Markerfreies Motion Capture: Neue Wege zur Analyse menschlicher Bewegungen

Markerless motion capture: new approaches to analysing human motion

Rosenhahn, Bodo

Max-Planck-Institut für Informatik, Saarbrücken

Korrespondierender Autor/in

Zusammenfassung

Mit Motion Capture bezeichnet man die Aufnahme und Analyse menschlicher Bewegungen, die z.B bei medizinischen Fragestellungen eingesetzt werden. Kommerzielle Systeme verwenden dazu künstliche Marker, die an die Person angebracht und mittels Kameras verfolgt und rekonstruiert werden. Von diesen Markern lässt sich dann auf die Position und Orientierung des Menschen, sowie die Konfiguration der Gelenke schließen. Im Gegensatz zu recht zuverlässigen markerbasierten Systemen sind markerfreie Systeme noch nahezu unerforscht und werfen viele ungelöste Fragen aus den Bereichen der Computergraphik, Bildverarbeitung und Biomechanik auf. Der Beitrag beschreibt Forschungsergebnisse, die es ermöglichen, Personen ohne manuelle Intervention, z.B. Marker, erfolgreich über Videodaten zu verfolgen.

Summary

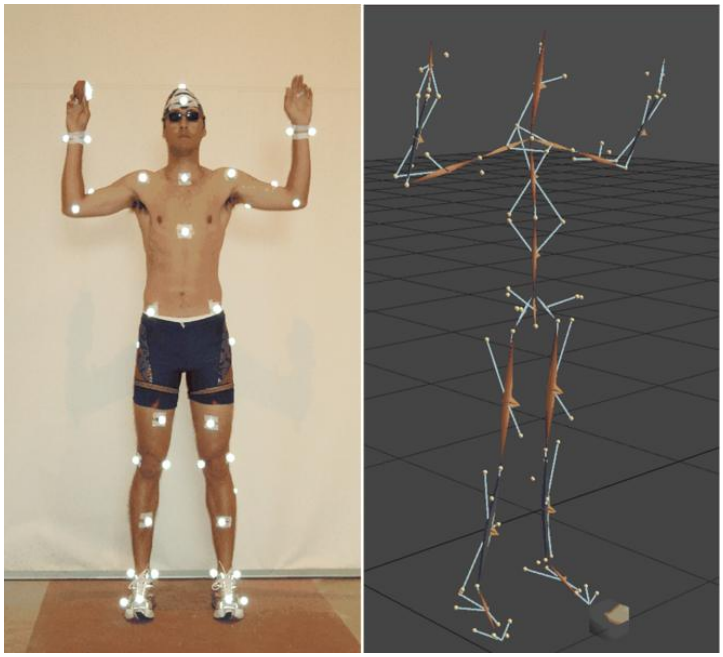
Motion Capture (MoCap) means to record and analyze human movements. Nearly all commercial products in this field rely on artificial markers, which have to be attached to the person. These markers are usually reconstructed from cameras and used to determine the position and orientation as well as the involved joint angles of the person. In contrast to these reliable marker based tracking systems exist so-called marker-less systems subject to recent research in the fields of Computergraphics, Computer vision and Biomechanics. In this article, we describe current aspects of research which allow for pose tracking of persons without manual intervention, e.g. the use of markers, just by using image data.

Einleitung

Viele Anwendungen erfordern die Aufnahme und Analyse menschlicher Bewegungen. Menschliche Bewegungsdaten werden in der Medizin und der modernen Sportwissenschaft untersucht und zur Unterstützung der Diagnosefindung z.B. bei orthopädischen Erkrankungen und ihren Heilungsverläufen oder zur Leistungsoptimierung von Sportlern eingesetzt. Auch in der Film- und Spieleindustrie werden Bewegungsdaten zur Animation von Charakteren in Spielen oder von virtuellen Wesen (King Kong oder Gollum) genutzt. Menschliche Bewegungen aufzunehmen oder zu analysieren erfordert, aus Sensordaten, z.B. Videos, die 3D-Bewegung des Körpers und die Konfiguration der Extremitäten zu schätzen. Um dies zu

© 2007 Max-Planck-Gesellschaft

ermöglichen, werden meistens so genannte markerbasierte Trackingverfahren verwendet, die von kommerziellen Anbietern (z.B. Vicon, Simi, Motion Analysis) bezogen werden können. Markerbasiert bedeutet in diesem Fall, dass Marken, meistens kleine retroflektierende Bälle, an die Person geklebt und in speziellen Aufnahmevorrichtungen von mehreren Kameras verfolgt werden.



Links: Eine mit Markern beklebte Person. Rechts: Die Marker werden von mehreren Kameras beobachtet und über eine Triangulation rekonstruiert. Diese 3D.Punkte werden zu der Rekonstruktion eines virtuellen 3D-Skeletts herangezogen.
© Max-Planck-Institut für Informatik

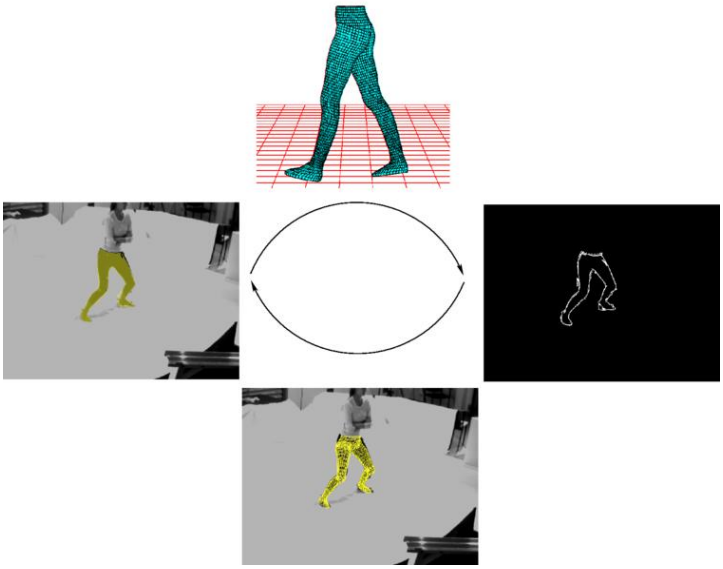
Zur Animation und Analyse menschlicher Bewegungen können dann aus den 3D-Markerpositionen die Gelenkwinkel zurückgerechnet werden (**Abb. 1**). Außerdem können die gewonnenen Positionsdaten auf virtuelle Charaktere (Avatare) übertragen werden, um diese mit Leben zu füllen und beliebig in einer Szene zu bewegen. Auf diese Weise wird ein Avatar, wie z.B. Gollum, mit realistischen Bewegungen animiert.

Markerbasierte Verfahren haben Nachteile, die dazu führen, dass die Produktionskosten für Filme relativ hoch und zeitintensiv werden und sie bei medizinischen Fragestellungen für den klinischen Alltagsgebrauch eher selten eingesetzt werden: Zunächst muss eine Person mit Marker an anatomischen Landmarken beklebt werden. Dieser Prozess kann zwischen 30 und 45 Minuten (bei einem Vollkörpermodell) benötigen und dazu führen, dass sich Sportler wieder aufwärmen müssen. Desweiteren wirken die Marker störend auf das zu analysierende Bewegungsmuster. Insbesondere Probanden, die Marker nicht gewöhnt sind (z.B. ältere Patienten) haben Angst, Marker zu zerstören oder zu verlieren und bewegen sich wesentlich vorsichtiger oder steifer, als sie es in normaler Kleidung würden. Allein diese psychologische Komponente führt zu einem kaum kalkulierbaren Bias der Daten. Weiterhin wird eine spezielle Laborumgebung mit Kameras und Stroboskoplicht benötigt, um die Marker stabil semi-interaktiv verfolgen zu können. Die Analyse von Aufnahmen im Außenbereich ist dadurch nahezu unmöglich.

Am Max-Planck Institut für Informatik arbeitet in der Computergraphik eine Nachwuchsgruppe an Methoden zur Aufnahme und Analyse menschlicher Bewegungen aus Videodaten ohne Hinzunahme externer Marker. Man spricht hier von einem markerfreien Ansatz, der Aufgrund der Verkopplung verschiedener Methoden aus den Bereichen der Computergraphik und des Maschinensehens außergewöhnlich stabil und effizient ist. Insbesondere ist es mit den Verfahren möglich, Personen in Außenbereichen, bei variierender Beleuchtung,

leichten Eigenbewegungen von Kameras oder sogar sehr schnellen Bewegungsmustern, z.B. Sportbewegungen, zu verfolgen. Die wesentlichen Ideen und Vorraussetzungen sollen im Folgenden kurz dargestellt werden.

Das Grundverfahren zur markerfreien menschlichen Bewegungsschätzung

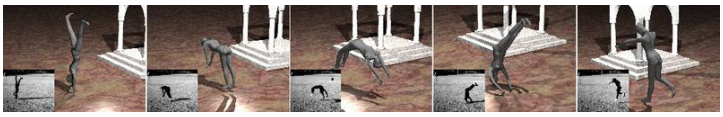


Vorrausgesetzt wird ein Modell der Person als artikulierbare Oberfläche, sowie Bilder synchronisierter Kameras. Das Grundverfahren kombiniert drei wesentliche Schritte: Zunächst wird das Bild segmentiert, dann werden Korrespondenzen zwischen dem Modell und dem Bild hergestellt und am Schluss über die gegebenen Korrespondenzen die Pose berechnet.
© Max-Planck-Institut für Informatik

Vorraussetzung unserer Verfahren sind zunächst Videodaten, die eine Person und ihre Bewegungen in mehreren Kameras parallel aufzeichnen (**Abb. 2**). Die Kameras müssen weiterhin kalibriert sein, d.h. die relative Position der Kameras zueinander und ihre internen Kameraparameter sind als bekannt vorausgesetzt. Weiterhin ist ein Modell der Person gegeben, z.B. in Form eines Oberflächennetzes. Da die Person Gelenke hat, nehmen wir diese ebenfalls als gegeben an um Gliedmassen des Oberflächennetzes adequat mit Gelenken transformieren zu können. Diese Informationen lassen sich a-priori über einen Body-Laserscanner gewinnen.

Ausgehend von den kalibrierten Bilddaten und dem Oberflächenmodell müssen diese zueinander in Bezug gebracht werden. Um dies zu ermöglichen, verfolgt die Nachwuchsgruppe einen so genannten silhouettenbasierten Ansatz, d.h. in einem Segmentierungsschritt wird eine Region im Bild extrahiert, die die Person vom Hintergrund trennt. Hierbei kommen so genannte Levelsetfunktionen zum Einsatz, die so verändert (evolviert) werden, dass Merkmalsverteilungen innerhalb und außerhalb der Regionen maximiert werden. Der Rand der extrahierten Kontur wird in einem anschließenden Registrierungsschritt mit Punkten auf dem Oberflächenmodell korreliert, so dass ein transformiertes Oberflächennetz nach einer sogenannten Poseschätzung möglichst gut mit den gemessenen Bilddaten in Deckung gebracht werden kann [1, 5]. Insbesondere die Integration von Vorwissen über das zu segmentierende Oberflächennetz (man spricht auch von Gestaltsvorwissen oder einem 3D shape prior) ermöglicht eine stabile Konturextraktion auch in komplexen Laborumgebungen. Die folgende **Abbildung 3** zeigt eine Person im Außenbereich, die eine komplexe, dynamische Sprungbewegung durchführt. Die am Max-Planck Institut für Informatik entwickelten Verfahren sind in der Lage, solche komplexen Bewegungen aus Bilddaten zu extrahieren und wiederum zu reanimieren

oder in virtuellen Umgebungen darzustellen.



Eine dynamische Sportsequenz, die von dem markerfreien System erfolgreich verfolgt und rekonstruiert wurde.
© Max-Planck-Institut für Informatik

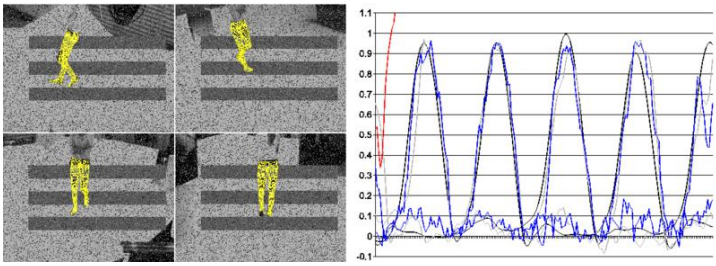
Kleidungsmodelle

Verfolgung bekleideter Personen.
© Max-Planck-Institut für Informatik

Das dargestellte Poseschätzverfahren erfordert einen eng anliegenden Körperanzug. Dies ist bei bestimmten Sportarten durchaus gegeben (z.B. beim Schwimmen), aber bei klinischen Untersuchungen (z.B. an Patienten) meistens unpraktisch. Aus diesem Grund wurde in einem weiteren Forschungsprojekt eine Kleidungssimulation in den Poseschätzprozess integriert. Unter Hinzunahme eines Kleidungsmodells und der sichtbaren Extremitäten ist es möglich, die physikalische Konfiguration der Kleidung und ihre Dynamik aus den Bilddaten zu schätzen sowie die Konfiguration der Gelenke, die durch die Kleidung verdeckt werden (**Abb. 4**). Derartige Verfahren sind weltweit neu und sind ein erster Schritt zu einer Personenanalyse aus Videodaten, in denen die zu verfolgende Person in ihrer natürlichen Umgebung und mit natürlicher Kleidung mit Videokameras verfolgt und ihre Bewegung rekonstruiert wird [2]

Statistische Lernverfahren

Bei vielen Anwendungen ist es möglich, a-priori eine grobe Idee des zu verfolgenden Bewegungsmusters zu erhalten. So werden bei medizinischen Fragestellungen z.B. gerne Gehbewegungen herangezogen. Dieses Vorwissen lässt sich ebenfalls in den Trackingprozess einbinden. So lässt sich z.B. eine Datenbank mit Gehbewegungen heranziehen, um einen Unterraum möglicher Gelenkkonfigurationen zu spezifizieren, die die zu erwartenden Gelenkmuster beinhalten. Es ist z.B. extrem unwahrscheinlich, dass während des Trackings die Beine verdreht werden, die Knie sich durchdrücken oder überschneiden. Um dieses Vorwissen in die Poseschätzung zu integrieren, wird aus einer Trainingsmenge eine Parzendichteverteilung berechnet, die über die gegebenen Trainingsdaten eine Wahrscheinlichkeitsdichte approximiert. Damit werden Gleichungen in einer Optimierungsschleife generiert, die die Gelenkkonfigurationen in „bekannte" Posen lenkt. Dadurch wird ein Soft-Constraint eingeführt, der anatomisch unmögliche Konfigurationen (z.B. der Arm wird durch den Körper geführt) bestraft und gleichzeitig die Gleichungen regularisiert. Außerdem können (auch bei fehlender Bildinformation) nicht sichtbare Gelenke propagiert werden und es kann mit sehr starkem Bildrauschen und Störungen umgegangen werden, wie die folgenden Beispiele zeigen[3].



Das Diagramm (rechts) zeigt einen Vergleich zwischen unserem System und einem kommerziellen markerbasierten Trackingsystem (von Motion Analysis) um eine quantitative Fehleranalyse zu ermöglichen. Die rote Kurve veranschaulicht, dass ohne die Integration des Priors das Tracking fehlschlägt. © Max-Planck-Institut für Informatik

Dies lässt sich nicht nur für statische Konfigurationen implementieren, sondern auch für dynamische Muster, um z.B. geeignete Vorhersagen der Gelenkstellungen bei schnellen Bewegungen durchzuführen oder um Fehler bei der Bildübertragung (Frame drops) zu kompensieren.

Zusammenfassung

Markerfreies Motion Capturing ist ein aktuelles Forschungsgebiet, das interdisziplinäre Kooperationen zwischen den Disziplinen der Computergraphik, Computer Vision und Biomechanik ermöglicht. Am Max-Planck-Institut für Informatik wird seit einigen Jahren erfolgreich ein solches System entwickelt, welches aus Bilddaten und einem gegebenem Modell die Position, Orientierung und die Konfiguration der Gelenke ermittelt [4]. Hierfür wird ein silhouettenbasierter Ansatz verfolgt, bei dem die Kontur des Körpermodells im Bild zur Lageschätzung der Person herangezogen wird. Das hier vorgestellte System ist in der Lage, Personen im Außenbereich erfolgreich zu verfolgen. Insbesondere bei komplexen Hintergründen (Schatten, wechselnde Beleuchtung, nicht homogener Hintergrund) erweist sich die Verwendung von Levelsetfunktionen als sehr hilfreich. In diesem Rahmen wurden noch zwei aktuelle Erweiterungen vorgestellt: (a) das Verwenden von Kleidungssimulatoren und (b) statistische Lernverfahren. Beide Erweiterungen sind für zukünftige Anwendungen wichtig, um Störgrößen bei der Bildaufnahme zu kompensieren und Menschen in ihrer natürlichen Umgebung (mit Ihrer Alltagskleidung) in Zukunft verfolgen zu können.

Danksagung

Die Arbeiten der Nachwuchsgruppe wurden gefördert durch das Max-Planck-Center for Visual Computing and Communication. In diesem Zusammenhang danken die Autoren Kollegen und Studenten, die die Arbeit unterstützt und gefördert haben, u.a. Thomas Brox (Universität Dresden), Daniel Cremers (Universität Bonn), Jürgen Gall (MPI-Inf Saarbrücken), Nils Hasler (MPI-Inf Saarbrücken), Uwe Kersting (Auckland University), Reinhard Klette (Auckland University), Gerald Sommer (Universität Kiel), Christian Schmaltz (Saarland University), Drew Smith (Auckland University), Martin Sunkel (MPI-Inf Saarbrücken) und Joachim Weickert (Saarland University).

Originalveröffentlichungen

Nach Erweiterungen suchenAbsatzBilderweiterungDateilisteHTML-ErweiterungJobtickerKalendererweiterungLinkerweiterungMPG.PuRe-ReferenzMitarbeiter (Employee

Editor)MitarbeiterlistenerweiterungPersonenerweiterungPublikationserweiterungRSS tickerSoziale Medien BoxTaglistenerweiterungTeaser mit BildTextblockerweiterungVeranstaltungstickererweiterungVideoerweiterungYouTube-Erweiterung

[1] **Rosenhahn, B., T. Brox and J. Weickert:**

Three-Dimensional Shape Knowledge for Joint Image Segmentation and Pose Tracking
International Journal of Computer Vision (IJCV) **73**, 243-262 (2007).

[2] **Rosenhahn, B., U. Kersting, K. Powell, R. Klette, G. Klette and H.-P. Seidel:**

A system for articulated tracking incorporating a clothing model
Machine Vision and Applications (MVA) **18**, 25-40 (2007).

[3] **Brox, T. B. Rosenhahn and D. Cremers:**

Contours, Optic Flow, and Prior Knowledge: Cues for Capturing 3D Human Motion in Videos
Human Motion - Understanding, Modelling, Capture and Animation, Springer, Computational Imaging an Vision, **Vol 36**, Rosenhahn, Klette, Metaxas (Eds.), Springer Verlag, Berlin 2007, p. 265-293.

[4] **Rosenhahn, B., T. Brox, U. Kersting, D. Smith, J. Gurney and R. Klette:**

A system for marker-less human motion estimation
Künstliche Intelligenz (KI) **1**, 45-51 (2006).

[5] **BRosenhahn, B., Ch. Perwass and G. Pose Sommer:**

Pose Estimation of Free-form Contours
International Journal of Computer Vision (IJCV) **62**, 267-289 (2007).