

Ablauf eines *Motion Capture* - Prozesses

Peter Votruba

Abstract

Diese Arbeit beschreibt den Ablauf eines kompletten Motion Capture - Prozesses eines markerbasierten optischen Systems. Nach einem kurzen Überblick über die verschiedenen Arten, wird auf die einzelnen Arbeitsschritte, einschließlich Vorbereitung und Nachbearbeitung, eingegangen.

1. Einleitung

Motion Capture findet heutzutage große Verwendung bei der Herstellung von Computerspielen (z.B. *Legends of Might and Magic*), um die Bewegungen computergenerierter Charaktere realistischer zu gestalten. Auch in der Filmindustrie wird diese Technik immer mehr verwendet - bei Special-Effects von Actionfilmen (z.B. *Total Recall*) oder bei komplett computergenerierten Filmen (z.B. *Toy Story*).

Was ist eigentlich *Motion Capture*?

“Motion capture involves measuring an object's position and orientation in physical space, then recording that information in a computer-usable form.” [2]

Motion Capture nimmt Bewegung auf und überträgt diese in eine computer-nutzbare Form, durch Verfolgung von Schlüsselpunkten im 3-dimensionalen Raum. [1]

Es ist zwar genauso möglich die Bewegung von Tieren zu messen. Da aber bei Computer Spielen hauptsächlich *Human Motion Capture* - die Aufnahme menschlicher Bewegung - angewandt wird, beschränkt sich dieses Paper darauf. Dieses Computermodell kann dann auf beliebige virtuelle Charaktere angewandt werden, um diese zu animieren.

Motion Capture ist trotz der rasanten technischen Entwicklung ein komplexer und kostspieliger Vorgang, bei dem es einige Punkte zu beachten gibt. Dieses Paper wird daher im folgenden die wichtigsten Arbeitsschritte eines *Human Motion Capture* - Prozesses beschreiben und die Probleme, die dabei auftreten können, erläutern.

2. Vergleich von *Motion Capture* - Systemen

Motion Capture - Systeme lassen sich nach 3 Gesichtspunkten in Kategorien einteilen:

direkte Weiterverwendung möglich?: [3]

- ***real-time*** oder ***on-line***: Das Ergebnis kann weiterverwendet werden um direkt virtuelle Charaktere zu steuern. Beispiel: elektromagnetische *Motion Capture* - Systeme.
- ***non-real-time*** oder ***off-line***: Die Auswertung der aufgenommenen Bewegung erfolgt in einem separaten Arbeitsschritt - dafür ist das Endergebnis genauer und es können größere und komplexere Bewegungen aufgenommen werden. Beispiel: optische *Motion Capture* - Systeme.

Position der Sensoren und Datenquellen: [1, 6]

- **outside-in:** externe Sensoren ermitteln Daten von Datenquellen, die am Körper der Akteure platziert sind, z.B. optische *Motion Capture* - Systeme, wobei die Kameras die Sensoren sind und die reflektierenden Marker die Datenquellen
- **inside-out:** Sensoren befinden sich am Körper der Akteure, z.B. elektromagnetische *Motion Capture* – Systeme.
- **inside-in:** Sensoren und Datenquellen befinden sich am Körper der Akteure, z.B. elektromechanische *Motion Capture* - Systeme

Art der Technik: [1, 2, 7]

- **optisch:** Basiert generell auf mehreren Kameras, die den Akteur von mehreren Seiten aufnehmen, um so seine 3-dimensionale Position zu bestimmen. Nachteil: nicht Echtzeit.
 - markerbasiert: Am Körper der Akteure sind reflektierende Marker angebracht, deren Position durch die Kameras bestimmt werden. Vorteil: hohe Genauigkeit.
 - markerlos: Ein relativ neuer Ansatz bei optischen *Motion Capture* - Systemen, bei dem versucht wird die Position und die Körperhaltung des Akteurs durch seinen Umriss zu identifizieren. Vorteil: größtmögliche Bewegungsfreiheit.
- **elektromagnetisch:** Sensoren am Körper bestimmen die Position in einem elektromagnetischen Feld. Vorteil: Echtzeit. Nachteil: Probleme bei metallischen Gegenständen im zu messenden Raum.
- **elektromechanisch:** Die Akteure tragen einen speziellen Anzug mit Sensoren (Potentiometer) an den Gelenken, welche die Bewegung der Gelenke messen. Vorteil: Sensoren werden nie verdeckt oder behindert. Nachteil: starke Einschränkung der Bewegungsfreiheit.

Einen ausführlichen Überblick über die verschiedenen *Motion Capture* - Systeme bietet die Arbeit "Human Motion Capture" von Christiane Ulbricht.

Aufgrund der hohen Genauigkeit und der großen Bewegungsfreiheit für die Akteure gehört optischen *Motion Capture* - Systemen die Zukunft – zumindest bei nicht-echtzeit Anwendungen. Daher wird dieses Paper die wichtigsten Schritte anhand von optischem *Motion Capture* erklären.

3. Überblick über optisches *Motion Capture* [1]

Optisches *Motion Capture* liefert zwar im Vergleich zu anderen Methoden die genauesten Ergebnisse, aber - noch - nicht in Echtzeit, da eine umfangreiche Nachbearbeitung der aufgenommenen Daten notwendig ist.

Ein typisches System besteht neben einem Computer aus mindestens 4-6 Kameras und 20-30 Markern. Bei den Markern wird zwischen aktiven und passiven Markern unterschieden. Bei passiven Markern - meist kleine, mit reflektierendem Klebeband überzogene Kugeln - benötigen die Kameras eigene (Infrarot-)Lichtquellen, damit die Marker erkannt werden können. Aktive Marker strahlen von selbst Licht aus - brauchen daher aber auch eine eigene Stromversorgung.

Ein Beispiel für ein markerbasiertes optisches *Motion Capture* - System ist *Vicon 8* [8], das bis zu 24 Kameras unterstützt und speziell für Computergraphik Anwendungen entwickelt wurde.

4. Vorbereitungen [1]

Der allererste Schritt in der Vorbereitungsphase einer *Human Motion Capture* - Aufnahme besteht darin, sich genau zu überlegen, welche Bewegungen benötigt werden und dafür jeweils geeignete Akteure zu finden. Weiters sollten, um störende Reflektionen bei der Aufnahmen zu vermeiden, die Beleuchtung, der Hintergrund und die Kleidung der Akteure aufeinander abgestimmt werden.

Wenn die computergenerierte Figur in der geplanten Animation, mit Objekten interagieren soll, muss dies bereits in der Vorbereitungsphase berücksichtigt werden. Dabei unterscheidet man, je nachdem ob sich die Objekte bewegen (bzw. bewegt werden), zwischen statischen und dynamischen Objekten.

Die spätere Position von statischen Objekten muss für den Akteur gekennzeichnet werden, z.B. durch Markierungen am Boden des Aufnahmestudios, damit beispielsweise der Akteur die computergenerierte Figur nicht durch eine Wand gehen lässt. Solche Kennzeichnungen, wie Bodenmarkierungen, werden ebenfalls verwendet, wenn der Akteur spezielle Bewegungen genau ausführen soll, z.B. eine bestimmte Kurve.

Um dynamische Objekte, wie z.B. Gegenstände, die die computergenerierte Figur in der Hand halten soll, realistischer darstellen zu können, müssen diese, genauso wie der Akteur, mit Markern ausgestattet werden (siehe Abbildung 1).



Abbildung 1 - Akteurin und Gegenstand (Waffe) mit Markern [1]

Vor Beginn jeder Aufnahme müssen die Kameras kalibriert werden. Dazu lässt man die Kameras ein Objekt mit Markern aufnehmen, dessen Abmessungen bekannt sind. Aus den Kamerabildern und den Objektabmessungen kann die exakte Position jeder Kamera bestimmt werden, die später benötigt wird, um die Koordinaten der Marker zu berechnen.

Ein weiterer wichtiger Punkt in der Vorbereitungsphase ist die Positionierung der Marker am Körper des Akteurs. Meistens wird dabei aber ein Standardmarkermodell (siehe Abbildung 2) verwendet, das alle Bewegungsmöglichkeiten des menschlichen Körpers abdeckt.

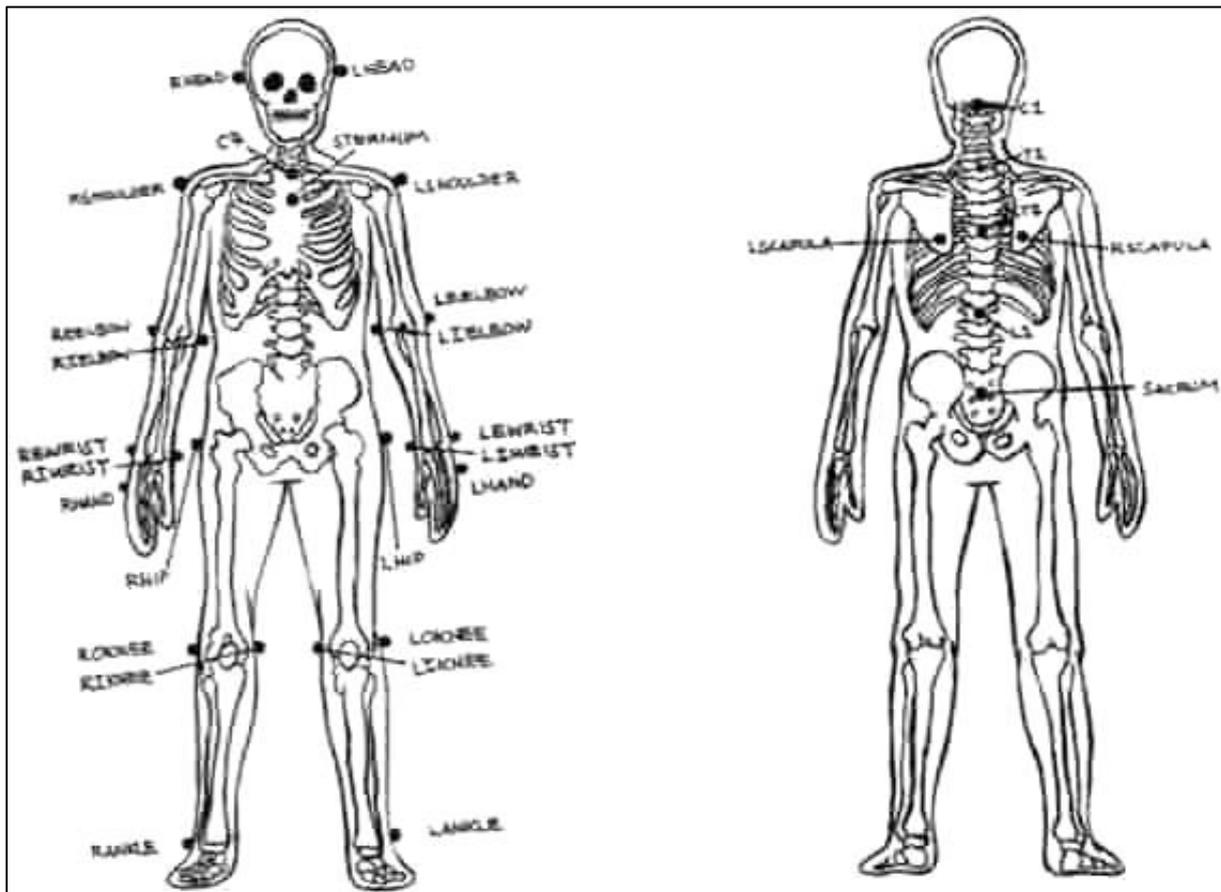


Abbildung 2 - Skelett mit Markerpositionen [1]

Diese Markerkonfiguration besteht aus 31 Markern - je einer an jedem Fußgelenk, 2 am Knie, je einer an jeder Seite der Hüfte, einer an jeder Hand, 2 an jedem Handgelenk und Ellenbogengelenk, einer auf jeder Schulter, einer am Brustbein und Hals, 3 am Kopf (Schläfen und Rückseite) und 6 am Rücken (Nacken, Schulterblätter, Kreuzbein und 3 an der Wirbelsäule).

Diese Anordnung ist speziell für Ganzkörper - *Motion Capture* - Aufnahmen gedacht, um die Bewegungen jedes Körperteils messen zu können. Detailaufnahmen von Händen oder vom Gesicht des Akteurs müssen mit eigenen Markerkonfigurationen (und kleineren Markern) realisiert werden.

In Abbildung 3 ist der Zusammenhang zwischen den beweglichen Körperteilen und den dazugehörigen Markern, die jeweils ein Dreieck bilden, dargestellt. Durch Gruppierung bestimmter Marker können auch die Bewegungen von "inneren" Gelenken (Wirbelsäule) gemessen werden.

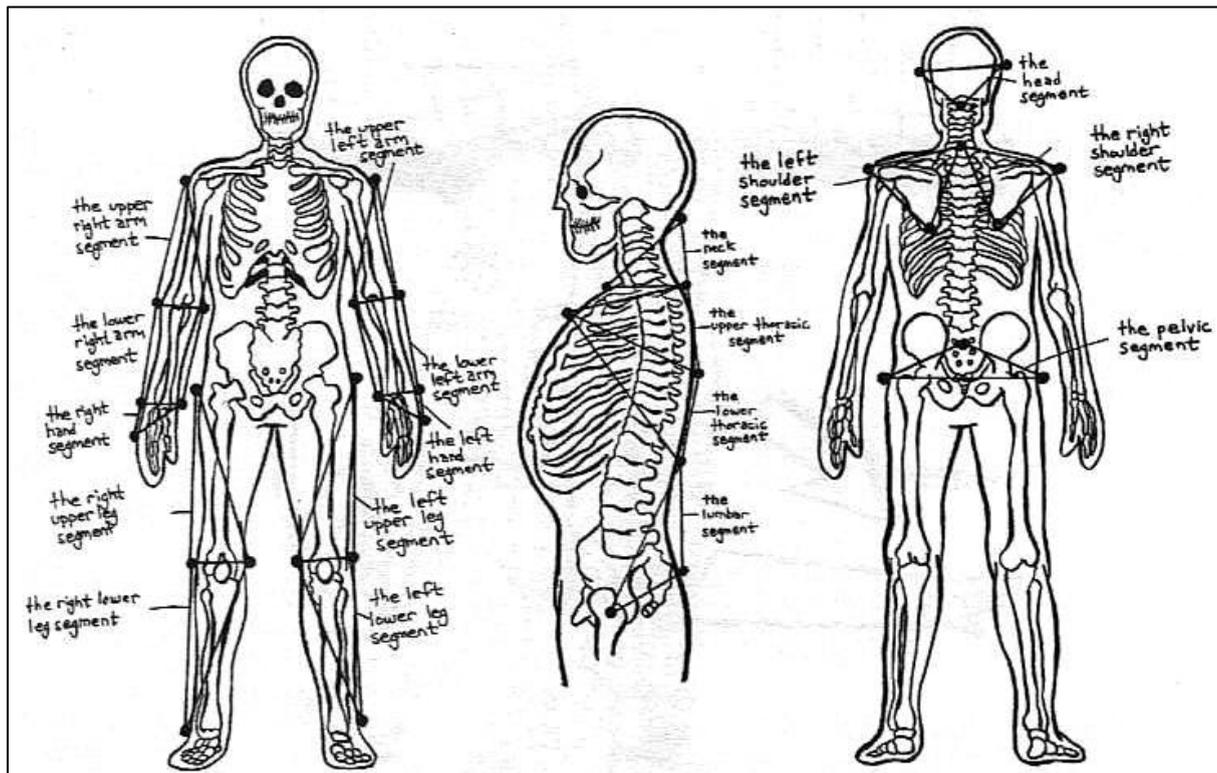


Abbildung 3 - Knochengruppen, die durch die Markerkonfiguration abgedeckt sind [1]

Ausschlaggebend für eine realistische Markerkonfiguration ist, unter anderem, das Verstehen der biomechanischen Eigenschaften des Armes. So ist zum Beispiel für das Drehen der Hand nicht das Ellenbogen- oder das Schultergelenk verantwortlich, sondern das Kreuzen der Speiche über die Elle. Daher sind, um die komplexeren Bewegungen des Unterarms zu messen, 4 Marker notwendig (siehe Abbildung 4).

Beim Anbringen der Marker sollte man darauf achten, dass sich möglichst wenig Abstand zwischen den Markern und den Knochen befindet. Nachdem alle Marker, nach vorher festgelegter Markerkonfiguration, am Körper des Akteurs positioniert sind, müssen die genauen Positionen der Marker, gemeinsam mit den Körpermaßen, vermerkt werden. Diese Daten werden später in der Nachbearbeitungsphase gebraucht.

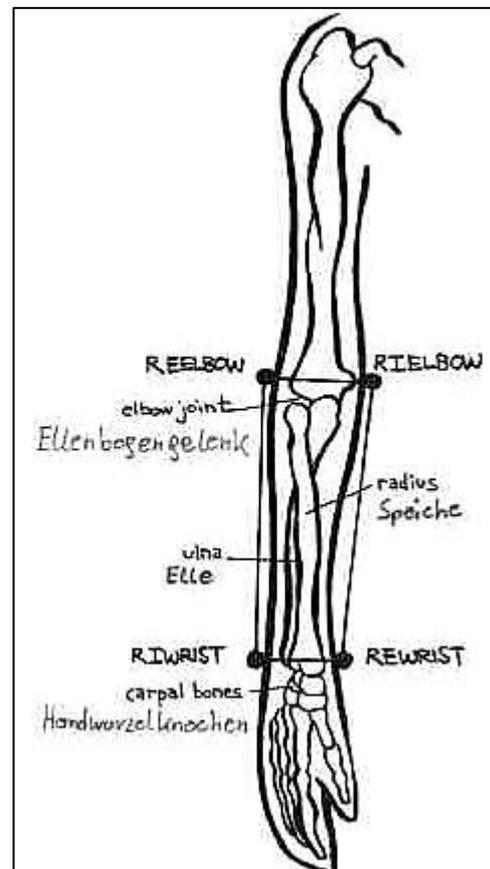


Abbildung 4 - Unterarm mit Marker [1]

5. Aufzeichnen der Bewegungen

Dieser Schritt ist zwar der wichtigste in einem *Motion Capture* - Prozess, aber die Hauptverantwortung für das Gelingen liegt hauptsächlich beim Akteur. Während der Aufnahme der Bewegungen ist vor allem darauf zu achten, dass zu viele Marker verdeckt sind. Außerdem müssen regelmäßig die Markerkonfiguration und die Kamerapositionen überprüft und gegebenenfalls rekali­briert werden.

6. Nachbearbeitung [1]

Die Nachbearbeitung ist die aufwändigste Phase. Nach der Aufnahme der Bewegungen müssen die Bilder der einzelnen Kameras in ein Modell eines menschlichen Skeletts umgewandelt werden. Dabei werden vom *Motion Capture* - System der Reihe nach folgende Schritte durchgeführt:

1. Marker extrahieren

Der erste Schritt besteht darin, eine saubere Wiedergabe der Bewegung zu produzieren, die nur einzelne Marker enthält. Verschiedene Bildverarbeitungs­methoden werden angewandt, um jeden Marker von der Umgebung und anderen Markern zu isolieren. Dabei wird zuerst das Bildrauschen mit Hilfe von Glättungsfiltern reduziert. Danach wird versucht, die Marker zu lokalisieren, indem nach Gruppen von Pixeln gesucht wird, deren Grauwert über einem bestimmten Schwellwert liegt. Je heller der Grauwert der Pixel ist, desto größer ist die Wahrscheinlichkeit, dass es sich dabei um die Reflektionen eines Markers handelt. Die Größe der auf diese Art ermittelten Pixelgruppen wird danach mit der tatsächlichen Größe der Marker verglichen. Unterscheiden sich diese Werte zu stark, so kann davon ausgegangen werden, dass es sich bei dieser Pixelgruppe nicht um einen Marker handelt. Trotzdem ist nach diesem Vorgang nicht sichergestellt, dass es sich bei den erkannten Pixelgruppen tatsächlich ausschließlich um Marker handelt. Reflektionen, die während des Messvorganges ungewollt auftreten, erschweren die Rekonstruktion der Markerpositionen. Ein weiteres Problem, das während des Rekonstruktionsvorganges ergeben kann, ist die Verdeckung von Markern, wodurch auf einem Einzelbild nicht alle Marker lokalisiert werden können. Aus diesem Grund ist es wichtig, eine ausreichende Anzahl von Kameras zu verwenden, damit jeder Marker auf zumindest zwei Einzelbildern identifiziert werden kann. Das Ergebnis dieses Schrittes ist für jede Kamera eine Folge von Bildern auf denen sich nur noch die Marker befinden (siehe Abbildung 5).

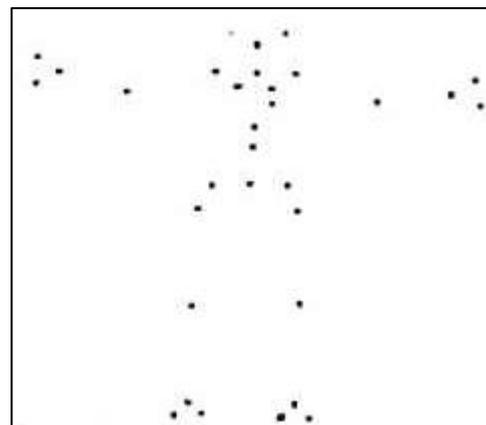


Abbildung 5 - Beispiel einer Punktwolke

2. Koordinaten der Marker berechnen

Im zweiten Schritt wird das Ergebnis des ersten Schrittes hergenommen und die Position für jeden sichtbaren Marker bestimmt. Dazu werden zuerst auf jedem Bild die 2-dimensionalen Koordinaten jedes Markers berechnet.

Danach muss versucht werden, auf je 2 Kamerabildern die korrespondierenden Markerabbildungen zu finden (*stereo marker matching*). Da dieses *correspondence problem* nicht eindeutig ist, werden verschiedene *constraints* angewandt. Das wichtigste davon ist das sogenannte *epipolar constraint* (siehe Abbildung 6):

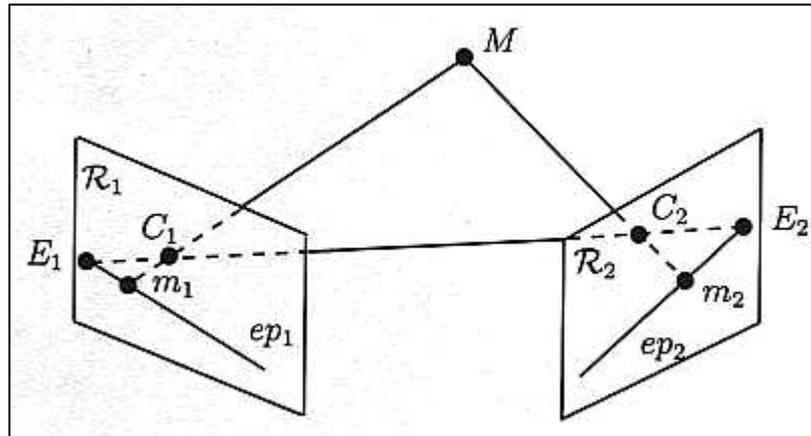


Abbildung 6 - *epipolar constraint*

Der 3-dimensionale Punkt M , der die 2-dimensionale Abbildung m_1 in der Bildebene R_1 erzeugt hat, muss auf der Geraden $m_1 - C_1$ liegen, wobei C_1 das optische Zentrum der linken Kamera in der Abbildung darstellt. Daraus folgt, dass alle möglichen zu m_1 korrespondierenden Punkte m_2 der Bildebene R_2 auf der Abbildung dieser Geraden liegen. Diese Abbildung, die Gerade ep_2 , ist die sogenannte *epipolar line*. Das *epipolar constraint* besagt, dass für einen gegebenen Punkt m_1 , alle dazugehörenden Punkte m_2 der Bildebene R_2 auf der *epipolar line* liegen. Dadurch wurde die Suche von einem 2-dimensionalen auf einen 1-dimensionalen Bereich eingeschränkt.

Aus den Ergebnissen aller Kameras zusammen mit den Kamerakoordinaten können dann die 3-dimensionalen Koordinaten jedes Markers, durch Schnitt der Geraden $m_1 - C_1$ und $m_2 - C_2$, berechnet werden.

Dieser Schritt macht aus jedem aufgenommenen Bild jeder Kamera eine Punktwolke, deren Punkte die Marker repräsentieren.

3. Marker identifizieren (3D marker matching)

In diesem Schritt wird versucht die einzelnen Punkte der Punktwolken realen Markern zuzuordnen. Dieser Schritt erfordert die meiste menschliche Unterstützung, da jeder Marker einmal manuell identifiziert werden muß. Danach versucht die Software, den Rest der Sequenz selber zuzuordnen. Dabei kann es leicht zu Problemen kommen, wenn einzelne Marker fehlen, weil sie in zu vielen Kamerasichten verdeckt waren. Dann muß versucht werden die Position, der fehlenden Marker zu rekonstruieren. Um das nicht jedesmal manuell machen zu müssen, gibt es verschiedene Lösungsansätze, die es dem System ermöglichen die Marker ohne menschlicher Hilfe zu bestimmen.

Eine Möglichkeit ist es, die Punktwolken der Marker mit einem biomechanischen Modell des Körpers zu vergleichen (*fitting*) [5].

4. Positionen der Gelenke bestimmen

Bevor man die gewonnenen Daten auf eine computergenerierte Figur anwenden kann, muss man aus den Koordinaten der Marker die Positionen der Gelenke berechnen. Dabei wird auf die, in der Vorbereitungsphase vermessenen, Markerpositionen und auf die Körpermaße des Akteurs zurückgegriffen. Das Ergebnis ist dann das Modell eines menschlichen Skeletts.

5. Übertragen des Modells auf die computergenerierte Figur (*retargeting*)

Wenn der Akteur und die computergenerierte Figur verschieden groß sind, muss das Computermodell zuerst noch skaliert werden. Dabei kann es zu Problemen kommen, falls sich die Proportionen der Körperteile des Akteurs und der Computerfigur stark unterscheiden. Wenn eine computergenerierte Figur (z.B. ein Monster in einem Computerspiel) mit langen Armen in der Animationsszene einen Gegenstand nehmen soll, kann es passieren, dass sie an dem Gegenstand vorbeigreift. Um das zu vermeiden, werden beim Skalieren Verzerrungen zugelassen. Dadurch wirken die Bewegungen dann aber nicht mehr so realistisch.

7. Fazit und Ausblick

Es gibt verschiedene Ansätze um die Probleme, die bei optischen *Motion Capture* auftreten können, zu minimieren. Die meisten Probleme treten bei der Identifizierung der Marker (Abschnitt 6.3) auf, vor allem, wenn Marker verdeckt werden. Neben mathematischen Methoden gibt es auch innovative technische Lösungsansätze für dieses Problem. Ein Beispiel dafür ist "*ReActor*" von *Ascension* [9], ein optisches *Motion Capture* - Verfahren, das aktive Marker verwendet, wobei zu einem bestimmten Zeitpunkt, jeweils nur ein "Marker" aufleuchtet. Da daher auf jedem Kamerabild nur jeweils ein Marker zu sehen ist, entfällt der aufwändige und fehleranfällige Schritt der Markeridentifizierung. Ein großer Nachteil dabei ist, dass sich die Anzahl der Marker auf die Aufnahmegeschwindigkeit (*Sampling-Rate*) auswirkt, d.h. je mehr Marker verwendet werden, umso niedriger ist die *Sampling-Rate*. Ein weiterer Nachteil ist auch die Stromversorgung der aktiven Marker.

Verbesserungsansätze bzw. Forschungsziele für optisches *Motion Capture*: [4]

- Vergrößerung des Aufnahmebereichs
- Minimierung der manuellen Nachbearbeitung - Ziel wäre ein vollautomatisches optisches *Motion Capture* - System
- Minimierung des Aufwandes bei der Kalibrierung bzw. Rekalibrierung
- Erhöhung der Geschwindigkeit bei der Nachbearbeitung: um optisches *Motion Capture* auch bei Echtzeitanwendungen, wie z.B. Interaktion bei Computerspielen, einsetzen zu können
- Senkung der Kosten: das wohl beste Mittel um die Verbreitung von *Motion Capture* zu fördern
- Möglichkeit, die Bewegung mehrerer Akteure aufzunehmen: besonders optische *Motion Capture* - Systeme sind in diesem Punkt stark eingeschränkt

Die Zukunft gehört aber sicher dem markerlosen optischen *Motion Capture*, da es keinerlei Hilfsmittel, wie z.B. Sensoren oder Marker, am Körper der Akteure benötigt. Dadurch sind viel realistischere Aufnahmen möglich, da die Akteure in ihren Bewegungen vollkommen uneingeschränkt sind.

8. Referenzen

Bücher:

- [1] A. Menache (2000, Morgan Kaufmann):
"Understanding Motion Capture for Computer Animation and Video Games"

Papers:

- [2] S. Dyer, J. Martin, J. Zulauf (1995):
"Motion Capture White Paper"
http://reality.sgi.com/jam_sb/mocap/MoCapWP_v2.0.html
- [3] L. Herda, P. Fua, R. Plänkers, R. Boulic, D. Thalmann:
"Skeleton-Based Motion Capture for Robust Reconstruction of Human Motion"
- [4] M. Furniss (1999):
"Motion Capture"
<http://media-in-transition.mit.edu/articles/furniss.html>
- [5] "Using Bio-mechanical constraints for better fitting of body models to dynamic 3d data"
<http://ligwww.epfl.ch/~lorna/MyProposal.htm>

Webseiten:

- [6] A. Mulder: "Human movement tracking technology"
<http://www.cs.sfu.ca/~amulder/personal/vmi/HMTT.pub.html>
- [7] J. Kolozs: "Position Trackers for Virtual Reality"
<http://www.cs.utah.edu/classes/cs6360/Kolozs/trackers.html>
- [8] V I C O N | Motion Capture
<http://www.vicon.com/animation>
- [9] "ReActor" von Ascension (aktive Marker)
<http://www.ascension-tech.com/products/reactor/index.html>
<http://www.vrnews.com/issuearchive/vrn1003/vrn1003vent.html>

weiterführende Literatur:

- L. Campbell, A. Bobick (1995):
"Recognition of Human Body Motion Using Phase Space Constraints"
- L. Herda, P. Fua, R. Plänkers, R. Boulic:
"Human shape and motion recovery using animation models"
- W. Tang, M. Cavalza, D. Mountain, R. Earnshaw (1999):
"A constrained inverse kinematics technique for real-time motion capture animation"
- R. Bowden, T.A: Mitchell, M. Sarhadi (1999):
"Non-linear statistical models for the 3D reconstruction of human pose and motion from monocular image sequences" (markerlos)
- R. Bowden, T. A. Mitchell, M. Sarhadi (1998):
"Reconstructing 3D Pose and Motion from a Single Camera View"
<http://www.brunel.ac.uk/~emstrrb/publications/bmvc98/h114.htm>
- A. Bottino:
"Non-intrusive Motion Capture" (markerlos)
<http://www.polito.it/~bottino/MotionCapture/>
- U. Neumann, S. You (1999):
"Natural Feature Tracking for Augmented-Reality" (markerlos)
<http://citeseer.nj.nec.com/330642.html>