Mechanismen, welche die Natur hierzu entwickelt hat, werden im Folgenden vorgestellt und beschrieben. Nachdem zunächst allgemein die Bildinterpretation, bezogen auf statische Bilder, diskutiert wird, erfolgt darauf aufbauend der Übergang zur Erkennung von bewegten Objekten.

#### **3.1.1 Neuronale Bildinterpretation**

Durch die Umsetzung von Umweltreizen ist es dem Menschen möglich äußere Reize zu erkennen, zu verarbeiten und darauf zu reagieren. Dabei werden physikalische Umweltreize in physiologische umgesetzt. Für diese Reizumsetzung stehen Rezeptoren zur Verfügung in denen elektrische Potenziale durch Änderung der Konzentration von Kationen (Kalium- und Natriumionen) und Anionen (z.B. Chlorionen) zur Reizweiterleitung entlang von sog. Axonen über Synapsen an Neuronenzellen führen. Wobei hierfür an der Membran der Rezeptorzelle durch Depolarisation eine bestimmte Schwelle überschritten werden muss. Erst dann liegt das sog. Aktionspotenzial an und die Zelle „feuert“. Innerhalb des Nervensystems können die Signale über weitere Neuronen weitergeleitet und verarbeitet werden.

Für die Detektion visueller Reize stehen im Auge Fotorezeptoren zur Verfügung. Es umfasst ca. 125 Millionen Fotorezeptoren an der Rückseite der Retina. Das Grundprinzip des retinalen Informationsflusses ist in Abbildung 3.1 gezeigt. Der direkteste Weg der visuellen Information ist von Fotorezeptoren durch Bipolarzellen zu den Ganglionzellen. Die Bipolarzellen werden in zwei Kategorien unterteilt, dies sind die ON und die OFF Bipolarzellen. ON bedeutet, dass die Zelle ein Signal gibt, wenn Licht auf den Rezeptor fällt; OFF bedeutet entsprechend, dass die Zelle ein Signal liefert, wenn kein Licht auf den Rezeptor fällt.

Die Ganglionzellen feuern entsprechend des empfundenen Lichts. Diese Impulse werden über den Sehnerv zum Gehirn geleitet. Neben diesem direkten Weg zum Gehirn, kann die retinale Informationsverarbeitung durch zwei weitere Zelltypen beeinflusst werden. Horizontal angeordnete Zellen empfangen Signale von den Rezeptoren und feuern an die Bipolarzellen bzw. an die Ganglionzellen. Durch diese Verschaltung ist es möglich, Signale umliegender Rezeptoren zu berücksichtigen.

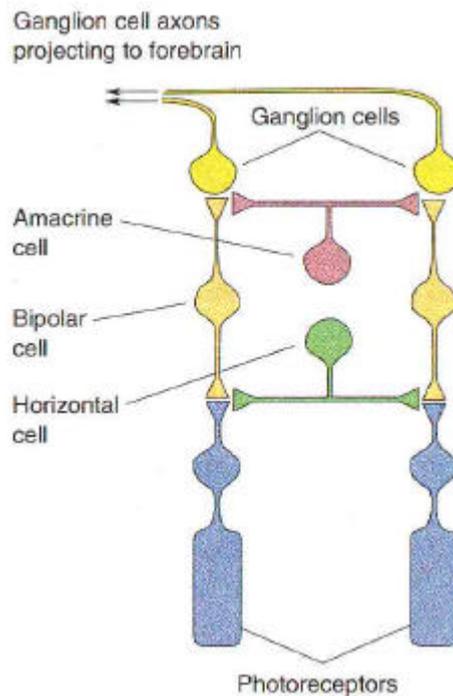


Abbildung 3.1: Grundsystem des Informationsflusses in der Retina (aus Bear, Connors und Paradiso [7]).

Die meisten Rezeptoren sind durch horizontale und bipolare Zellen verschaltet, so dass sich ein Zentrum aus einigen Rezeptoren ergibt, das konzentrisch durch weitere Rezeptoren umgeben ist. Diese Bereiche werden rezeptive Felder genannt. Am Ende jedes rezeptiven Feldes ist eine Ganglionzelle, welche die verschalteten Signale zum Gehirn weiterleitet. Bewegt sich z.B. eine hell-dunkel Kante über diesen Bereich, so liefert eine OFF-Ganglionzelle kein Signal, wenn der dunkle Bereich nur die Umgebung abschattet, nicht jedoch das Zentrum des rezeptiven Feldes. Wandert der hell-dunkel Bereich weiter und bedeckt das Zentrum, so liefert die Ganglionzelle ein starkes Signal, d.h. ein Signal mit höchster Frequenz, wie Abbildung 3.2 c) zeigt. Bedeckt der dunkle Bereich die gesamte Fläche, so liefert die Ganglionzelle ein mäßiges Signal. Entscheidend bei dieser Verschaltung ist, dass hiermit bereits in der Retina hell-dunkel-Unterschiede, gegenüber absoluten Farb- oder Grauwerten neuronal

verstärkt werden (Abbildung 3.2), wenn das Zentrum des rezeptiven Feldes komplett und die Umgebung nur zum Teil bedeckt ist. Unser visuelles System ist also darauf spezialisiert lokale Helligkeitsunterschiede zu detektieren.

Die eben gezeigte einfache Kantendetektion durch rezeptive Felder wird durch weitere Verschaltungen im visuellen Kortex, also jenem Bereich des Gehirns, an den der Reiz vom Auge weitergeleitet wird. Auf mehreren Ebenen werden dort optische Signale benachbarter rezeptiver Felder durch spezialisierte Zellen weiterverarbeitet. Mit Hilfe dieser komplexen und hyperkomplexen Zellen realisiert die Natur neuronale Kantendetektoren, Detektoren für die Länge einer Kante, Detektoren für die Orientierung von Kanten (Abbildung 3.3) und Bewegungsdetektoren [9].

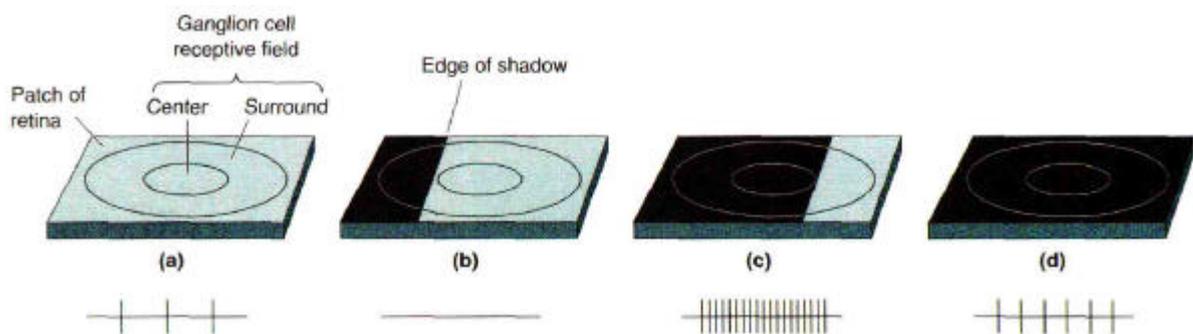


Abbildung 3.2: Ganglionsignal in Abhängigkeit der Lage eines hell-dunkel-Unterschieds. Bei der Lage, wie in c) gezeigt, liefert die Zelle das stärkste Signal, d.h. das Ganglion feuert mit höchster Frequenz. Das jeweils resultierende Aktionspotenzial ist unterhalb der Bilder dargestellt (aus [7]).

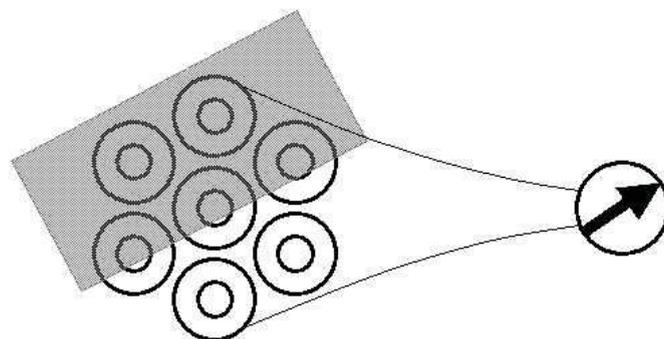


Abbildung 3.3: Mehrere rezeptive Bereiche, die zusammenschaltet sind, können die Orientierung einer Kante messen, wie dies durch den Pfeil rechts symbolisiert ist. Dabei liefert eine entsprechend spezialisierte Zelle, ein Signal, das proportional zum Winkel der Kante ist.

### 3.1.2 Interpretation von Bewegung

Die Empfindung von Bewegung wird in einem eigenen neuronalen Bereich des visuellen Systems bearbeitet. Form, Farbe und Bewegung werden in räumlich separierten Arealen des visuellen Kortex bearbeitet. Die Verarbeitung visueller Information erfolgt in drei Ebenen. Die erste Ebene der Verarbeitung erfolgt auf der Retina, die zweite (einfache Zellen) und dritte (komplexe Zellen) im Kortex. Daneben fließen in höheren Ebenen der Verarbeitung Kontextinformationen hinzu.

Bewegung wird von der Retina als zeitlicher Unterschied zwischen dem Signalmuster einzelner rezeptiver Felder an den Kortex codiert. Im visuellen Kortex existieren hierauf spezialisierte Neuronen, die auf bestimmte Eigenschaften der Bewegung, wie Geschwindigkeit, Orientierung oder Richtung der Bewegung ansprechen. Ein Teil der Neuronen sind nur sensitiv auf Stimulierung verschiedener rezeptiver Felder zur selben Zeit, so dass hiermit die Bewegung eines (zusammenhängenden) Objekts als solches erkannt werden kann.

Im Folgenden wird das Problem behandelt, wie das visuelle System ein Objekt identifiziert und entscheidet, ob das Objekt sich aufgrund von Bewegung oder durch andere Umstände, z.B. Änderung der Beleuchtung, verändert. Dieses Problem wird im Allgemeinen als Korrespondenzproblem bezeichnet, wobei gezeigt wird, dass einfache Bewegungserkennung ein autonomer Prozess ist, der auf einer sehr niedrigen Verarbeitungsebene stattfindet. Mit Hilfe des Korrespondenzprozesses, kann das visuelle System die Verbindung zwischen der zeitlichen Entwicklung von Objekten herstellen, wie Ullmann [69] zeigte.

Das Hauptproblem, welches das visuelle System lösen muss, ist die Identifikation bestimmter Elemente, die in der zeitlichen Folge der eintreffenden Bilder, ein und das selbe physikalische Objekt repräsentieren; d.h. für ein gegebenes Element  $X$  in einem visuellen Feld  $A(t)$  zur Zeit  $t$  muss das entsprechende Gegenstück  $X'$  im visuellen Feld  $A(t')$  zum späteren Zeitpunkt  $t'$  gefunden werden. Die Art dieser Elemente, anhand derer ein zeitlicher Zusammenhang zwischen Objekten hergestellt wird, wird als Korrespondenzmerkmal bezeichnet. Es wird angenommen, dass das Bild in kleine Bereiche zerlegt wird, die in kleinen Gruppen organisiert sind. Auf dieser Ebene erfolgt die Herstellung von Korrespondenz. Die angesprochenen Gruppen oder Einheiten sind vermutlich hierarchisch organisiert, in der Form, dass Einheiten wie Kanten, Fragmente, Linien oder andere kleine Bildbereiche zuerst erkannt werden und nachher in größere Einheiten und schließlich zu Objekten gruppiert werden. Die Einheiten, welche die Korrespondenzmerkmale beinhalten, befinden sich hierarchisch auf einer sehr niedrigen Ebene, sie werden deshalb als „low-level units“ bezeichnet. Das visuelle System findet auf Wahrscheinlichkeitsbasis heraus, ob ein Element  $X$  identisch ist mit einem

Element  $X'$  zu einem späteren Zeitpunkt, wobei zunächst die Korrespondenz auf Basis kleinster Korrespondenzmerkmale gesucht wird. Später werden aus diesen kleinen Grundeinheiten größere Objekte zusammengesetzt. Die optischen Grundbausteine, wie sie auch schon im vorherigen Abschnitt beschrieben wurden, sind die Bausteine aller größeren Objekte und Figuren.

Anhand einfacher Beispiele zeigt Ullmann, dass das visuelle System durch Korrelation die Korrespondenz zwischen Grundeinheiten herstellt und Bewegung detektiert bzw. interpretiert, wobei es sich bei der Interpretation streng an die naheliegendste Bildinterpretation hält, unabhängig davon, ob diese Interpretation im Widerspruch zu höher geordnetem Modellwissen steht.

In Abbildung 3.4 sind die Speichen eines Wagenrades in Form einer schwarzweißen Linienzeichnung dargestellt, wobei jene Speichen, die diagonal liegen, unterbrochen sind. Wird die Zeichnung um den Winkel  $\beta$  im Gegenuhrzeigersinn gedreht, resultiert die Stellung, die in Abbildung 3.4 b) gestrichelt dargestellt ist. Wird das Rad weiter um  $\beta$  gedreht und einem Probanden die Bilder so schnell präsentiert, dass dieser den Eindruck einer kontinuierlichen Bewegung hat, so nimmt der Proband nicht die Drehung eines einzelnen Objektes „Wagenrad“ wahr, sondern drei verschiedene Rotationen. Das Wagenrad erscheint in drei Ringe zerteilt, wobei sich der innerste und äußerste langsam im Uhrzeigersinn dreht, während der mittlere Ring sich schnell gegen den Uhrzeigersinn dreht. Dieser Eindruck, der dem tatsächlichen physikalischen Verhalten widerspricht, denn definitiv wurde das gesamte Objekt gedreht, ist mit Hilfe der angesprochenen Grundelemente, die das visuelle System zur Korrespondenzbildung nimmt, zu erklären. Ein solches Grundelement sei durch ein Liniensegment der Länge  $x$  gegeben, wie in Abbildung 3.4 b) eingezeichnet. Die neuronale Bildverarbeitung sucht dann das Liniensegment, das am besten zu dem davor gesehenen passt. Dabei werden Wahrscheinlichkeiten gebildet, wobei dasjenige Element, das am nächsten an der vorherigen Position liegt, die ähnlichste Orientierung und die ähnlichste Länge aufweist, als das wahrscheinlichste interpretiert wird. So kommt es, dass beim inneren und äußeren Ring jene Liniensegmente, in Korrespondenz gesetzt werden, die einer Drehung im Uhrzeigersinn entsprechen, während die Korrespondenz im mittleren Ring bei einer Drehung gegen den Uhrzeigersinn dem visuellen System am wahrscheinlichsten erscheint. Diese Bildinterpretation findet auf einer sehr niedrigen Ebene im Kortex statt und lässt sich durch Modellwissen, das der Proband hat (er weiß, dass das gesamte Bild gedreht wird und es ein zusammenhängendes Objekt ist) nicht beeinflussen.

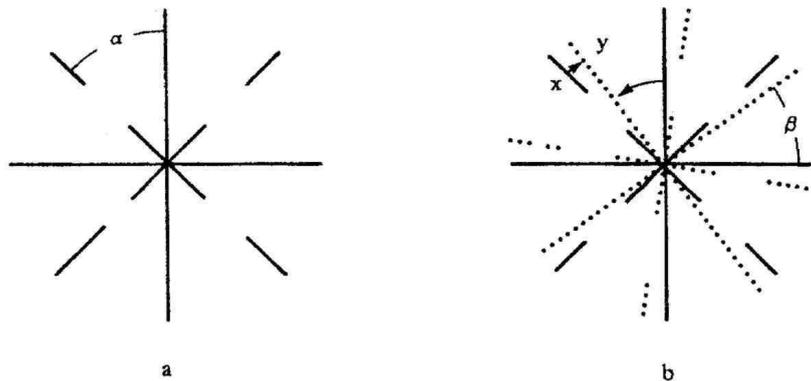


Abbildung 3.4: „Brocken Wheel“. Diese Abbildung erinnert an die Speichen eines Rads. Die Speichen werden [wie in b) angedeutet] jeweils um den Winkel  $\beta$  in Filmgeschwindigkeit im Uhrzeigersinn gedreht und einem Probanden präsentiert. Dadurch entsteht der Eindruck, als ob sich der äußere und innere Ring im Uhrzeigersinn drehen, während sich der mittlere Ring gegen den Uhrzeiger dreht.

Das folgende Beispiel lässt sich ebenfalls nur erklären durch die Annahme, dass die Korrespondenzbildung mit Hilfe einfachster geometrischer Grundelemente erfolgt. Abbildung 3.5 zeigt einen abstrahierten Zug mit Fenstern, den sog. „Block Train“. Der horizontale Abstand aller vertikalen Linien ist immer  $x$ . Wird der Zug kontinuierlich um die Strecke  $y$  nach rechts verschoben, entsteht der Eindruck, dass die Wagen des Zuges nach rechts laufen, die Fenster hingegen nach links. Dies ist wiederum dadurch zu erklären, dass durch die bewusst gewählte Verschiebung um  $y$ , jede vertikale Linie näher an einer Linie rechts von ihr, als links von ihr ist. So erscheint es dem visuellen System wahrscheinlicher, dass die kurzen vertikalen Linien sich nach links bewegen, während alle Stoß- und Kreuzungspunkte von Linien dichter an einem links liegenden Partner liegen, so dass diese nach rechts wandern.

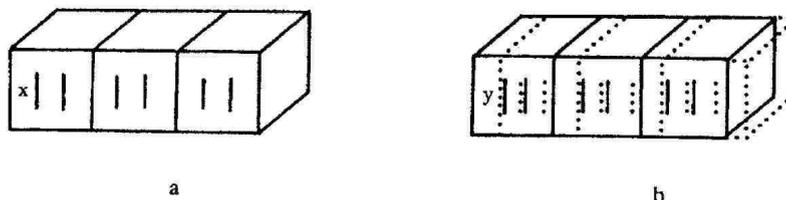


Abbildung 3.5: „Block Train“. Diese einfache geometrische Figur wird einem Probanden ebenfalls als animiertes Objekt präsentiert und dabei bei jedem Bild der Animation um die Strecke  $y$  nach rechts verschoben. Es entsteht der Eindruck, als ob sich der „Zug“ nach rechts bewegt, die „Fenster“ aber nach links.

Durch relativ einfache Versuche fand Ullmann heraus, inwieweit die Korrespondenzherstellung bei simplen Linien von Abstand, Länge oder Orientierung zueinander abhängt. So stellt sich z.B. auch die Frage, was geschieht, wenn dem visuellen System gleichwertige Lösungen präsentiert werden. Ein Beispiel hierzu zeigt Abbildung 3.6. Zunächst ist eine Linie zu sehen. Im Folgebild wird die Linie ersetzt und rechts und links der alten Position jeweils eine neue Linie gezeigt, die identisch zu der alten in Länge und Strichstärke ist. Der Abstand der rechten und linken Linie zur alten Linie ist gleich. Somit ist die linke Linie gleichberechtigt zur rechten und genau so wird es vom visuellen System interpretiert. Der Betrachter gewinnt den Eindruck, als ob sich die ursprüngliche Linie in eine Linie rechts und links teilt, die dann weg wandern. Wird jedoch sukzessive der Abstand, z.B. der rechten Linie zur Mittellinie erhöht, bricht dieser Eindruck irgendwann ab. Es entsteht der Eindruck, als ob nur noch eine Linie wandert, nämlich die rechte. Das bedeutet, dass das visuelle System nur zu solchen Merkmalen eine Korrespondenz herstellt, für die das mit einer gewissen Wahrscheinlichkeit gerechtfertigt ist. Scheinbar sinnlose Lösungen werden vom visuellen System unterdrückt.

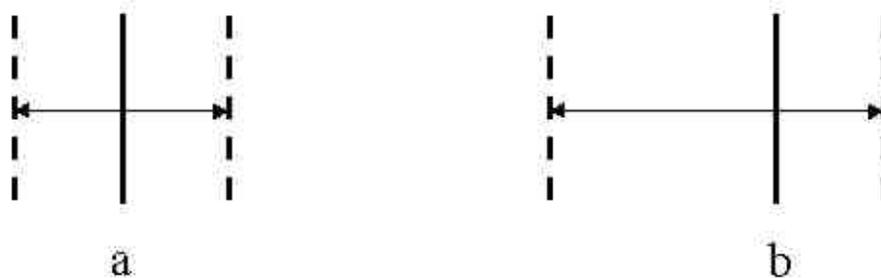


Abbildung 3.6: Korrespondenz zwischen zwei Merkmalen.

Die Korrespondenzerzeugung bei den gezeigten Liniensegmenten wird mit zunehmendem Orientierungsunterschied kleiner. In Abbildung 3.7 ist die Linie c gezeigt, die sich in zwei Linien zerteilt. Das Liniensegment l hat dabei dieselbe Orientierung, wie Linie c. Linie r hingegen weist eine andere Orientierung als die Ursprungslinie c auf. Helligkeit und Linienstärke sind wieder vergleichbar. Unter den gezeigten Bedingungen bevorzugt das visuelle System die Bewegung zum linken Linienelement l, wohingegen der Bewegungseindruck zum rechten unterdrückt wird. Dieser Effekt hängt vom Unterschied der Orientierung ab.

Bei Längenänderung der Liniensegmente und Beibehaltung der Orientierung wird ebenfalls eine Korrespondenz hergestellt, so dass ein Liniensegment in Bewegung gesehen wird und gleichzeitig zu schrumpfen/wachsen scheint. Wie oben, hängt auch hier die Leistung des visuellen Systems Korrespondenz herzustellen stark von den Bewegungs- und Längenunter-

schieden ab. Der Eindruck des Wachsens oder Schrumpfens bricht zusammen, wenn das Linienelement, das sich in der Länge ändert zu dicht am Ursprungselement ist. Es entsteht hierbei der Eindruck, als ob nur ein kleiner Teil des Ursprungselements aus diesem „ausbrechen“ würde.

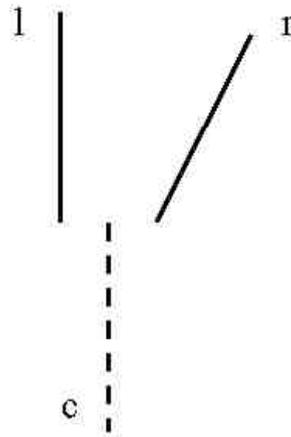


Abbildung 3.7: Korrespondenz zwischen Linien unterschiedlicher Orientierung.

Die gezeigten Beispiele zeigen, dass das visuelle System zum einen nur kleine Bildbereiche interpretiert und zum anderen vergleicht, wie ähnlich ein Objekt zum momentanen Zeitpunkt einem Objekt zum Zeitpunkt davor war. Wenn das Ähnlichkeitsmaß hoch ist, wird das Objekt als ein und dasselbe identifiziert. Bei konkurrierenden Interpretationsmöglichkeiten gewinnt das, mit dem größten Ähnlichkeitsmaß.

### 3.1.3 Kognitive Verarbeitung

Bisher wurde die reine sensorische Verarbeitung visueller Information betrachtet. Was geschieht jedoch zusätzlich aus psychologischer Sicht? Eingehende sensorische Informationen, wie oben beschrieben, werden in geeigneter Weise vom Gehirn kategorisiert und mit Informationen im Langzeitgedächtnis in Übereinstimmung gebracht. Im Wesentlichen ist dies ein Vorgang von Interpretationen und Zuordnungen von Wissensbeständen unserer Umwelt. Dieser Vorgang wird in der kognitiven Psychologie als Mustererkennung bezeichnet. Die Mustererkennung ist eine kognitive Leistung eines biologischen Systems, die sich an eine Stimulation anschließt. Die zugehörigen Mechanismen werden im Folgenden erläutert [74].

#### 3.1.3.1 Ikonisches Gedächtnis

Viele Umweltreize sind zeitlich überdauernd, was die Wahrnehmung stützt, denn diese braucht Zeit. Andere Stimuli sind sehr kurzlebig, sind aber nichtsdestotrotz ebenso von Be-

deutung für weitere Überlegungen oder weiteres Handeln. Beispielsweise verhilft ein kurzer Lichtblitz in der Dunkelheit dazu, sich in der Umwelt zu orientieren, obwohl der Reiz sehr kurz ist, kann der resultierende Eindruck praktisch unbegrenzt im Gedächtnis bleiben. Dieses und andere Beispiele zeigen, dass höhere Organismen, hier insbesondere der Mensch, in der Lage sind, auf kurzfristige Stimuli zu reagieren, deren Wahrnehmung nicht zeitgleich mit dem Reiz erfolgt. D.h. der Reiz wird so lange im Gedächtnis gehalten, bis eine Verarbeitung erfolgen kann. Im Allgemeinen spricht man bei dieser Art Gedächtnis vom „sensorischen Gedächtnis“; im speziellen Fall der visuellen Wahrnehmung vom „ikonischen Gedächtnis“. Das ikonische Gedächtnis kann neben physikalischen Informationen über Größe, Umfang, Farbe, Helligkeit und Bewegung auch kategoriale Informationen, z.B. handelt es sich beim Reiz um Buchstaben oder Zahlen, zum Stimulus einbeziehen. Bereits die Stäbchen und Zäpfchen leisten einen Beitrag zum sensorischen Gedächtnis (Sperling [66]).

Das Erkennen von Mustern involviert zahlreiche Prozesse. Es gibt psychologische und physiologische Untersuchungen, die nahe legen, dass Muster (z.B. Buchstaben) dadurch erkannt werden, dass elementare Merkmale, wie horizontale, vertikale und gebogene Linien detektiert werden. Allerdings deckt das Erkennen solcher Merkmale nicht den gesamten Bereich der Mustererkennung ab, deshalb können Reize auch ganzheitlich wahrgenommen werden. Hierbei wird der Stimulus mit bereits im Gedächtnis abgespeicherten typischen Repräsentanten des Reizes verglichen. Diese Repräsentanten werden als Prototypen bezeichnet. Der zentrale Prototyp beschreibt dabei den typischsten Vertreter einer Kategorie (z.B. Prototyp „Mensch“, Merkmale: hat Kopf, Arme, Beine, Torso, usw.). Atypische Vertreter weichen immer mehr vom typischen Vertreter ab, werden aber noch zur betreffenden Kategorie hinzugezählt. Diese Art Datenbank wird durch Alltagserfahrung stetig erweitert und ggf. neu spezifiziert. Ob bei der Mustererkennung der holistische Ansatz oder die Suche nach definierten Merkmalen dominiert hängt von Faktoren wie der Art der vorgegebenen Stimuli und der Trainiertheit ab.

### 3.1.3.2 *Kontextinformation*

Die Verarbeitung von Information basiert nicht allein auf der Analyse von Merkmalen oder Kategorisierung. In bestimmten Kontexten treten bestimmte Muster eher auf als andere. Mit welcher Wahrscheinlichkeit in bestimmten Zusammenhängen welche Muster auftreten wird durch Erfahrungen bestimmt. So erwartet man beim Lesen eines Satzes aufgrund des Zusammenhangs u.U. an einer bestimmten Stelle schon ein bestimmtes Wort, bevor man es tatsächlich anhand visueller Merkmale detektiert hat. Auch die Art und Weise, wie die Analy-

se eines Musters erfolgt, wird durch den Kontext bestimmt. So reicht es, beispielsweise ein Wort, das man beim Lesen erwartet, anhand globaler Merkmale, die schnell erfassbar sind, zu detektieren. Es kann auch der Fall eintreten, dass der Kontext dazu führt, dass nur bestimmte Merkmale erkannt werden müssen, andere hingegen vernachlässigbar sind. Der Kontext beeinflusst die Aspekte der sensorischen Daten, die wir verarbeiten, kann aber auch die Perzeption dieser Daten selbst determinieren.

### **3.1.4 Zusammenfassung**

Die psychologisch-physiologischen Mechanismen der Bewegungserkennung basieren zum einen auf klar determinierten neuronalen Verschaltungen, die z.T. bereits auf der Netzhaut des Auges ablaufen und nicht beeinflussbar sind. Zum anderen finden sich komplexere Verfahren, bei denen Gedächtnisinhalte, sog. Kontextinformationen, abgerufen werden und mit Reizen verglichen und kategorisiert werden. Technisch ausgedrückt, heißt das, dass Bewegung in einer unteren Verarbeitungsebene mit einfachen Bildverarbeitungsoperatoren detektiert und quantifiziert wird. In einer höheren Ebene, wird jedoch Modellwissen, also die Kontextinformation, einbezogen und die resultierenden Messwerte damit abgeglichen und im Sinne des gewählten Modells interpretiert. Deshalb sehen wir bei menschlichen Bewegungen nicht einfach unzusammenhängende Bewegungen, sondern sind in der Lage, die diversen Bewegungen einzelner Körperelemente (in eventuell unterschiedliche Richtungen) als die Bewegung eines Objektes mit mehreren Gliedern, die aber gekoppelt sind, zu interpretieren. Unser Modellwissen über den Menschen und dessen Bewegungsmuster, erlaubt solche Interpretationen. Fazit ist also, dass die Natur beim Bewegungssehen einfache Algorithmen zur Mustererkennung, die der Korrespondenzherstellung zwischen ähnlichen Objekten dient, mit grundsätzlichem, modellhaftem Wissen über Objekte kombiniert.

Dieses Analogon zum menschlichen visuellen System ist die Basis des realisierten Bewegungsmesssystems, dass wie das menschliche visuelle System, ein Bild zunächst anhand analytisch zugänglicher Merkmale interpretiert, die Merkmale dann jedoch im Sinne des Modellwissens mit Hilfe des Modells RAMSIS interpretiert.

## **3.2 Bildverarbeitung**

Im Folgenden wird die Methodik zur bildverarbeitungs-basierten Objektverfolgung betrachtet. Sie ist die Grundlage der technischen Umsetzung der natürlichen Prinzipien des Bewegungssehens, wie es in Kapitel 3.1 erläutert wurde.

### 3.2.1 Bildvorverarbeitung

Als Aufnahmegerät für die Aufzeichnung von Bewegungsfilmern wurden Videokameras benutzt. Die aufgezeichneten Filme wurden mit Hilfe einer Videokarte an einem Rechner digitalisiert und in synchrone Einzelbildpaare zerlegt. Da zum einen der Detektor, also die Kamera bzw. deren CCD-Chip, als auch das digitalisieren zu Störungen auf den Bildern führt, müssen die Daten, vor einer Weiterverarbeitung, aufbereitet werden. Dabei treten im Wesentlichen zwei Effekte auf, das Rauschen und der Interlace-Effekt.

#### 3.2.1.1 Digitalisierung

Entsprechend der Videonorm (PAL in Europa) liefern Videokameras einen Bildstrom mit einer Wiederholrate von 25 Halbbildern pro Sekunde. Mit Hilfe einer Video- oder Framegrabberkarte werden diese analogen Signale digitalisiert. Da es nur mit extrem teuren Geräten möglich ist, jedes digitalisierte Einzelbild in einem Rohdatenformat (wie BMP oder TIFF), schnell genug auf ein Speichermedium zu schreiben, wurde hier der übliche Umweg über das datenreduzierende AVI-Format gewählt. Somit werden zwar Bildverluste in Kauf genommen, dafür können aber auch längere Filme gehandhabt werden.

#### 3.2.1.2 Glättung

##### Rauschen

Im Allgemeinen sind Rohbilddaten mit Rauschen behaftet. Selbst modernste, gekühlte CCD-Kameras, wie sie in der Astronomie zum Einsatz kommen sind nicht frei von Rauschen. Üblicherweise besitzt ein CCD mehrere Rauschquellen in der Größenordnung von 0,01 – 0,001 Prozent seiner maximalen Grauwerte. Eine handelsübliche 8-Bit-Kamera mit 256 Graustufen besitzt demnach bei absoluter Dunkelheit ein Rauschen in der Größenordnung von rund 2,5 bis 0,25 ADU<sup>2</sup> [38]. Bei langen Belichtungszeiten wird vor allem durch thermische Anregung im CCD Rauschen erzeugt, dass sich zu den Grauwerten, die durch den eigentlichen Lichteinfall erzeugt werden, hinzuaddiert. Bei der weiteren Verarbeitung kommen zusätzliche Rauschquellen hinzu, wie z.B. das Digitalisiererauschen beim Übertragen eines Bildes vom Videoband in einen Rechner.

---

<sup>2</sup> ADU ist die Abkürzung für *Analog to Digital Unit* und ist die Einheit des Grauwerts, der nach der Digitalisierung jedem Pixel zugeordnet wird. Typischerweise liefern gängige A/D-Wandler 8-Bit-Werte.

Im Hinblick auf eine Untersuchung von feinen Bildmerkmalen und –strukturen erscheint eine Bildglättung unerlässlich. Die Beseitigung hochfrequenten Rauschens entspricht im Wesentlichen einer Tiefpassfilterung. Allerdings hat die Glättung auch zur Folge, dass feine Strukturen im Bild verwaschen werden und Kanten unscharf erscheinen. Ziel der Glättung muss es daher sein, Störungen zu unterdrücken, Kanten- und Strukturinformationen jedoch möglichst zu erhalten. Es existieren verschiedene Methoden zur Glättung von Bildern.

Die einfachste Methode zu Glätten besteht in der Mittelwertbildung. Ein Grauwert wird dabei durch den Mittelwert aller Grauwerte einer  $N \times N$  Nachbarschaft ersetzt. Mit solch einem Mittelwertoperator kann Rauschen sehr gut unterdrückt werden (Abbildung 3.8). Allerdings werden auch hohe Ortsfrequenzen abgeschwächt, was zu einem Verlust von Details und zu unscharfen Kanten führt. Die Größe der  $N \times N$  Nachbarschaft hat erheblichen Einfluss auf das Ergebnis. Neben dem klassischen Mittelwertfilter gibt es noch Binomialfilter. Diese enthalten als Maskenelemente die Werte der diskreten Binomialverteilung. Binomialfilter sind wesentlich bessere Glättungsfilter als Rechteckfilter, wie der Mittelwertfilter. Diese Filterart wird als gewichtete Mittelung bezeichnet. Der Binomialfilter ist im Übrigen eine Näherung der Gaußschen Filterung. Er besitzt fast dieselben Eigenschaften, ist aber leichter zu realisieren.

Leider verwischen alle linearen Filter klar getrennte Bildbereiche, was meistens nicht erwünscht ist. Die Mittelung wird ausgeführt, egal, ob die Grauwertunterschiede groß oder klein sind. Gerade bei Kanten findet man jedoch große Grauwertunterschiede, die erhalten bleiben sollen. Mit Hilfe adaptiver Filter, also Filter die ihre Glättung von lokalen Gradienten abhängig machen, kann Rauschen unterdrückt und können Kanten erhalten bleiben. Leider ist es äußerst schwierig Kriterien für die zu unterdrückenden und nicht zu unterdrückenden Diskontinuitäten aufzustellen bzw. sind die Rechenzeiten erheblich, so dass diese Filter hier keine Bedeutung haben.

Mit Mittelwert- und Binomialfilter lässt sich statistisches Rauschen gut unterdrücken. Um Fehlstellen im Bild zu eliminieren sind diese Filter weniger geeignet. Hier bietet sich der Medianfilter an, der einen Grauwert durch den Medianwert einer  $N \times N$  Nachbarschaft ersetzt. Allerdings ist diese Filterart aufgrund der nötigen Sortierung der Grauwerte, wesentlich rechenintensiver als lineare Filter.

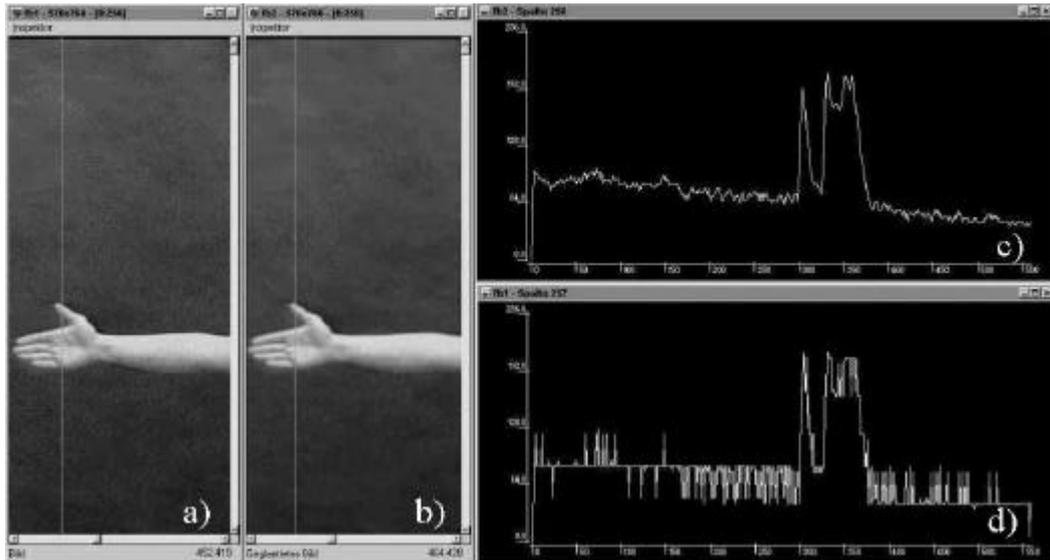


Abbildung 3.8: Rohbild und geglättetes Bild. Bild a) zeigt das Rohbild, Bild b) das durch Mittelwertfilter geglättete Bild. Bild c) die Grauwertverteilung entlang der hellen Linie von Bild b), Bild d) zeigt die Grauwertverteilung von Bild a). Deutlich ist die Unterdrückung des Rauschens in Diagramm c) zu sehen.

### Interlace-Effekt

Die Videotechnik arbeitet mit Halbbildern, d.h. anstatt das gesamte Bild zu belichten, wird dies nur für jede zweite Zeile des CCD-Chips getan, während die anderen Zeilen ausgelesen werden. Mit diesem Trick behalf man sich bei der Etablierung von Fernseh- und Videonormen, da die damalige Technik nicht schnell genug war, um Vollbilder auszulesen. Da sich die Videonorm nicht veränderte, ist diese Technik noch heute in allen Videokameras zu finden. Die Probleme, die sich mit der Halbbildtechnik ergeben bleiben dem Betrachter im Normalfall verborgen. Doch bei der Betrachtung eines Einzelbildes, auf dem eine schnelle horizontale Bewegung zu sehen ist, tritt der Effekt klar hervor. Die Kanten des sich bewegenden Objekts sind in jeder zweiten Bildzeile horizontal versetzt. Die Kante wirkt unscharf und verfranzt. Dieser Effekt wird als Interlace-Effekt bezeichnet. Zur Beseitigung kann man das ganze Bild glätten, was allerdings zum Verlust feiner Strukturen führt oder man eliminiert eins der Halbbilder und ersetzt die gelöschten Zeilen durch eine Kopie der darüber liegenden Zeilen. Dadurch verringert sich allerdings die vertikale Auflösung des Bildes. Kanten treten dadurch allerdings weiterhin scharf hervor.

Kanten bewegter Objekte werden nicht nur durch den Interlace-Effekt verschmiert, sondern auch durch eine zu lange Shutterzeit. Deshalb muss für eine erfolgreiche Interlace-Effekt-Beseitigung auch eine kurze Shutterzeit gewählt werden, so dass sich keine Bewegungsunschärfe ergibt.



Abbildung 3.9: Beseitigung des Interlace-Effekts. Das linke Bild zeigt die durch das Interlacing verwaschene Kontur des bewegten Arms. Rechts wurde jede ungerade Bildzeile durch eine Kopie der darüber liegenden Zeile ersetzt. Dadurch werden die Kanten schärfer. Die vertikale Bildauflösung reduziert sich allerdings um Faktor zwei.

### 3.2.2 Bewegungsmessung

Bei der Bewegungsmessung muss ein definierter Bildbereich zuverlässig im Folgebild gefunden werden. Als Randbedingung gilt dabei, dass die Veränderung des Bildbereichs nicht so groß sein darf, dass dieses Wiederfinden nicht ermöglicht wird, d.h. Bild-zu-Bild-Unterschiede sollen klein sein.

Zum Wiederfinden von Bildmustern bieten sich zwei Methoden an. Das sind zum einen die Differenzbildung, wobei die Differenz Null ist, wenn zwei Bildregionen gleich sind, und zum anderen die Korrelationsmethode, die gegenüber der Differenzbildung etwas sensitiver ist. Beide Verfahren wurden implementiert und zeigen befriedigende Ergebnisse.

### 3.2.2.1 Differenzbildung (Template-Matching)

Die Differenzbildung ist eine der einfachsten Methoden, um identische Bildmuster zu detektieren. Dabei wird die Differenz zwischen Bildbereichen gebildet und bewertet. Wenn  $I_1(i,j)$  und  $I_2(k,l)$  Grauwerte eines Bildes sind, wobei  $i, j, k$  und  $l$  den Bildpunkt in x- und y-Richtung bezeichnen und  $(2M + 1) \times (2N + 1)$  die Größe des Suchbereichs um  $I_1(i,j)$  und  $I_2(k,l)$  ist, wobei  $M$  und  $N$  natürliche Zahlen sind, dann beschreibt

$$d(k,l) = \sum_{u=-M}^M \sum_{v=-N}^N |I_1(i+u, j+v) - I_2(k+u, l+v)| \quad (3.1)$$

die Differenz  $d$  zwischen den Bildbereichen um  $I_1$  und  $I_2$ . Wird  $d(k,l)$  Null, so heben sich die Grauwerte der Umgebung von  $I_1$  und  $I_2$  gerade auf, das heißt, der Bereich um  $I_1$  sieht genauso aus wie der um  $I_2$ . In Bildfolgen, die frei von Bildstörungen, etwa durch Rauschen, oder Formveränderungen sind, ist die Differenzmethode eine gute und schnelle Möglichkeit, identische Bildbereiche zu finden.

Nachteil der oben beschriebenen Differenzmethode ist allerdings, dass sie relativ unsensitiv ist, wenn Bildmuster nicht identisch sind sondern sich nur ähneln. Das sich Bildmuster nur ähneln ist jedoch aufgrund von Bildstörungen der häufigste Fall. Zur Steigerung der Sensitivität bietet sich die Korrelationsmethode an.

### 3.2.2.2 Bewegungsmessung durch Kreuz-Korrelation

Korrelationsverfahren sind gut geeignet, um Bildmerkmale wieder zu finden, insbesondere wenn sich diese in Struktur und Farbe nicht verändern, sondern nur im Ort. Damit kann mit Hilfe der Korrelation die Korrespondenz zwischen Bildmerkmalen in zeitlich aufeinander folgenden Bildern hergestellt werden.

Um in Bildfolgen herauszufinden welcher Bildpunkt  $(i,j)$  eines Bildes mit dem entsprechenden Bildpunkt  $(k,l)$  des Folgebildes korrespondiert, wird jede mögliche Paarung mit einer Korrelationsfunktion bewertet. Als Zuordnung wird jener Bildpunkt gewählt, der den optimalen Korrelationswert aufweist. Da einzelne Bildpunkte in der Regel zu wenig Informationen bieten, wird die Nachbarschaft um einen Punkt miteinander verglichen. Bei rechteckigen Nachbarschaften der Größe  $(2M + 1) \times (2N + 1)$  ist

$$c(k, l) = \frac{1}{K} \sum_{u=-M}^M \sum_{v=-N}^N [(I_1(i+u, j+v) - \bar{I}_1(i, j)) \times (I_2(k+u, l+v) - \bar{I}_2(k, l))] \quad (3.2)$$

mit

$$K = K_1 K_2 \quad (3.3)$$

$$K_1 = \sqrt{\sum_{u=-M}^M \sum_{v=-N}^N (I_1(i+u, j+v) - \bar{I}_1(i, j))^2} \quad (3.4)$$

$$K_2 = \sqrt{\sum_{u=-M}^M \sum_{v=-N}^N (I_2(k+u, l+v) - \bar{I}_2(k, l))^2} \quad (3.5)$$

wobei  $c(k, l)$  als Kreuz-Korrelationskoeffizient bezeichnet wird (Jiang und Bunke [31]). Der Index 1 bezeichnet das aktuelle Bild und 2 das Folgebild.  $\bar{I}_1(i, j)$  und  $\bar{I}_2(k, l)$  beschreiben den Durchschnittsgrauwert der Nachbarschaft um den Punkt  $(i, j)$ . Durch den Normierungsfaktor  $K$  liegen die Werte von  $C$  zwischen null und eins. Eins bedeutet optimale und null sehr schlechte Korrelation.

Wird mit dieser Methode die Bewegung eines Menschen verfolgt, so muss bedacht werden, dass Änderungen von Bild zu Bild das oben genannte Kriterium, dass nämlich Struktur und Farbe der Nachbarschaft erhalten bleibt, in der Regel nicht gilt. Da normalerweise die Aufnahme einer Bewegung mit Kameras erfolgt, die eine konstante Anzahl Bilder pro Sekunde erzeugt, hängt die Möglichkeit Korrespondenzen zu Bildpunkten herzustellen stark von der Geschwindigkeit der Bewegung im Verhältnis zu der Aufnahmegeschwindigkeit der Kamera ab, da Struktur bzw. Farbänderungen um so wahrscheinlicher sind, um so größer der Bewegungsunterschied zwischen zwei Bildern ist. Idealerweise sollte das Objekt, das wieder erkannt werden soll, nur minimalen Veränderungen von Bild zu Bild unterliegen, so dass die wesentlichen Erkennungsmerkmale unangetastet bleiben.

Ein wesentliches Problem bei der Korrespondenzherstellung durch Korrelation ist die optimale Wahl der Größe der Nachbarschaft und die Bewertung, ob die Nachbarschaft ausreichende Strukturmerkmale bietet, um identifiziert werden zu können. Wegen des letzten Punktes werden bis heute in der Regel Marker am Probanden befestigt. Diese Marker haben ein definiertes Erscheinungsbild, das sich mit der Zeit nicht ändert. Bei Verwendung körpereigener Merkmale, wie das hier angestrebt ist, ist jedoch damit zu rechnen, dass ein gewähltes Merkmal strukturellen Veränderungen unterliegt. Zur Bewertung einer Nachbarschaft kann

z.B. deren Grauwertvarianz benutzt werden. Für die Korrespondenzherstellung eignen sich inhomogene Nachbarschaften, während homogene ungeeignet sind.

Eng mit diesem Problem ist die Wahl der richtigen Größe der Nachbarschaft verbunden. Eine kleine Nachbarschaft enthält oft zu wenig Strukturmerkmale erlaubt aber die Verfolgung kleiner Details. Eine Große Nachbarschaft enthält genügend Strukturmerkmale, löst aber Details nicht ausreichend auf. Bei der Verfolgung von Körperelementen ergibt sich die Größe der Nachbarschaft jedoch automatisch durch die scheinbare Größe des zu beobachtenden Körperelementes. Soll beispielsweise die Position einer Hand verfolgt werden, darf die Größe der Nachbarschaft den Bildbereich den die Hand abdeckt nicht wesentlich überschreiten. Abbildung 3.10 zeigt einen Vergleich zwischen der Analyse einer Handbewegung mit zu großer Nachbarschaft und optimaler Nachbarschaft. Die linke Bildspalte zeigt von oben nach unten die Bewegung, wobei das betrachtete Pixel mit einem Kreuz markiert ist. Um dieses zentrale Merkmal liegt die Nachbarschaft, die im linken Beispiel  $71 \times 71$  Pixel groß ist, im rechten  $31 \times 31$  Pixel. Da die Nachbarschaftsgröße links Strukturmerkmale der Hand und des Hintergrunds enthält, ist die Bewegungsanalyse sehr instabil Die Strukturmerkmale des Hintergrunds haben ein entsprechend hohes Gewicht, so dass von Bild zu Bild die detektierte Pixelposition weiter in den Hintergrund wandert. Nach wenigen Bildern wird die Position der Hand nicht mehr korrekt detektiert. Je nachdem, wie der Hintergrund beschaffen ist, wirkt sich sein Einfluss unterschiedlich aus. Da im gezeigten Beispiel der Hintergrund relativ homogen ist, verliert der Algorithmus die Hand nicht komplett, jedoch, wie im dritten Bild von oben zu sehen, resultiert ein großer Offset zur wahren Position. Um den Einfluss des Hintergrunds bzw. von Nicht-Objekt-Merkmalen zu eliminieren, muss die Nachbarschaft ans Objekt adaptiert werden. Die rechte Bildspalte zeigt die ideale Bewegungsverfolgung. Die Nachbarschaft ist gerade so groß gewählt, dass möglichst viele Strukturmerkmale der Hand in ihr liegen. Die Bildreihe zeigt, mit welcher Stabilität hier eine Bewegungsverfolgung realisiert wird. Die dargestellte Bildfolge umfasst ca. 50 Bilder, dies entspricht 2 Sekunden Film.

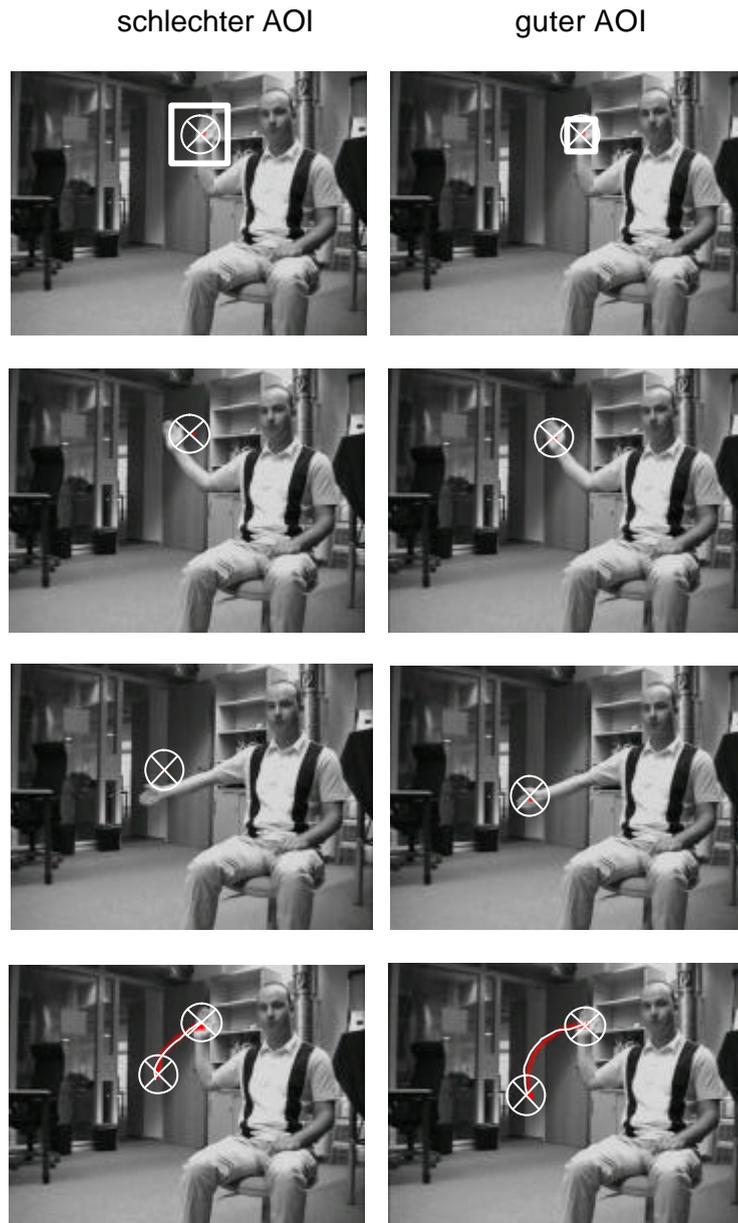


Abbildung 3.10: Einfluss der Größe der Nachbarschaft (Quadrat) auf die Korrelationsmethode. Die linke Bildspalte zeigt die detektierte Position der Hand bei Benutzung einer Nachbarschaft, die zu groß gewählt ist. Die rechte Bildspalte zeigt die gleiche Analyse, bei angepasster Größe der Nachbarschaft. Die zwei mittleren Reihen sind jeweils zeitliche Schnappschüsse aus der gesamten Sequenz und zeigen die dort detektierte Position der Hand. In der linken Spalte stimmt die detektierte Position mit der wahren nicht überein. Die untersten Bilder zeigen jeweils die resultierenden Bewegungsbahnen mit Start und Endpunkt. Die Startposition war jeweils identisch.

### 3.2.3 Kantenbilder

Zur Analyse von Bewegungen bieten sich auch Kantenbilder an. Die Konturen einer Versuchsperson können mit einem Modell verglichen werden, so dass die Haltung bestimmt werden kann. Kantenbilder sind die binäre Darstellung von segmentierten Gradientenbildern. Wird ein Gradienten- bzw. Ableitungsfilter

$$\nabla = \frac{\partial}{\partial x} + \frac{\partial}{\partial y} \quad (3.6)$$

auf ein Grauwertbild angewendet, treten alle Regionen, in denen hohe Grauwertunterschiede auftreten, hervor. Da meistens Grauwertunterschiede daher rühren, dass Übergänge zwischen Objekten stattfinden, dient die Gradientenfilterung zur Hervorhebung von Objektgrenzen, in diesem Fall also der Abgrenzung zwischen Proband und Hintergrund (Jähne [32]).

Die einfachste diskrete Realisierung eines Ableitungsoperators ist der Differenzoperator

$$D_x = \frac{1}{2} [1 \quad 0 \quad -1] \text{ und } D_y = \frac{1}{2} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \quad (3.7)$$

der hauptsächlich Kanten, die senkrecht zur Operatorrichtung stehen, detektiert. Unabhängig von der Orientierung einer Kante arbeitet (der hier verwendete) optimierte Sobeloperator,

$$S_x = \begin{bmatrix} 3 & 0 & -3 \\ 10 & 0 & -10 \\ 3 & 0 & -3 \end{bmatrix} \text{ und } S_y = \begin{bmatrix} 3 & 10 & 3 \\ 0 & 0 & 0 \\ -3 & -10 & -3 \end{bmatrix}, \quad (3.8)$$

der zusätzlich senkrecht zur Ableitungsrichtung eine Bildmittelung durchführt (Scharr [56]). Durch eine Kantendarstellung werden die Konturen der Versuchsperson sichtbar gemacht. Allerdings zeichnen sich auch viele andere Konturen des Hintergrunds im Kantenbild ab (vgl. Abbildung 3.11).



Abbildung 3.11: Links: Rohbild. Rechts: Kantenbild.

### 3.2.4 Zusammenfassung

Zur technischen Nachahmung der Korrespondenzherstellung zwischen Bildmerkmalen bewegter Objekte, wie es durch das visuelle System des Menschen, geleistet wird, kommen von algorithmischer Seite im Wesentlichen zwei Methoden in Frage. Es sind dies die gezeigte Differenzmethode und die Kreuz-Korrelation. Bei beiden Methoden ist zu beachten, dass die Größe des Bildbereichs, der das zu verfolgende Objekt zeigt, richtig gewählt wird. Ferner müssen die Bildmuster ausreichend prägnant sein, so dass sie wieder gefunden werden können. Letztlich erlauben diese Methoden auch nur sehr geringe Änderungen in der Struktur der Nachbarschaft. Bildänderungen müssen stetig und langsam ablaufen. Prinzipiell sind dies die gleichen Einschränkungen, die das menschliche visuelle System auch hat. Da der Rechenaufwand für die Differenzmethode geringer ist, ist sie schneller als die Korrelationsmethode. Allerdings ist die Korrelationsmethode sensitiver.

Zur Hervorhebung von Kanteninformation existieren Standardfilter, die hier zum Einsatz kommen. Somit kann eine weitere wesentliche Funktion des menschlichen visuellen Systems, die bei der Bildinterpretation eine Rolle spielt, mitberücksichtigt werden.

Um die Verfahren zuverlässiger anwenden zu können, müssen die aufgezeichneten Videobilder zunächst einer Vorverarbeitung unterzogen werden. Standard hierbei ist zunächst das Glätten der Bilder, wodurch hochfrequentes Rauschen unterdrückt wird (Hochpassfilter). Da die gängige Videotechnik mit Halbbildern arbeitet, ist es ferner nötig den sog. „Interlace-Effekt“ zu beseitigen. Dies reduziert jedoch die vertikale Auflösung der Bilder um den Faktor zwei.

## 4 Realisierung

Der realisierte Ansatz für ein Haltungs- und Bewegungsmesssystem kombiniert die Vorteile der Benutzung eines detaillierten, anthropometrisch korrekten und dreidimensionalen Menschmodells mit Techniken der digitalen Bildverarbeitung, so dass die Messung menschlicher Haltungen ohne Anbringung externer Marker unter bestimmten Randbedingungen möglich wird.

Dabei bietet das System die Möglichkeit zum einen reine zweidimensionale Bewegungsdaten zu erheben, die frei von Modellinformationen sind und deshalb mit geringerem Aufwand erhoben werden können. Zum anderen können für modellbasierte Untersuchungen dreidimensionale Daten erhoben werden, die an das Menschmodell RAMSIS gekoppelt sind.

### 4.1 Zweidimensionale Bewegungsmessung ohne Modell

Bei vielen Fragestellungen in der Ergonomie ist es nicht notwendig eine dreidimensionale Repräsentation eines Bewegungsablaufs zu messen. Weiterhin ist es nicht immer nötig, dass Bewegungsdaten modellkonform vorliegen. Vielmehr ist es häufig wichtig schnell und unkompliziert Daten zu ermitteln, die den prinzipiellen Vergleich und Aussagen über Bewegungen ermöglichen. Mit dem Ansatz, wie in Kapitel 3.2.2.2 beschrieben, lassen sich sehr gut Objekte in Bildern verfolgen. Es bietet sich also an, mit dieser Methode ein einfaches System zur Bewegungsmessung zu realisieren. In Abbildung 4.1 ist das prinzipielle Vorgehen dieser Anwendung gezeigt.

In Abbildung 4.2 ist ein Proband beim Zähneputzen dargestellt. Das Kind sollte in diesem Beispiel die Kaufläche der Backenzähne putzen (Seitz et al. [65]). Die Seitenansicht wurde gewählt, da dann die Bildebene auch die Hauptbewegungsebene der Putzbewegung definiert und eine Reduzierung auf zwei Dimensionen möglich ist. Abbildung 4.3 zeigt dasselbe Kind aber mit überlagerter Position der Marker aus der gesamten Bewegungssequenz. Weil die Marker ihre Erscheinungsform nicht wesentlich ändern, kann die Bildanalyse sehr präzise über lange Zeiten die Markermittelpunkte detektieren. Dabei werden als Basis der Korrelationsanalyse (siehe auch Kapitel 4.2.2) jeweils die Grauwerte jeden Markers aus dem ersten Bild genommen, d.h. die Folgebilder werden immer mit dem ersten Bild der Sequenz verglichen.

Dass der Markermittelpunkt stabil detektiert wird, zeigt Abbildung 4.4. Es ist zu erkennen, dass der Marker auch am Ende der Sequenz nach 6,72 Sekunden noch stabil detektiert wird, obwohl die Bildqualität aufgrund der schnellen Bewegung nicht optimal ist.

Das vorgestellte Werkzeug ermittelt die Position der Marker in Bildkoordinaten. Deshalb können zunächst keine absoluten Messungen gemacht werden. Dennoch lassen sich mit der Methode schnell und unkompliziert vergleichende Bewegungsmessungen realisieren. Bei der Anwendung dieses zweidimensionalen Verfahrens ist darauf zu achten, dass die zu analysierende Bewegung in der Bildebene verläuft.

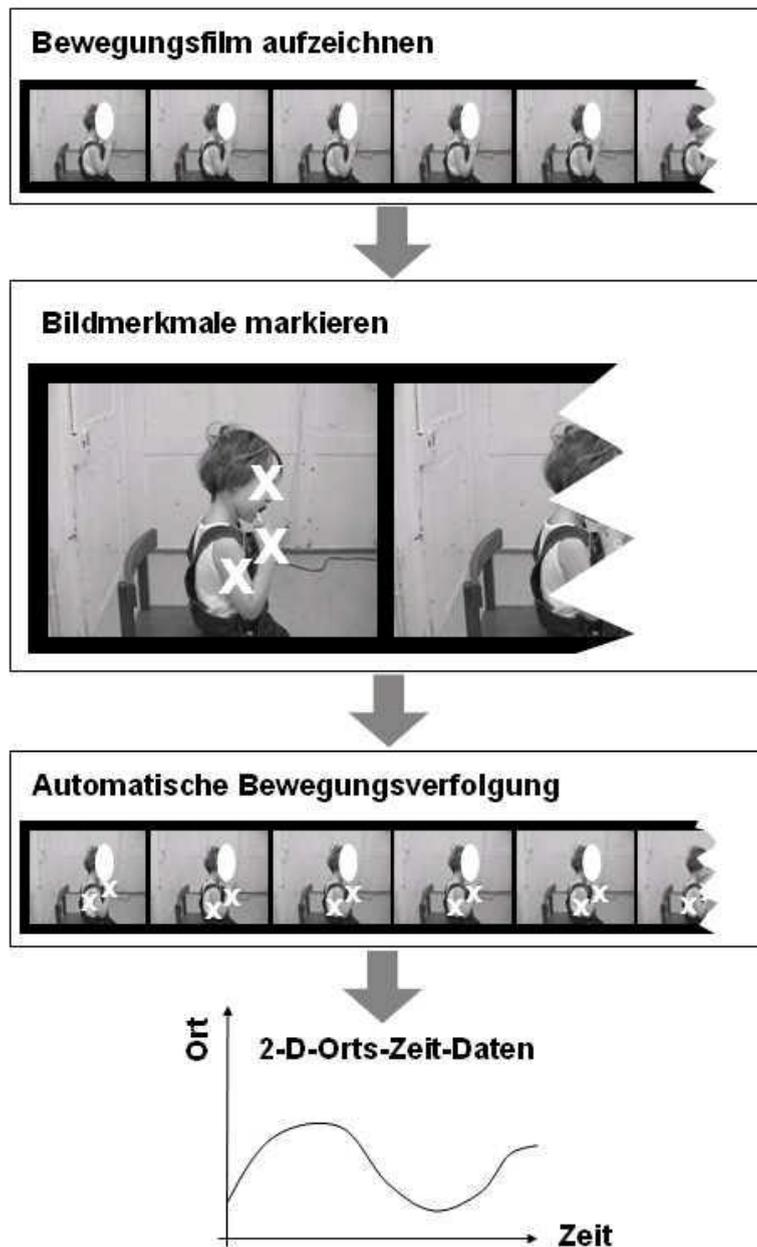


Abbildung 4.1: Prinzipielles Vorgehen bei der zweidimensionalen Bewegungsmessung.



Abbildung 4.2: Seitenansicht eines Kleinkindes beim Zähneputzen. Die Markierungen an Kopf, Oberarm, Unterarm und Hand sind deutlich zu erkennen.



Abbildung 4.3: Selbe Versuchsperson wie oben, aber mit überlagerter Position der Marker (Punktewolken) für alle Bilder der Bewegungssequenz.



Abbildung 4.4: Detektierte Position des Handmarkers. Im ersten Bild (ganz links) ist die manuell markierte Position zur Zeit  $t = 0s$  dargestellt. Die anderen Bilder zeigen von links nach rechts die Markerposition nach eins, zwei, vier, sechs Sekunden und am Ende der Bildsequenz nach 6,72 Sekunden.

## 4.2 Dreidimensionale Bewegungsmessung mit Modell

Bei der dreidimensionalen Bewegungsmessung mit Modell kann ein Bewegungsfilm analysiert werden der durch zwei Kameras erzeugt wird. Die Bilder des Bewegungsfilms müssen dabei als zeitsynchrone Einzelbildpaare vorliegen. In das erste Bildpaar des Films wird das Menschmodell überlagert und exakt an die Probandenhaltung adaptiert. Nach dieser Präpositionierung des Modells bestimmt die Software die Bewegung einzelner Körperelemente anhand der Musterverschiebungen im Folgebildpaar. Sobald die Bewegung für das Bildpaar bestimmt ist, wird aus den Verschiebungen eine dreidimensionale Haltung berechnet und dem Bildpaar überlagert. Durch Vergleichen von Körperkanten im Bild mit der Modellhaut wird die durch das Menschmodell bekannte Anthropometrie des Probanden bei der Analyse für die Haltungsberechnung ausgenutzt. Nach erfolgreicher Haltungsmessung wird die Prozedur für das nächste Bildpaar erneut gestartet, so dass sukzessive alle Bildpaare des Films analysiert werden (vgl. Abbildung 4.6). Im Folgenden wird das Vorgehen beschrieben.

### 4.2.1 Präpositionierung des Modells

Das realisierte System benötigt für die modellbasierte Bildfolgenanalyse die Haltung und Position des Probanden im ersten Bildpaar zur Initialisierung der Prozedur. Die Platzierung des Modells erfolgt halbautomatisch, d.h. der Benutzer gibt per Mausklick die Position relevanter Punkte in den Bildern an. Diese relevanten Punkte werden von der Software abgefragt und korrespondieren mit Gelenkpunkten oder sonstigen prägnanten Punkten am Probanden, wie z.B. der Augenpunkt. Durch dieses Vorgehen wird vom Benutzer ein Satz dreidimensionaler Punkte erzeugt, der die Lage im Raum der folgenden Körperpunkte definiert:

- Augenpunkt

- Schultergelenke
- Ellenbogengelenke
- Handgelenke
- Handspitzen
- Hüftgelenke
- Kniegelenke
- Knöchel
- Fußspitzen

Ziel der Präpositionierung des Modells ist es, das Menschmodell in eine Haltung zu bringen, die durch die Lage der vom Benutzer angegebenen Punkte definiert wird. Während dies bei den Extremitäten relativ einfach ist, da die Lage der Gelenkpunkte, die der Benutzer vorgibt, in der Regel eine eindeutige Haltung beschreibt, ist es für die Haltung des Oberkörpers schwieriger. Die Haltung des Oberkörpers wird durch die Orientierung eines jeden Wirbels der Wirbelsäule definiert. Da die Raumlage einzelner Wirbel vom Benutzer nicht in den Bildern erkannt werden kann wird hier eine Haltungsschätzung vorgenommen. Aus der Lage der Hüftgelenke und der Schultern wird die Krümmung der Wirbelsäule berechnet. Ferner wird die Neigung des Beckens geschätzt.

Im Detail gliedert sich der Ablauf des Algorithmus wie folgt:

1. Modell verschieben: Die Hüftgelenke stellen die Basis der Positionierung dar. Deshalb werden zunächst die Hüftgelenke des Modells an die vom Benutzer angegebenen Punkte im Raum angepasst. Hierzu wird das Menschmodell in  $x$ ,  $y$  und  $z$ -Richtung verschoben, so dass der euklidische Abstand zwischen Modellgelenkpunkt und vom Benutzer geschätztem Punkt am geringsten ist.
2. Torso krümmen: Ausgehend von der Lage des Beckens aus Punkt 1 wird der Winkel zwischen Hüftgelenk und Schulterpunkte ermittelt. Daraus resultiert die Beckenneigung des Modells. Durch Bewegen der Oberschenkel, so dass die Kniegelenke sich mit den vom Benutzer markierten Positionen decken, wird die Oberschenkelhaltung angepasst und hieraus der Öffnungswinkel zwischen Torso und Oberschenkel ermittelt. Aus diesem Winkel wird die Krümmung der Wirbelsäule geschätzt, wobei der lordotische (dorsal konkave Krümmung der Wirbelsäule; Rundrücken) und kyphotische (dorsal konvexe Krümmung der Wirbelsäule; Hohlkreuz) Fall unterschieden wird. Bei Öffnungswinkeln zwischen Oberschenkel und Torso kleiner 180 Grad ist der konkave

Fall wahrscheinlich, bei Öffnungswinkeln größer 180 Grad der konvexe (Lordotische und kyphotische Sitzhaltungen werden nicht berücksichtigt). Mit Festlegung der Krümmungsart der Wirbelsäule können dann die Wirbelsäulengelenkwinkel so variiert werden, dass die Schulterpunkte erreicht werden.

3. Extremitäten anpassen: Kniegelenk, Knöchel, Fußspitze, Ellbogen, Handgelenk, Handspitze und Augenpunkt werden in der angegebenen Reihenfolge an die vom Benutzer vorgegebenen Raumpunkte angenähert.

Bei der beschriebenen Anpassung werden für jedes beteiligte Gelenk alle möglichen Winkelstellungen berechnet und ausgehend von der resultierenden Haltung der euklidische Abstand zum Zielpunkt bestimmt. Für die endgültige Haltung werden jene Winkel ausgewählt, die zu einem minimalen Abstand zum Zielpunkt führen. Um den Vorgang zu Beschleunigen wird dabei zunächst die Haltung in großen Winkelschritten variiert. Sobald die ungefähre Lage bestimmt wurde, wird in kleinen Winkelschritten die Haltung verfeinert. Dieses Vorgehen wirkt sich positiv auf die Berechnungsgeschwindigkeit aus.

Die resultierende Haltung dient als Basis für die weitere Analyse. Falls sich die Haltung nicht exakt mit dem Probanden deckt, so kann manuell korrigiert werden.



Abbildung 4.5: Bildfolge aus zwei Kameraansichten. Für die Bewegungsverfolgung wird Bild für Bild abgearbeitet und die Haltung des Menschmodells berechnet und in die Bildfolge überlagert.

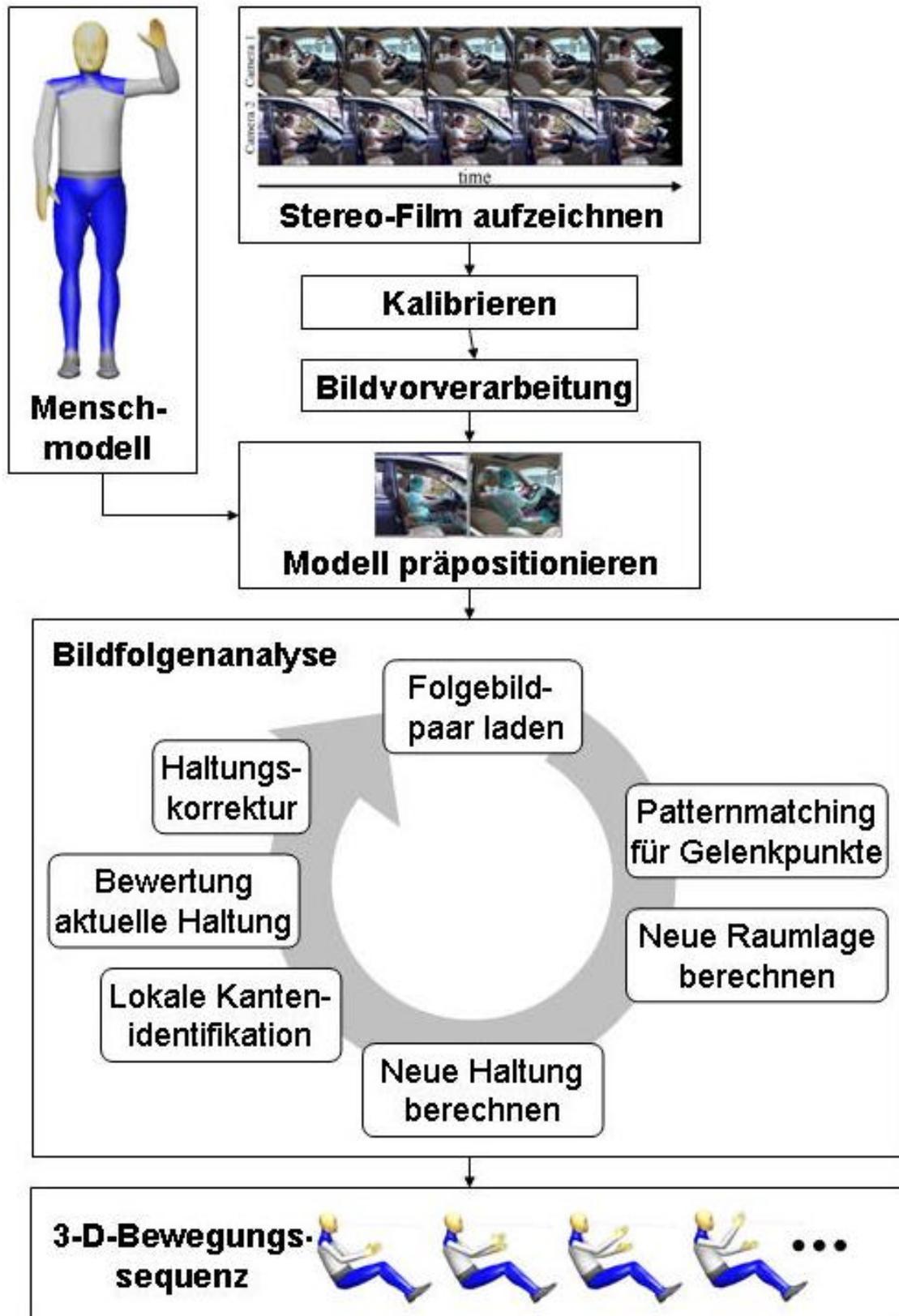


Abbildung 4.6: Übersichtsdiagramm der Funktionsweise der modellbasierten dreidimensionalen Bewegungsmessung.

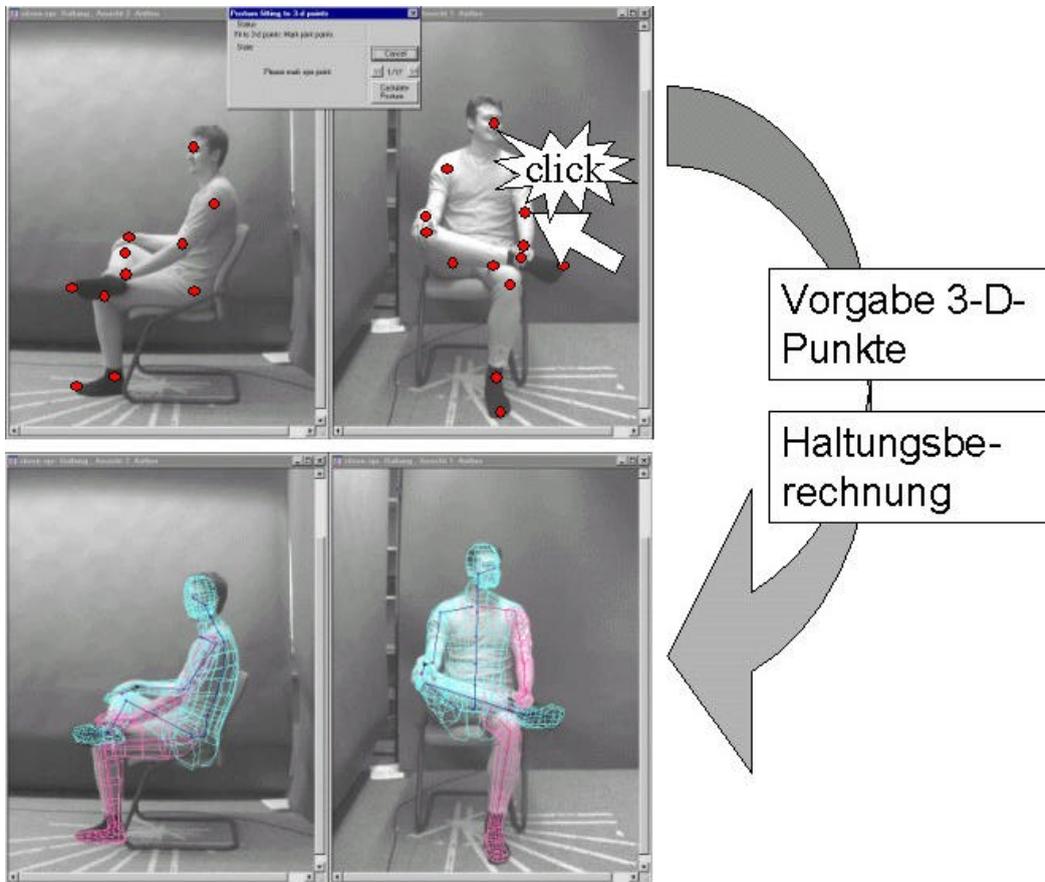


Abbildung 4.7: Haltungsberechnung durch Vorgabe relevanter Punkte. Die Haltung der Wirbelsäule wird dabei aufgrund der Lage von Schulter und Hüftgelenke bestimmt.

#### 4.2.2 Gelenkpunktverschiebung detektieren

Das überlagerte Menschmodell beschreibt die Haltung der Versuchsperson dreidimensional und definiert jene Bildbereiche (AOI), die mit Eigenschaften der Versuchsperson korrespondieren. Durch die Initialisierung kann ein Satz  $G_n$  von Vektoren  $\vec{g}_{n,i,j} = (x, y)^T$  erzeugt werden, welche die Position eines jeden AOIs auf einem Bild beschreibt, wobei  $n$  die Nummer des Bildes der Bildfolge (erstes Bild  $n=0$ ),  $i$  die Nummer der Kamera,  $j$  eine Modelleigenschaft und  $x$  sowie  $y$  die Koordinate einer korrespondierenden Modelleigenschaft  $j$  ist. Jeder AOI  $A_{n,i,j}$  ist eine  $(2M+1) \times (2N+1)$  Matrix und gehört zu je einem Vektor  $\vec{g}_{n,i,j}$ . Die Matrix  $A_{n,i,j}$

trägt Grauwertinformationen  $I$  als Matrixelemente und beschreibt damit das Bildmuster in der lokalen Umgebung eines jeden Vektors  $\bar{g}_{n,i,j}$ .

Für alle Folgebilder (siehe Abbildung 4.5) wird ein neuer Satz von Vektoren  $G_n$  berechnet, welche die neue Position korrespondierender Modelleigenschaften beschreibt. Die Verschiebung  $V_{n+1}$  zwischen Modelleigenschaften aus Bild  $n$  und Bild  $n+1$  berechnet sich zu  $V_{n+1} = G_{n+1} - G_n$ . Auf Basis der Korrelationsanalyse zur Mustererkennung (wie in Kapitel 3.2.2.2 beschrieben) wird eine Korrelationsmatrix  $C$  berechnet. Die Koeffizienten  $c_{k,l}$  dieser  $(2M+1) \times (2N+1)$  Matrix  $C$  wurden zu

$$c_{k,l} = \frac{1}{K} \sum_{u=-M}^M \sum_{v=-N}^N [(I_1(x+u, y+v) - \bar{I}_1(x, y)) \times (I_2(k+u, l+v) - \bar{I}_2(k, l))] \quad (4.1)$$

mit

$$K = K_1 K_2$$

$$K_1 = \sqrt{\sum_{u=-M}^M \sum_{v=-N}^N (I_1(x+u, y+v) - \bar{I}_1(x, y))^2} \quad (4.2)$$

$$K_2 = \sqrt{\sum_{u=-M}^M \sum_{v=-N}^N (I_2(k+u, l+v) - \bar{I}_2(k, l))^2} \quad (4.3)$$

berechnet, wobei  $I_2$  die Grauwerte eines AOI im Folgebild repräsentiert und  $I_1$  die Grauwerte des Vergleichsbildes. Wenn

$$\bar{x}_{n+1,i,j} = \min(C_{n+1,i,j}), \quad (4.4)$$

ein Vektor, der zur Position des Minimums zeigt, dann ist der Bewegungsvektor  $\bar{v}_{n+1,i,j}$  gegeben durch

$$\bar{v}_{n+1,i,j} = \bar{x}_{n+1,i,j} - \bar{g}_{n,i,j}. \quad (4.5)$$

$C_{n,i,j}$  beschreibt ein "pattern matching" für Bild  $n$ , Kamera  $i$  und Modelleigenschaft  $j$  (vgl. auch Seitz und Bubb [63]).

### 4.2.3 Resultierende Haltung im Raum bestimmen

Basierend auf den resultierenden zweidimensionalen Vektoren werden dreidimensionale Vektoren berechnet, die dann die Position von Modellpunkten im Raum beschreiben. Durch die Korrelationsanalyse sind für jeden beobachteten Gelenkpunkt Bildkoordinaten ermittelt worden, die seine neue Position beschreiben. Diese Bildvektoren  $\vec{x}_{n+1,i,j}$  beschreiben einen Punkt auf der Bildebene, der, legt man eine Gerade durch diesen Punkt und den Fokus der Kamera, eine Sichtgerade beschreibt, entlang derer sich der wahre Gelenkort im Raum befinden kann. Da Sichtgeraden aus zwei oder mehr Kameraperspektiven für identische Modellpunkte vorliegen, ergibt sich durch Schnitt dieser Geraden die Raumkoordinate des entsprechenden Punkts. In Abbildung 4.8 sind zwei Sichtgeraden  $g_1$  und  $g_2$  dargestellt. Da sich die Geraden in der Regel nicht wirklich schneiden (wegen Messungenauigkeiten und der diskreten Pixelmatrix des CCD) muss der Ort der größten Annäherung ermittelt werden. Seien

$$\vec{r}_1 = \vec{a}_1 + s_0 \vec{b}_1 \quad (4.6)$$

und

$$\vec{r}_2 = \vec{a}_2 + t_0 \vec{b}_2 \quad (4.7)$$

Vektoren, die jeweils einen Punkt auf den Sichtgeraden beschreiben. Dann ergibt sich aus der Bedingung, dass  $\vec{r}_2 - \vec{r}_1$  minimal werden soll eine Gleichung für den Vektor  $\vec{r}$ , der einen Raumpunkt beschreibt, der in der Mitte der kürzesten Strecke zwischen  $g_1$  und  $g_2$  liegt. Für das Beispiel aus Abbildung 4.8 mit zwei Sichtgeraden ergibt sich der Vektor  $\vec{r}$  zu

$$2\vec{r} = \vec{a}_1 + \vec{a}_2 + \frac{[(\vec{b}_1 \times \vec{b}_2) \times \vec{b}_2](\vec{a}_1 - \vec{a}_2) \cdot \vec{b}_1}{\|\vec{b}_1 \times \vec{b}_2\|^2} + \frac{[(\vec{b}_1 \times \vec{b}_2) \times \vec{b}_1](\vec{a}_1 - \vec{a}_2) \cdot \vec{b}_2}{\|\vec{b}_1 \times \vec{b}_2\|^2}. \quad (4.8)$$

Ausgehend von diesen Gelenkpositionen errechnet der in Abschnitt 4.2.1 erwähnte Algorithmus die Haltung, die durch diese Gelenkorte am besten beschrieben wird, wobei durch Einschränkung des Suchbereichs gewährleistet wird, dass nur solche Haltungen akzeptiert werden, die der vorherigen Haltung ähnlich sind, d.h. es wird erwartet, dass die Folgebilder nur kleine Bewegungsänderungen zeigen. Die resultierenden Haltungen werden den Bildern anschließend überlagert, so dass die Qualität der Messung begutachtet werden kann. Gegebenenfalls können nun noch manuelle Korrekturen vorgenommen werden.

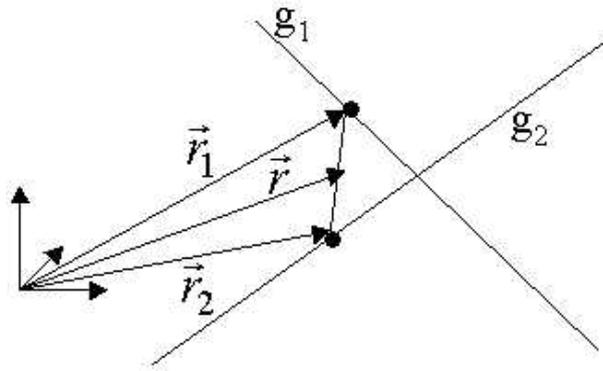


Abbildung 4.8: Bestimmung eines Raumpunktes mit Hilfe der Sichtgeraden  $g_1$  und  $g_2$ .

#### 4.2.4 Integration von Kanteninformation

Die reine Verwendung der Korrelationsmethode zum Detektieren der Körperhaltungsveränderungen hat allerdings zwei entscheidende Nachteile. Zum einen ist die Qualität der Körperverfolgung stark von der Ausprägtheit der Bildmerkmale die für die Bildverfolgung verwendet werden abhängig und zum anderen fehlt die Möglichkeit, eine mögliche Abweichung zwischen wahrer und berechneter Körperhaltung zu messen und zu bewerten. Dies hat zur Folge, dass die Bewegungsanalyse Haltungen ermittelt, die nach einiger Zeit deutliche Abweichungen zur tatsächlichen Haltung aufweisen, ohne dass dies vom System bewertet bzw. korrigiert werden kann. Abbildung 4.9 zeigt die Qualität der Bildverfolgung am Beispiel einer Schaltbewegung im PKW. Es ist zu erkennen, dass die Position des RAMSIS-Arms eine geringfügig falsche Position aufweist. Gerade bei ergonomischen Untersuchungen ist jedoch eine höhere als die hier erreichte Genauigkeit bei Bewegungs- bzw. Haltungsmessungen gefordert. Die Abweichung rührt vor allem von Verschiebungen bei der Bildkorrelation her, da der Arm keine ausreichend prägnanten Bildmuster liefert. Diese Drift beim Korrelieren von Bildmerkmalen in langen Bildsequenzen gilt es durch einen geeigneten Mechanismus zu erkennen und zu korrigieren. Der Abstand zwischen der Kontur des Modells zur Kontur des realen Körpers gibt hierbei ein Maß für die Größe der Fehlplatzierung.

Eine Möglichkeit, Konturen in Bildern zu detektieren, ist die Erzeugung von Kantenbildern, wie es in Kapitel 3.2.3 dargestellt ist. Die Information über die Lage von Körperkonturen ermöglicht es einen Abgleich mit Modellkonturen durchzuführen. Da die Bilder nach einer Kantendetektion sehr viele Kanten zeigen, beschränkt sich die Vorgehensweise hier darauf, Kanten in der Nähe von Gelenkpunkten zu detektieren. Durch diese lokale Kantendetektion wird zum einen der Rechenaufwand reduziert und zum anderen der Einfluss des Hintergrunds

minimiert. Wie in Abbildung 4.10 dargestellt, sucht der Algorithmus in der Nachbarschaft, des Modellgelenkpunktes nach Kanten. Wenn der Abstand  $x'$  der detektierten Kanten zueinander vergleichbar des vom Modell vorgeschriebenen Abstands  $x$  ist (Abbildung 4.11), wird das Modell entlang der Kanten ausgerichtet. Durch diesen Vergleich wird gewährleistet, dass nur Kanten berücksichtigt werden, die wahrscheinlich zur Person gehören. Mit diesem Ansatz können kleine Fehler der Korrelationsanalyse ausgeglichen werden (Seidl [59], Werner [73]).

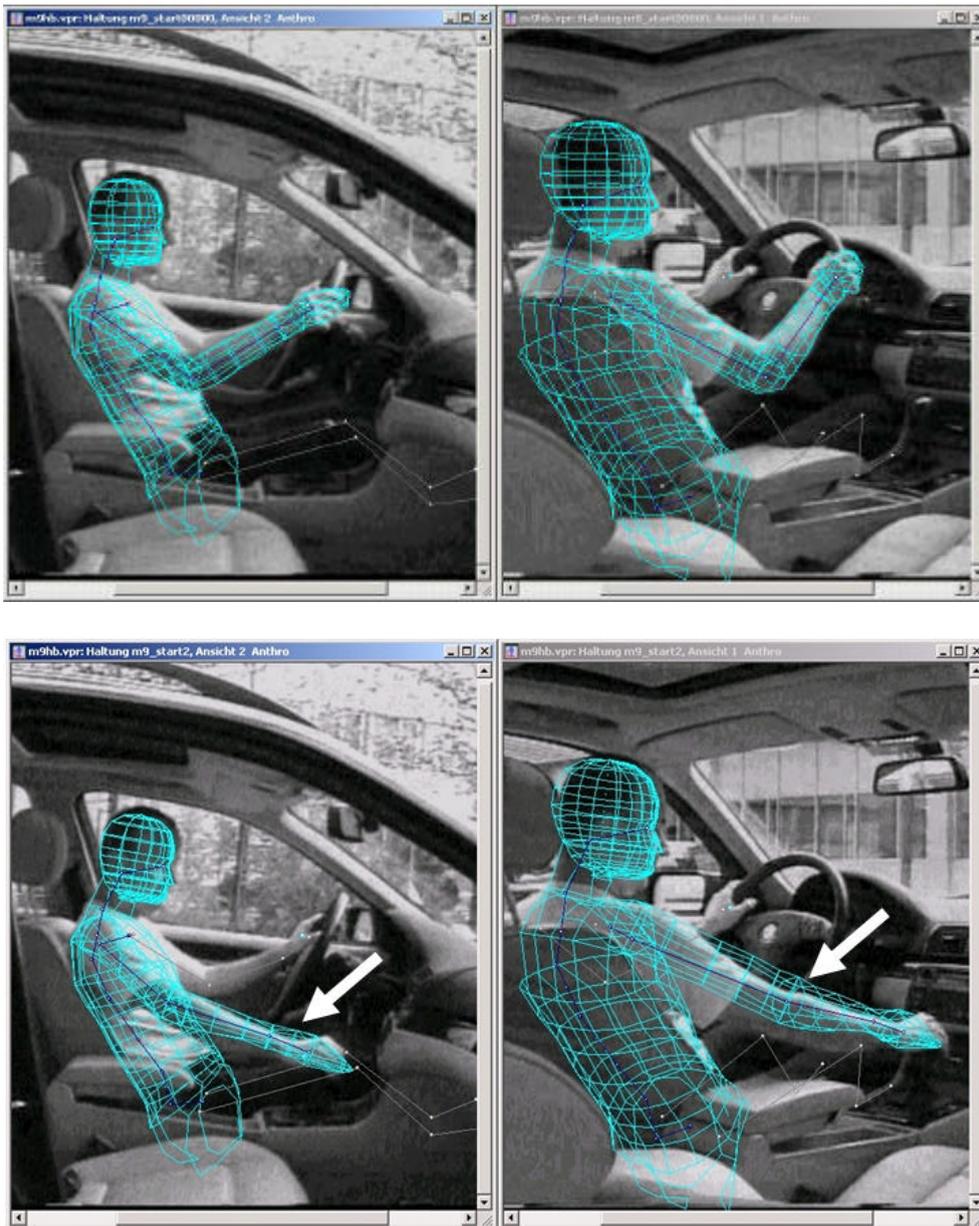


Abbildung 4.9: Oben: Startbild für die Analyse. Unten: Nach der Verfolgung von 25 Bildern kann ein deutlicher Unterschied zwischen wahrer und überlagerter Armposition erkannt werden.

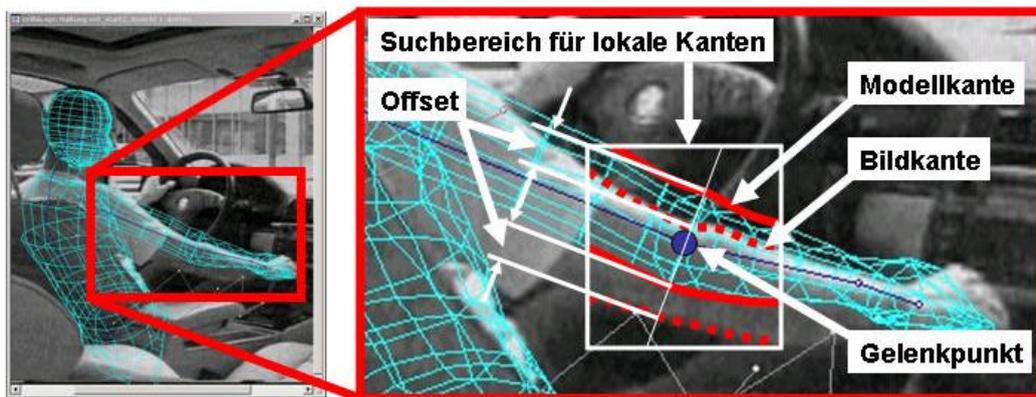


Abbildung 4.10: Durch Detektion lokaler Kanten in der Nachbarschaft von Gelenkpunkten werden Modellkanten mit Bildkanten abgeglichen, um das Modell präziser zu platzieren.

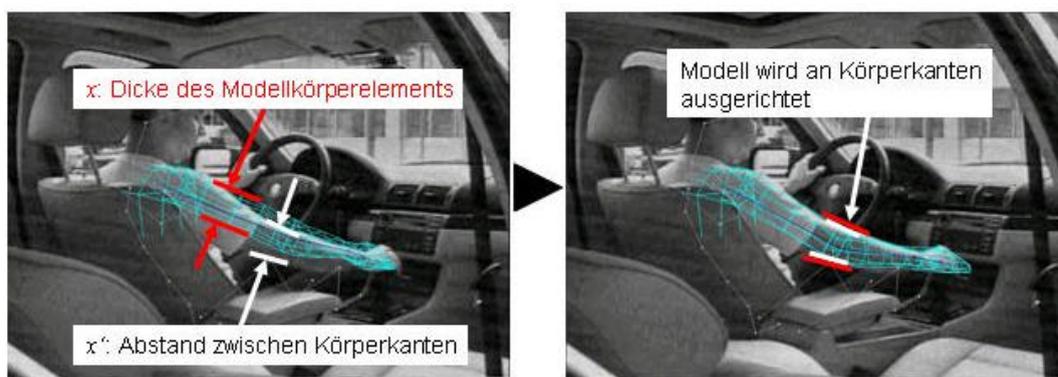


Abbildung 4.11: Durch den Vergleich des Abstands zwischen detektierten Körperkanten  $x'$  und dem Sollwert  $x$  wird gewährleistet, dass das Modell nur an echten Körperkanten ausgerichtet wird.

### 4.3 Probleme

Die genaue und stabile Messung menschlicher Bewegungen stellt bis heute eine große Herausforderung dar. Weltweit existieren viel versprechende Ansätze, die jedoch alle mit Tücken und Problemen behaftet sind. So ist auch der hier vorgestellte Ansatz ein weiterer Beitrag menschliche Bewegungen einfach zu erfassen, der aber naturgemäß seine Probleme und Grenzen hat.

### 4.3.1 Drift beim Verfolgen unprägnanter Strukturen

Das Wiederfinden von Körpermerkmalen in Bildfolgen mit Hilfe der Kreuz-Korrelation bzw. Differenzbildung funktioniert in der Regel sehr stabil, solange die gesuchte Struktur ausreichend prägnante Merkmale besitzt, anhand derer sie eindeutig im Folgebild wieder erkannt werden kann. Jedoch besitzt der menschliche Körper bzw. getragene Kleidung oft keine solchen Merkmale, so dass zwangsläufig Körperpunkte ungenau wieder gefunden werden. In Abbildung 4.12 ist die Drift beim Verfolgen des Ellenbogengelenks dargestellt. Obwohl bei der Bildfolge die Bild-zu-Bild-Veränderungen gering sind, ist schon nach kurzer Zeit das Merkmal fehldetektiert. Die vorgestellte Einführung der Kantenstabilisierung kann hier zwar eine Korrektur bewirken, doch problematisch bleibt die Tatsache, dass in der Regel in beiden Kameraansichten diese Drift zu beobachten ist und deshalb der resultierende 3-D-Punkt weit außerhalb des Körpervolumens liegen kann. Daraus folgt, dass die Haltungsberechnung eine Haltung prognostiziert, bei der das überlagerte Menschmodell außerhalb der Bildbereiche, die dem Probanden zugeordnet sind, liegt. Bei solch einer Fehldetektion ist die Kantenstabilisierung nicht mehr in der Lage, eine Korrektur durchzuführen. Denn für die Kantenstabilisierung ist es nötig, dass überlagerte Modellgelenke innerhalb des Bildbereichs des Probanden liegen.

Das beschriebene Problem führt dazu, dass eine manuelle Korrektur der Haltungsprognose vorgenommen werden muss. Ausgehend von der manuell bestimmten Lage erfolgt die weitere automatische Analyse der Folgebilder, bis wieder eine deutliche Differenz zwischen Soll- und Ist-Haltung auftritt.

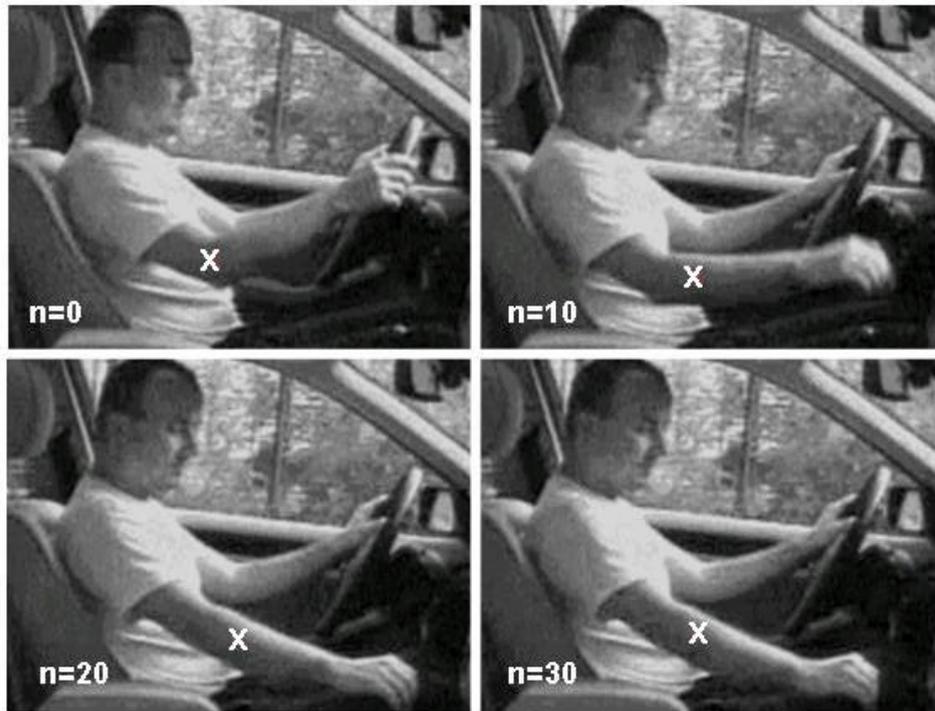


Abbildung 4.12: Drift beim Wiederfinden des Ellenbogengelenks. Die Kreuze markieren die von der Kreuz-Korrelation gelieferten Ergebnisse. Deutliche ist eine Wanderung der detektierten Punkte längs des Arms zu erkennen. n bezeichnet die laufende Bildnummer der Sequenz.

### 4.3.2 Nicht ausgeprägte Kanten

Die Kantenstabilisierung kann nur erfolgen, wenn

1. Kanten detektiert werden und wenn
2. die detektierten Kanten zum Körperumriss des Probanden gehören.

Beide Punkte sind in der Realität nicht immer erfüllbar. Abbildung 4.13 zeigt, dass Körperkanten in der Regel nicht über die ganze Sequenz gleich prägnant hervortreten. So treten hier die Konturen des bewegten rechten Arms mal mehr und weniger deutlich hervor. Die Kantenstabilisierung steht nun vor dem Problem prinzipiell zu erkennen, ob eine Kante überhaupt zur gesuchten Kontur gehört. Als Kriterium hierfür wird der Abstand von zwei Kanten zueinander verwendet. Wenn er in der Größenordnung des Durchmessers des korrespondierenden Körperelements ist, dann wird die Kante als Körperkante interpretiert und das Modell dazwischen platziert. Im Falle des Arms heißt dies, dass die obere und untere Armkante detektiert werden muss. Wird nur eine Kante gefunden, so kann nicht eindeutig gesagt werden,

ob die Kante ein Bildartefakt (Schatten, Hintergrund oder andere Strukturen) oder die tatsächliche Körperkante darstellt. Eine Anpassung an die Kontur bleibt aus. In die Haltungspgnose geht dann nur die Information aus der Korrelation ein. Bei einer fehlerhaften Detektion, muss manuell korrigiert werden.

### **4.3.3 Verdeckung**

Von Seiten der Bildverarbeitung kann eine Bewegung eines verdeckten Körperelements oder –bereichs nicht behandelt werden. In der Bewegung von Abbildung 4.9 beispielsweise kann die Bewegung der linken Schulter wegen der Verdeckung nicht ermittelt werden. Allerdings bietet das verwendete Menschmodell die Möglichkeit automatisch zu erkennen, welche Körperregionen verdeckt sind, so dass entweder verdeckte Körperteile starr bleiben bzw. der Benutzer diese ggf. manuell justiert. Mit Hilfe von Prognosemodellen (wie z.B. Kalman-Filter) könnten kurzzeitige Verdeckungen einzelner Körperelemente behandelt werden, so dass nach verlieren eines Körpermerkmals, dieses wieder gefunden und das Tracking fortgesetzt werden kann.

### **4.3.4 Rotation**

Der entwickelte Ansatz zur Bewegungsmessung benutzt die Detektion translatorischer Verschiebungen in Bildern, um hieraus eine Haltung zu berechnen. Dabei können Rotationen von Körperelementen entlang der Längsachse nur richtig interpretiert werden, wenn ein Winkel zwischen rotiertem Körperelement und nachfolgendem bzw. vorhergehendem Körperelement besteht. Die Rotation des Oberarms ist beispielsweise bei angewinkeltem Unterarm aus der Unterarmposition eindeutig zu bestimmen (siehe Abbildung 4.14). Bei gestrecktem Unterarm – Ober- und Unterarm bilden eine Linie – ist eine Rotation längs der Knochenachse nicht zu detektieren, da die Bildverarbeitung zwar eine translatorische Verschiebung detektieren würde (Punkte auf dem Arm wandern während der Rotation in der Bildebene), diese aber zu einer fehlerhaften Haltungspgnose führt.

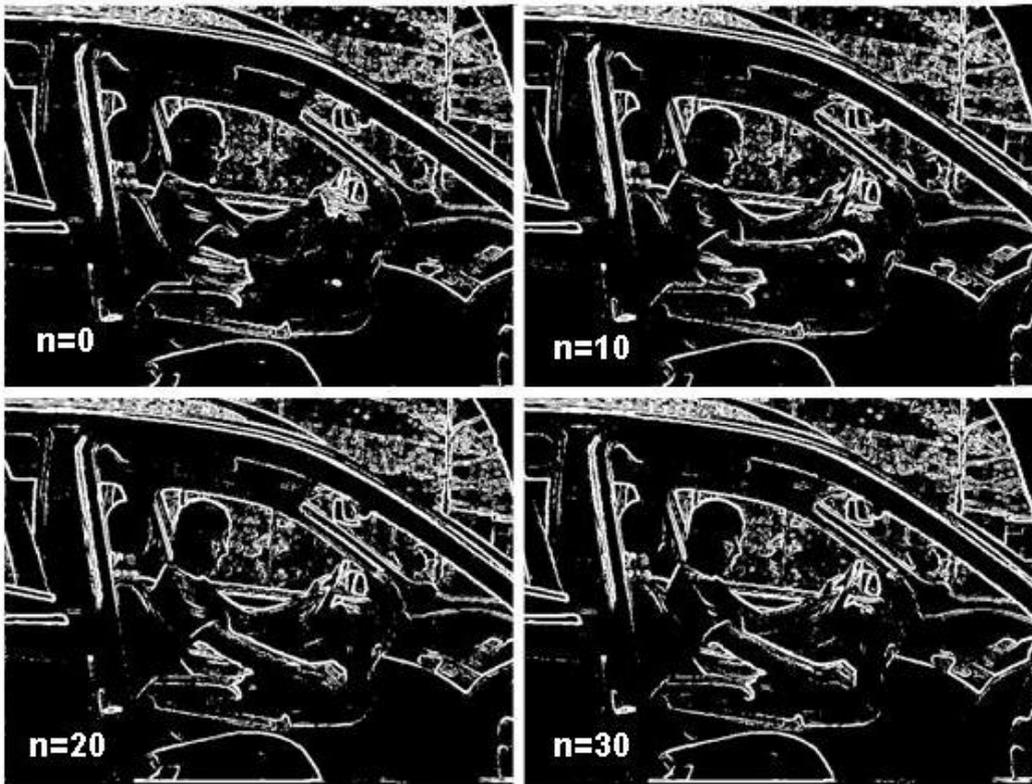


Abbildung 4.13: Innerhalb einer Bildsequenz variiert die Ausprägtheit von Kanten, z.B. weil der Hintergrund sich nicht stark vom Körper unterscheidet. So ist bei  $n=0$  die Armkontur praktisch verschwunden, bei  $n=10$  die obere Kontur des rechten Arms deutlich zu sehen und bei  $n=20$  und  $n=30$  die obere und untere Kontur ausgeprägt.

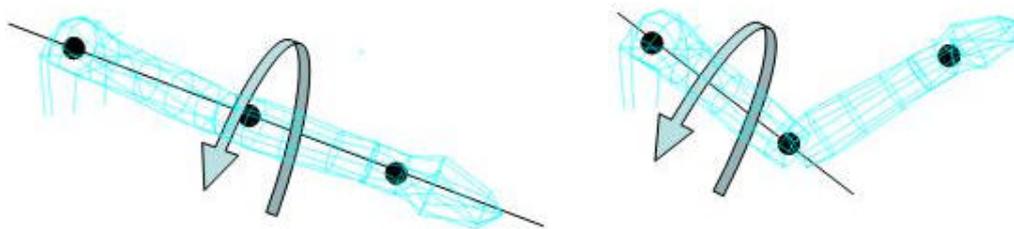


Abbildung 4.14: Eine Rotation um Körperelementlängsachsen ist mit dem vorgestellten Ansatz nur messbar, wenn ein Winkel zwischen zwei Körperelementen besteht (rechte Abbildung).

### 4.3.5 Genauigkeit

Ziel der vorgestellten Methode ist es, einen zeitlichen Winkelverlauf aller Modellgelenke zu erhalten. Aus Modellparametern sind ferner die räumlichen Orte von Gelenk- und Hautpunkten des Modells zu ermitteln. Die Genauigkeit, mit der die Messung eines Punktes im Raum erfolgen kann, hängt von der Auflösung der verwendeten Kamera, so wie dem Abstand zwischen Kamera und Person ab. Generell lässt sich sagen, dass bei doppeltem Abstand auch nur die halbe räumliche Genauigkeit erreicht werden kann. Bei Verwendung von Kameras mit Standardauflösung von 576x768 Pixeln kann bei der Vermessung einer Versuchsperson von 1,80 m Körpergröße eine Auflösung von 3 mm erreicht werden. Bei Messungen des Torso (ca. halbe Körperhöhe) können die Kameras entsprechend näher an der Versuchsperson platziert werden, so dass sich eine Auflösung von ca. 1,5 mm ergibt. Die durch die Anzahl der Pixel vorgegebene Auflösung definiert die kleinstmöglichen Distanzen, die vermessen werden können. Dieser Wert vergrößert sich jedoch durch leichte Kalibrierfehler bei der Korrektur der Linseneigenschaften sowie der Standortbestimmung der Kameras. Hinzu kommen ferner Fehler bei der Analyse mit Bildverarbeitung sowie beim Rückrechnen von Raumpositionen auf Bildpositionen und umgekehrt. Die Analyse einer Ganzkörperbewegung mit einem angenommenen Bildverarbeitungsfehler von 2 - 4 Pixeln würde zu Fehlern in der Positionierung der Körperelemente von minimal 9 - 15 mm im Raum führen. Die Verwendung hoch auflösender Kameras kann hierbei die Genauigkeit stark erhöhen.

## 4.4 Zusammenfassung

Ausgehend von der Idee biologische Prinzipien des Bewegungssehens nachzuahmen und Funktionen, wie Korrelation und Modell-Bild-Vergleich zu integrieren, entstand der vorgestellte Ansatz. Das vorgestellte Verfahren detektiert durch Bild-zu-Bild-Vergleich Bewegungen. Dabei können sowohl prägnante Markierungen ohne Verwendung eines überlagerten Menschmodells, als auch Texturmerkmale mit Unterstützung eines Menschmodells verfolgt werden.

Im ersten Fall, der im Wesentlichen die heutige kommerzielle Vorgehensweise in der Bewegungsanalyse repräsentiert, wird ein Proband gezielt mit Markierungen bestückt. Durch die Tatsache, dass sich die Markierungen in ihrer Erscheinung nicht verändern, wird deren Position zuverlässig verfolgt. Als Resultat ergeben sich markerabhängige Bewegungsdaten.

Im zweiten Fall wird die Bewegungsanalyse unterstützt durch das verwendete Menschmodell. Die Lage von angebrachten Markierungen oder Texturmerkmalen in Relation zum Modell ist dabei unerheblich, denn durch die Überlagerung des Modells wird diese Relation individuell neu erstellt. Dies hat zur Konsequenz, dass bei Verwendung von Markern diese an nicht näher definierten Stellen am Körper positioniert werden können, so dass bei Folgemessungen mit der selben Versuchsperson Marker nicht wieder an den selben Körperstellen angebracht werden müssen. Ferner kann ganz auf externe Markierungen verzichtet werden, da mit Hilfe der Bildverarbeitung auch körper- und kleidungseigene visuelle Merkmale detektiert werden. Allerdings ist dann die erzielte Genauigkeit geringer, da die Verfolgung von wenig ausgeprägten Körpermerkmalen zu einer unstabilen Bildverfolgung führt.

Um dieses Problem in den Griff zu bekommen wurden zusätzlich die Körperumrisse in der Nähe von Gelenkpunkten zur Stabilisierung der Bewegungsverfolgung hinzugezogen. Da aus dem individuellen Modell des Probanden die genaue Anthropometrie bekannt ist, kann die Abweichung zwischen Modellposition und realer Probandenposition ermittelt und die Lage des Modells korrigiert werden.

## 5 Ausblick

Auf Grundlage dieser Arbeit können in Zukunft Haltungs- und Bewegungsmessungen konform zum Menschmodell RAMSIS einfacher und schneller durchgeführt werden. Damit ist eine Erleichterung der Forschungen auf dem Gebiet des *Human Modeling*, insbesondere der Modellierung von Bewegungen, erreicht worden. Dieser Ansatz erlaubt es modellbasiert und ohne Verwendung von Markern die Bewegung von Menschen zu messen. Daher werden die Probanden in keiner Weise bei ihrer Bewegung behindert oder eingeschränkt, wie dies bei Markersystemen der Fall ist – ungestörte Bewegungen können beobachtet werden. Aufgrund der nicht zwingend nötigen Präparierung von Probanden können außerdem wesentlich mehr Versuche pro Zeiteinheit durchgeführt werden, als dies heute der Fall ist.

Allerdings ist die Genauigkeit und Stabilität des Ansatzes verbesserungsfähig. Dies kann vor allem durch eine Verfeinerung der vorgestellten Methoden erreicht werden, z.B. Betrachtung von Kanten über das ganze Modell. Da Kanten bzw. Konturen in der Regel nur unvollständig aus Bildern zu extrahieren sind, müssen hier stabilere Methoden verwendet werden. Hansen [26] beispielsweise detektierte mit Hilfe *aktiver Konturen* Körperumrisse von beliebig bekleideten Menschen. Eine Kombination des hier vorgestellten Ansatzes und dieser Arbeiten erscheint nahe liegend. Allerdings ist anzumerken, dass ganz generell die Genauigkeit durch die Auflösung des verwendeten CCD-Chips eingeschränkt ist. Bei der Aufzeichnung von Bildsequenzen tritt jedoch die Videotechnik zusehens in den Hintergrund, so dass hier zu erwarten ist, dass die heutige Technik bald durch hoch auflösende CCDs und digitale Technik abgelöst wird.

Die Kopplung von Videoaufnahmen an das Menschmodell RAMSIS erlaubt eine Reihe nützlicher Interpretationen der gemessenen Haltungen, die ergonomische Untersuchungen erleichtern. So könnten beispielsweise mit Hilfe des Systems dynamische Arbeitsplatzuntersuchungen durchgeführt und mit Hilfe der Haltungsinformationen im Sinne der in der Ergonomie etablierten OWAS-Methode analysiert werden. Ferner können mit Hilfe des individuellen Menschmodells kinematische Aspekte einer Bewegung untersucht werden, so dass z.B. aus den Trägheitssensoren einzelner Körperelemente und den gemessenen Bewegungsbahnen die dynamischen Belastungen des Körpers ermittelt werden können (Zauner [80]). Ein weiteres Entwicklungspotenzial stellt die Onlinefähigkeit dar, so dass z.B. der Einsatz in Virtual Reality Umgebungen in Frage käme oder allgemein beim Visualisieren von natürlichen Bewegungen, wie es für die Produktions- und Produktplanung nötig ist. Heute werden dort Tracking-Systeme benutzt, die ohne Modell und mit Markern arbeiten.

## 6 Literatur

1. Aggarwal, K.J. und Cai, Q., Human Motion Analysis: A Review. *Computer Vision and Image Understanding*, 73(3), S. 428-440, 1999.
2. Arlt, F. und March, A., CAD modelling of a human 3D child body, *Int. J. of Industrial Ergonomics* 22, 1998.
3. Badler, N.I., Philipps, C.B., und Webber, B.L., *Simulating Humans: Computer Graphics Animation and Control*, Oxford University Press, New York, 1993.
4. Bakic, V. and Stockman, G., Menu Selection by Facial Aspect, *Proceedings of Vision Interface*, Québec, 1999.
5. Barnes, R.B., *Motion and Time Study*, Wiley, New York, Seite 10ff, 1968.
6. Baum, E., *Motografie*, Bundesanstalt für Arbeitsschutz und Unfallforschung, Dortmund, 1980.
7. Bear, M. F., Connors, B. W. und Paradiso M.A., *Neuroscience – Exploring the Brain*, Williams & Wilkins, Baltimore, 1996.
8. Braune, W. und Fischer, O.: Der Gang des Menschen, in: *Abhandlungen der mathem.-phys. Classe der Königl. Sächsischen Gesellschaft der Wissenschaften*, Leipzig, 1895.
9. Bubb, H., *Ergonomie des Mensch-Maschine-Systems*, Habilitationsschrift, Technische Universität München, 1977.
10. Bubb, H., From Measurement to Design and Evaluation – the Idea of the CAD-tool RAMSIS, *Workshop on 3-D Anthropometry and Industrial Products Design*, Editor A. Coblentz, Paris, 1998.
11. Bubb, H., Human Modeling in the Past and Future – The Lines of European Developments, *SAE International Conference, Digital Human Modeling for Design and Engineering*, The Hague, The Netherlands, 1999.
12. Bubb, H., Küger, W., Schmidtke, H. und Speyer, H., *Ergonomie im Fahrzeuginnenraum – Rechnergestützte Insassensimulation auf der Basis mathematischer Modelle und aktueller ergonomischer Daten*, VDI-Berichte 816, S.459-468, VDI Verlag Düsseldorf, 1990.
13. Case, K., Porter, J.M. und Bonnes, M.C., SAMMIE: A Man and Workplace Modeling System, in Karwowski, W., Genaidy, A.M. und Asfour, S.S. (Eds.), *Computer-aided Ergonomics*, S. 47, Taylor & Francis, London, 1990.

14. Chaffin, D., Digital Human Modeling for Vehicle and Workplace Design, Society of Automotive Engineers Inc., Warrendale, 2001.
15. Chaffin, D.B., Andersson, G.B.J. und Martin, B.J., Occupational Biomechanics, 3<sup>rd</sup> Edition, John Wiley and Sons, New York, 1999.
16. Chung, J.-M., Nagata, t. und Ohnishi, N., Measuring Human Actions using Sequential Boundary Contour Pairs, In: Proceedings of the IEEE International Workshop on Robot and Human Communication, 1996.
17. Darell, T. Maes, P., Blumberg B. und Pentland, A., A novel environment for situated vision and behavior. In: IEEE Workshop for Visual Behaviors, 1994.
18. Davis, W., J. und Bobick, F., A., SIDEshow: A Silhouette-based Interactive Dual-screen Environment, M.I.T. Media Laboratory Perceptual Computing Section Technical Report No. 457, M.I.T., Cambridge, 1999.
19. Elgammal, A., Harwood, D. und Davis, L., Non-parametric Model for Background Subtraction, Publications of the Computer Vision Laboratory, University of Maryland, 2000.
20. Gavrilu, D. und Davis, L.S., 3-D model-based tracking of humans in action: a multi-view approach, In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, San Francisco, 1996.
21. Gavrilu, D. und Fritsche M., persönliche Kommunikation, DaimlerChrysler Research, Ulm, 2001.
22. Gavrilu, D.M., The Visual Analysis of Human Movement: A Survey. Computer-Vision and Image Understanding, 73(1), S. 82-98, 1999.
23. Gavrilu, D.M., Vision-based 3-D Tracking of Humans in Action, Dissertation an der University of Maryland, 1996.
24. Geuß, H., Entwicklung eines anthropometrischen Meßverfahrens für das CAD-Menschmodell RAMSIS, Dissertation, Lehrstuhl für Ergonomie, Technische Universität München, 1994.
25. Gilbreth, F.B. und Gilbreth L.M.: Ermüdungsstudium, Eine Einführung in das Gebiet des Bewegungsstudiums, Verlag des Vereins deutscher Ingenieure, Berlin, 1920.
26. Hansen, C., Modellgetriebene Verfolgung formvariabler Objekte in Videobildfolgen, Dissertation, Institut für Informatik, Technischen Universität München, 2002.

27. Haritaoglu I., Harwood D. und Davis, L.S., W4: Who? When? Where? What? A Real Time System for Detecting and Tracking People, In: Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition, Nara, 1998.
28. Heinrich, S., Real time fusion of motion and stereo using flow/depth constraint for fast obstacle detection, In: Proceedings of the 24<sup>th</sup> DAGM Symposium on Pattern Recognition, Zürich, 2002.
29. Hilf, H.H., Einführung in die Arbeitswissenschaft, de Gruyeter, Berlin, S. 28f, 1976.
30. Hogg, D., Model-based vision: a paradigm to see a walking person. Image and Vision Computing, 1(1), 1983.
31. Jiang, X. und Bunke, H., Dreidimensionales Computersehen, Springer Berlin Heidelberg, 1997.
32. Jähne, B., Digitale Bildverarbeitung, 4. Auflage, Springer Berlin Heidelberg, 1997.
33. Kamusella, Ch., Tommy – ein Konzept zur 3D Simulation des menschlichen Körpers, Wissenschaftliche Zeitschrift der Humboldt-Universität zu Berlin, R. Med. 38, 1989.
34. Karwowski, W., Genaidy, A.M., und Asfour, S. S. (Eds.), Computer-Aided Ergonomics, pp. 138 – 156, Taylor & Francis, London, 1990.
35. Kaibel, J., Kubbat, W., Albert, O., Balzulat, J. und Seitz, T., Posture Prediction Model for Pilots in Cockpits and the Appliance on the Human Model RAMSIS, VDI/SAE International Conference on Digital Human Modeling for Design and Engineering, VDI-Berichte Nr. 1675, München, 2002.
36. Kinzel, W., Präattentive und attentive Bildverarbeitungsschritte zur visuellen Erkennung von Fußgängern, Nummer 329 in Fortschr.-Ber. VDI Reihe 10, VDI-Verlag, Düsseldorf, 1994.
37. Kolling, J., Validierung und Weiterentwicklung eines CAD-Menschmodells für die Fahrzeuggestaltung, Dissertation, Lehrstuhl für Ergonomie, Technischen Universität München, 1998.
38. Léna, P., Observational Astrophysics, Springer Berlin Heidelberg, 1988.
39. McDaniel, J.W., Human Modeling: Yesterday, Today, and Tomorrow, SAE International Conference, Digital Human Modeling for Design and Engineering, Dayton, Ohio, 1998.
40. Mochimaru, M. und Yamazaki, N., The marker-less measurement method of the two-dimensional human motion, In: Proceedings of the 10<sup>th</sup> Conference of the ESB, 1996.

41. Mochimaru, M. und Yamazaki, N., The three-dimensional measurement of unconstrained motion using a model matching method, *Ergonomics*, 37(3), S. 493-510, 1994.
42. Morguet, P. und Lang, M., Comparison of Approaches to Continuous Hand Gesture Recognition for a Visual Dialog System, *Proceedings of ICASSP 99*, Phoenix, S. 3549-3552, 1999.
43. Oberseminar „Sprache, Mimik, Gestik“, persönliche Kommunikation, Wintersemester 2000/2001, Institut für Phonetik, Ludwig-Maximilians-Universität München.
44. Peacock, B. und Karwowski, W., *Automotive Ergonomics*, Taylor & Francis, Washington, D.C., 1993.
45. Pentland, A., Smart rooms, *Scientific American*, 274(4), S. 54-62, 1996.
46. Press, W. et al., *Numerical Recipes in C*, Cambridge University Press, 1995.
47. Produkt „Contour“ der Human Solutions GmbH, [www.human-solutions.com](http://www.human-solutions.com)
48. Quesada, P.M., Video-based measurements of human movement, *The Occupational Ergonomics Handbook*, CRC Press LCC, 1999.
49. Radig, B., et al., A Model-Driven Three-Dimensional Image-Interpretation System Applied to Person Detection in Video Images. In: Jähne, B., Haußecker, H. und Geißler, P. (Hrsg.), *Handbook of Computer Vision and Applications*, Band 3: Systems and Applications, Kapitel 22, Academic Press, San Diego, New York, Boston, London, Sydney, Tokyo, 1999.
50. Raschke, U., Schutte, L. und Chaffin, D.B., Simulating Humans: Ergonomic Analysis in Digital Environments, in Salvendy, G. (Ed.), *Handbook of Industrial Engineering*, J. Wiley & Sons, New York, 2001.
51. Rehg, J., und Kanade, T., Visual tracking of high dof articulated structures: an application to human hand tracking. In: *European Conference on Computer Vision*, S. 35-46, 1994.
52. Remlinger, W., Using RAMSIS in automotive conception and design at AUDI, *RAMSIS User Conference*, Technische Universität München, Garching, 2001.
53. Ridder, Ch., Interpretation von Videobildfolgen zur Beobachtung artikularer Bewegung von Personen anhand eines generischen 3D Objektmodells, *Dissertation*, Technische Universität München, 2000.

54. Rohr, K., Towards Model-Based Recognition of Human Movements in Image Sequences, *Computer Vision, Graphics and Image Processing: Image Understanding*, 59 (1), S. 94-115, 1994.
55. Ryan, P.W. und Springer, W.E., Cockpit Geometry Evaluation Final Report, Vol. V, JANAIR Report 69105, Office of Naval Research, Washington, D.C., 1969.
56. Scharr, H., Diplomarbeit, Ruprecht-Karls-Universität Heidelberg, 1996.
57. Schaub, K., Entwicklung eines modularen, rechnergestützten, dreidimensionalen man models mit interaktiver Anpassung an die Arbeitsgestaltungsaufgabe. Schriftenreihe Fortschrittsbericht Nr. 51, VDI-Reihe 17: Biotechnik, Düsseldorf, VDI-Verlag, 1988.
58. Schmidtke, H., *Ergonomie*, 3. Auflage, Hanser, München, 1993.
59. Seidl, A., Automatisierte Haltungsanalyse des Menschen, Diplomarbeit, Lehrstuhl für Ergonomie, Technische Universität München, 1989.
60. Seidl, A., Das Menschmodell RAMSIS – Analyse, Synthese und Simulation dreidimensionaler Körperhaltungen des Menschen, Dissertation, Lehrstuhl für Ergonomie, Technische Universität München, 1994.
61. Seitz, T., The optical measurement system PCMAN, Workshop on 3-D Anthropometry and Industrial Products Design, Editor A. Coblentz, Paris, 1998.
62. Seitz, T., Balzulat, J. und Bubb, H., Anthropometry and measurement of posture and motion, *Int. J. of Industrial Ergonomics* 25, 2000.
63. Seitz, T. und Bubb, H., Human model-based movement-capturing without markers for ergonomic studies, *Proceedings of the SAE Conference on Digital Human Modeling for Design and Engineering*, Washington D.C., 2001.
64. Seitz, T., PCMAN – ein Messsystem nicht nur zur Analyse von Fahrerarbeitsplätzen in Gabelstaplern, *REFA-Nachrichten*, Heft 6, 2002
65. Seitz, T., Garcia-Gil, H., Stüdeli, T. und Menozzi, M., Digital image processing for the determination of body movements in young children, Konferenzbeitrag zur Internationalen Frühjahrstagung der Gesellschaft für Arbeitswissenschaft, München, 2003.
66. Sperling, G., The information available in brief visual presentations, *Psychological Monographs*, 74: No. 498, 1960.
67. Speyer, H., Wirsching, H.-J., Devolder, S. und van Roemdonk, R., ALERT - Motion Capturing using a Suit, *VDI/SAE International Conference on Digital Human Modeling for Design and Engineering*, VDI-Berichte Nr. 1675, München, 2002.

68. Taylor, F.W., The Principles of Scientific Management, Harper and Bros., New York, 1929.
69. Ullmann, S., The Interpretation of Visual Motion, The Massachusetts Institute of Technology, 1979.
70. Wachter, S. und Nagel, H.-H., Tracking Persons in Monocular Image Sequences, Computer Vision and Image Understanding, 74 (3), S. 174-192, 1999.
71. Wachter, S., Verfolgung von Personen in monokularen Bildfolgen, Vice Versa Verlag, Berlin, 1997.
72. Weiß, D., Entwicklung eines Systems zur Analyse menschlicher Bewegung, Diplomarbeit, Lehrstuhl für Ergonomie, Technische Universität München, 1985.
73. Werner, C., Stabilisierung von Bewegungsanalyse mit Hilfe von Kanteninformation, Diplomarbeit, Lehrstuhl für Ergonomie, Technische Universität München, in Vorbereitung.
74. Wessels, M., G., Kognitive Psychologie, Ernst Reinhardt Verlag, München, 1994.
75. Wren, C.R. und Pentland, A.P., Dynamic Models of Human Motion, In: Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition, Nara, 1998.
76. [www.simi.com](http://www.simi.com)
77. [www.vicon.com](http://www.vicon.com)
78. Yamamoto, M. et al., Incremental Tracking of Humans in Action from Multiple Views, In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, Santa Barbara, 1998.
79. Yamamoto, M. et al., Human Action Tracking Guided by Key-Frames, In: Proceedings of the Fourth International Conference on Automatic Face and Gesture Recognition, Grenoble, 2000.
80. Zauner, C., Integration of the man model RAMSIS into MBS software Simpack – model comparison and biomechanical simulation for verification, Master Thesis, Lehrstuhl für Ergonomie, Technische Universität München, 2002.

## 7 Anhang

Exemplarisch werden hier zwei Bewegungssequenzen dargestellt, die mit Hilfe des dargestellten Verfahrens gemessen wurden.

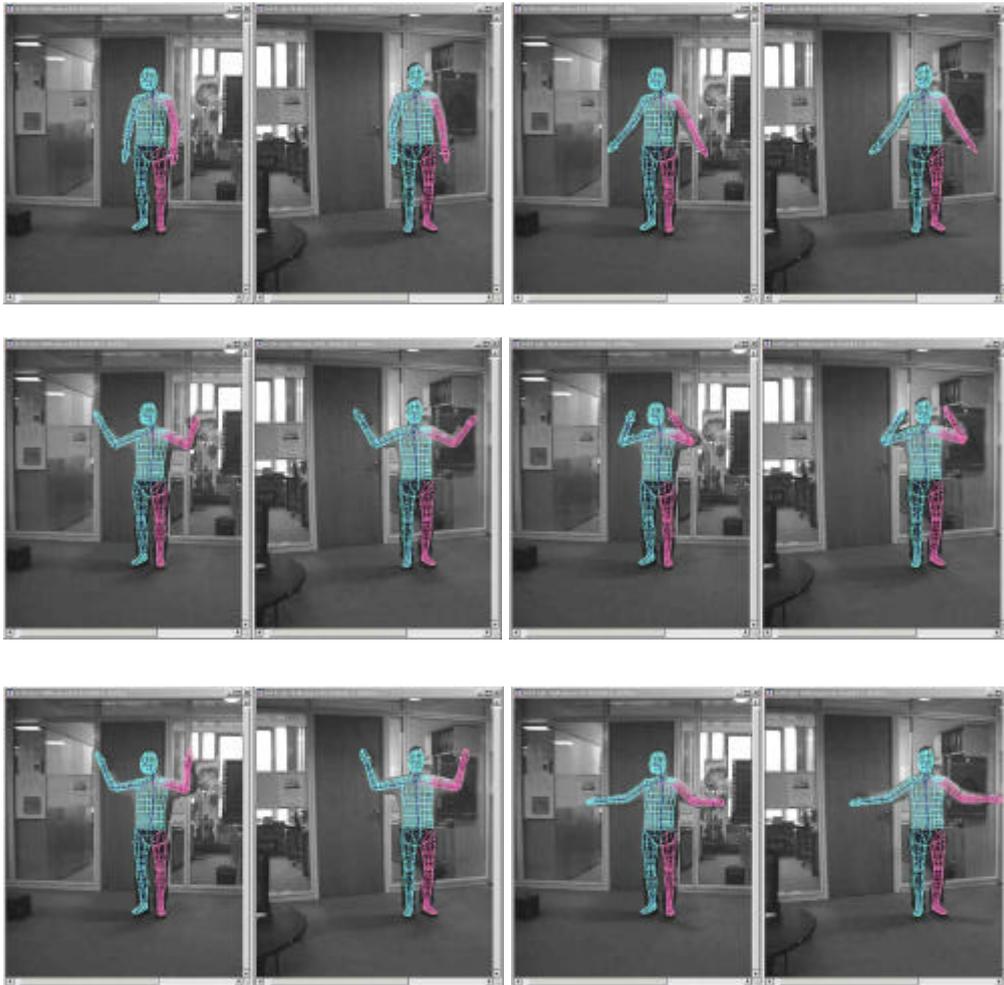


Abbildung 7.1: Beispiel Armbewegung.



Abbildung 7.2: Beispiel Schaltvorgang im KFZ.

