

Geometrische Relationen für die Bewegungsanalyse

Tido Röder, Meinard Müller, Michael Clausen
Institut für Informatik III, Universität Bonn

Einführung

Die Erforschung und Entwicklung von grundlegenden Techniken und Werkzeugen für den Vergleich, die Klassifikation und das Retrieval menschlicher Bewegungsdaten stellt in der Informatik ein aktuelles und anwendungsnahes Forschungsgebiet mit zahlreichen offenen wissenschaftlichen Fragestellungen dar. Die Erfassung solcher Daten erfolgt heutzutage oft unter Verwendung sogenannter optischer, mechanischer oder magnetischer Motion-Capture-Systeme. Motion-Capture-Daten, im folgenden auch MoCap-Daten genannt, kamen in einfacher Form bereits um 1980 in biomechanischen Anwendungen wie der Analyse von Anomalien des menschlichen Gangs zum Einsatz. In den 90er Jahren erfolgte dann der kommerzielle Durchbruch für MoCap-Anwendungen im Bereich der Computeranimation. Durch den Einsatz von automatisch gewonnenen MoCap-Daten gelang es hier erstmals im großen Stil, künstlichen Charakteren natürlich wirkende Bewegungen aufzuprägen. Neben Anwendungen in der Computeranimation und der Biomechanik stellen MoCap-Techniken mittlerweile wichtige Hilfsmittel in der Biometrie (Personenidentifikation), der Medizin und der Sportwissenschaft dar.

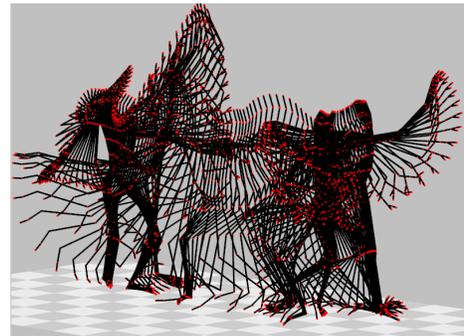


Abbildung 1: MoCap-Aufnahme einer Ballettbewegung als Folge von Einzelposen.

Im Gegensatz zum reinen Bild- oder Videodatenformat liegt dem MoCap-Datenformat ein

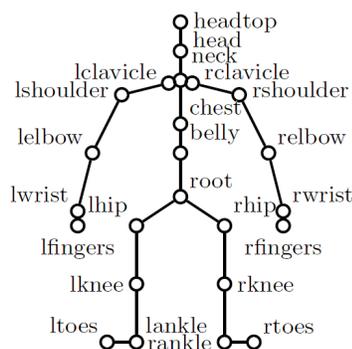


Abbildung 2: Skelettmodell.

Modell in Form einer kinematischen Kette zugrunde (siehe Abbildung 2). Analog zum menschlichen Skelett besteht eine solche kinematische Kette aus *Segmenten* (den „Knochen“), die durch *Gelenke* unterschiedlicher Typen verbunden sind. Da bei MoCap-Daten die 3D-Ortskoordinaten der Gelenke explizit und in hoher Genauigkeit vorliegen, spielen typische Computer-Vision-Probleme wie Tracking oder 3D-Rekonstruktion keine Rolle mehr.

Durch die strukturelle Modellinformation, also das Wissen, welcher 3D-Punkt zu welchem Körperteil korrespondiert, eröffnen MoCap-Daten neue Möglich-

keiten zur Bewegungsanalyse, -synthese und -klassifikation auf hohem semantischem Niveau. So geht es in aktuellen Forschungsprojekten z.B. um die Untersuchung von Bewegungsdetails, die natürlich aussehende von künstlichen Bewegungen unterscheiden [1]. Ebenso steht man bei der Erforschung von Parametern zur Beschreibung bestimmter Bewegungsstile vor zahlreichen offenen Fragestellungen [2, 3]. Große Probleme bereitet auch die Identifikation „logisch ähnlicher“ Bewegungsmuster – bisherige *quantitative* Ansätze wie der von Kovar und Gleicher in [4] verfolgte Zugang sind nur bedingt geeignet, die starken raum-zeitlichen Variabilitäten zu erfassen, die bei verschiedenen Ausführungen perzeptuell ähnlicher Bewegungsabläufe auftreten.

In diesem Beitrag skizzieren wir einen *qualitativen*, auf geometrischen Relationen basierenden, Ansatz zur Erfassung von großmotorischen Bewegungscharakteristika. Davon ausgehend zeigen wir, wie sich auf natürliche Weise semantisch sinnvolle Segmentierungen von Bewegungen in Bewegungsfragmente ergeben. Als Anwendungsbeispiel beschreiben wir abschließend eine Technik zur inhaltsbasierten Suche in Bewegungsdaten, siehe auch [5].

Geometrische Relationen

Aufbauend auf dem Skelettmodell, das den MoCap-Daten zugrunde liegt, führen wir nun informell das Konzept der geometrischen Relationen ein. Betrachten wir dazu beispielhaft die beiden Skelettposen in Abbildung 3(a). Die grünen Scheiben stellen die Ebene dar, die vom linken Sprunggelenk, vom linken Hüftgelenk und vom Hüftzentrum aufgespannt werden. Je nachdem ob nun das *rechte* Sprunggelenk vor oder hinter dieser Ebene liegt, ordnen wir der jeweiligen Pose den Wahrheitswert null (Fuß vor der Ebene) oder eins (Fuß hinter der Ebene) zu. Vereinfacht gesagt handelt es sich hierbei um einen Test, ob sich der rechte Fuß gerade vor oder hinter

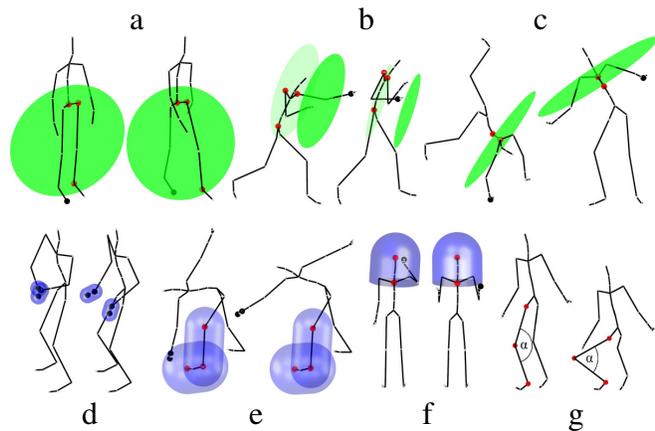


Abbildung 3: Beispiele geometrischer Relationen.

dem Körper befindet. Durch Vertauschung von links und rechts bei der Definition der Testebene ergibt sich die komplementäre Relation für den linken Fuß. Da diese Testebene am Körper fixiert ist, erhält man automatisch die Invarianz der Relationen gegenüber Translationen, Rotationen, Skalierungen, Änderungen der Skelettgeometrie und zu großen Teilen auch gegenüber Änderungen der Körperhaltung. Analog stellt Abbildung 3(b) einen Test auf eine vorwärtsgestreckte Hand

dar und Abbildung 3(c) einen Test auf eine angehobene Hand (Testkörperteile jeweils mit verdickten Gelenkpunkten markiert). Die Abbildungen 3(d-f) zeigen Tests auf Berührungen verschiedener Körperteile, während in 3(g) auf Anwinkelung eines Knies geprüft wird. Viele weitere derartige Tests sind denkbar.

Einzelne Skelettposen und Bewegungen, also zeitliche Abfolgen von Skelettposen, können nun auf einfache Weise qualitativ erfasst werden. Anstelle der quantitativen Spezifikation der Posen, z.B. mittels 3D-Trajektorien, wird jede Pose qualitativ durch einen *Merkmalsvektor* beschrieben, der die Wahrheitswerte einer geeigneten Menge geometrischer Relationen kodiert.

Als Beispiel betrachten wir einen Übergang vom Stehen ins Gehen und konzentrieren uns

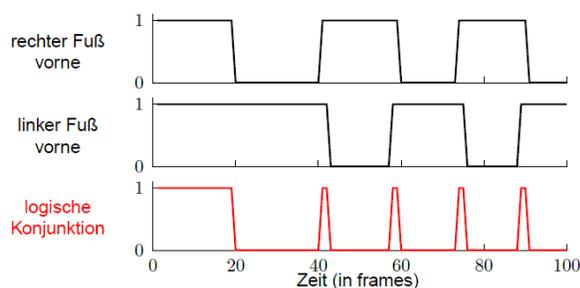


Abbildung 4: Logische Verknüpfung geometrischer Relationen für eine Gehbewegung (Schrittfolge links-rechts-links-rechts-links).

dabei auf die grobe Bewegung der Beine. Dazu wählen wir die aus Abbildung 3(a) bekannte Relation „rechter Fuß vorne“ und zusätzlich die komplementäre Relation „linker Fuß vorne“. Abbildung 4 (oben) zeigt die resultierenden zeitlichen Verläufe der Wahrheitswerte. Beim anfänglichen Stehen sind per Konstruktion der Ebenen beide Beine vor dem Körper, also ergibt sich der Merkmalsvektor (1,1). Bei Schritten mit dem linken Fuß ergibt sich (0,1), bei Schritten mit dem rechten Fuß dagegen (1,0).

Boolesche Ausdrücke in solchen Merkmalssignalen erlauben zusätzliche Interpretationen. So nimmt z.B. die logische Konjunktion der beiden Signale in Abbildung 4 genau dann den Wert eins an, wenn *beide* Füße vorne sind – die Spitzen im Signal treten also immer dann auf, wenn beim Gehen die Beine gerade aneinander vorbeiziehen. Anhand des Abstandes aufeinanderfolgender Signalspitzen kann man leicht die Schrittgeschwindigkeit bestimmen, einen semantisch sehr hochwertigen Parameter.

Als Fazit halten wir fest, dass der Einsatz geometrischer Relationen eine semantisch interpretierbare Vergrößerung von Bewegungen erlaubt, bei der durch geeignete Auswahl der Relationen gewisse Bewegungsaspekte betont und andere Aspekte ausgeblendet werden können.

Adaptive zeitliche Segmentierung

Bis jetzt haben wir einzelne Posen nur *räumlich* vergrößert – die *zeitliche* Dichte der Bewegungsrepräsentation wurde nicht angetastet. Es zeigt sich, dass eine noch kompaktere Beschreibung

möglich ist, wenn man aufeinanderfolgende Posen, deren Merkmalsvektoren sich nicht unterscheiden, zu *einem Segment* zusammenfasst. Indem man nur noch die Abfolge der Merkmale der einzelnen Segmente betrachtet, ohne die Segmentlängen zu berücksichtigen, erhält man so etwas wie einen „Fingerabdruck“ der Bewegung, der invariant unter räumlichen und zeitlichen Deformationen ist. Diese Invarianz rührt daher, dass die Segmentgrenzen per Konstruktion durch bewegungsdefinierte Ereignisse bestimmt sind, deren *Reihenfolge* sich bei unterschiedlichen Bewegungsausführungen nicht ändert. Langsames und schnelles Gehen, müdes oder fröhliches Gehen ergeben auf diese Weise alle denselben Fingerabdruck.

Betrachten wir zum Beispiel die Gehbewegung in Abbildung 5. Wir beschränken uns wie im

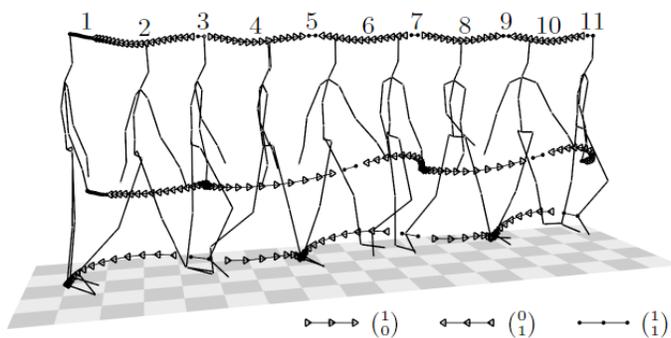


Abbildung 5: Adaptive Segmentierung einer Gehbewegung mit den Relationen „rechter Fuß vorne“ und „linker Fuß vorne“

vorhergehenden Beispiel auf die beiden geometrischen Relationen für die Beine. Die 3D-Trajektorien von Kopf, rechter Hand und rechtem Sprunggelenk wurden hier beispielhaft eingezeichnet und mit drei verschiedenen Markierungstypen versehen, je nachdem welche Merkmalsausprägung in der dazugehörigen Pose vorliegt. Dadurch ergeben sich elf Segmente mit der zugehörigen Merkmalsabfolge (1,1), (0,1), (1,1), (1,0), (1,1), (0,1), (1,1), (1,0), (1,1), (0,1), (1,1), die den Prototypen einer Gehbewegung charakterisiert.

vorhergehenden Beispiel auf die beiden geometrischen Relationen für die Beine. Die 3D-Trajektorien von Kopf, rechter Hand und rechtem Sprunggelenk wurden hier beispielhaft eingezeichnet und mit drei verschiedenen Markierungstypen versehen, je nachdem welche Merkmalsausprägung in der dazugehörigen Pose vorliegt. Dadurch ergeben sich elf Seg-

Wir zeigen nun eine mögliche Anwendung dieser Techniken zur Gewinnung eines Fingerabdrucks von Bewegungen auf, bei der die beschriebenen Invarianzen eine wichtige Rolle spielen.

Anwendung bei der inhaltsbasierten Suche in Bewegungsdaten

Das Problem der inhaltsbasierten Suche in Bewegungsdaten verdeutlicht man sich am besten aus Sicht eines Benutzers, der eine große Kollektionen von MoCap-Aufnahmen, wie sie z.B. bei der Produktion von Computeranimationsfilmen anfallen, durchsuchen möchte. Bei der manuellen Suche nach einer bestimmten Aufnahme oder einem bestimmten Bewegungstypen steht als einziger Hinweis auf den Inhalt einer Aufnahme häufig nur ein kryptischer Dateiname zur Verfügung. In diesem Fall scheint die Lösung zu sein, jede Aufnahme von Hand zu überprüfen, was aber aus Zeitgründen bei großen Datenmengen nicht machbar ist.

Hier können inhaltsbasierte Suchtechniken weiterhelfen. Vereinfacht ausgedrückt kann ein Computer die Aufgabe übernehmen, sich alle Aufnahmen vorab einmal anzusehen und wichtige Charakteristika der Bewegungen abzuspeichern, um anschließend Suchanfragen schnell beantworten zu können. Man nennt diesen einmaligen Vorgang *inhaltsbasierte Indexierung*. Um mit dem resultierenden Index suchen zu können, müssen Suchanfragen auf geeignete Weise formuliert werden. Häufig bedient man sich hier des „Query-By-Example“-Paradigmas, das es vorsieht, eine Suchanfrage in Form einer kurzen Beispielbewegung zu stellen. Aus dieser Beispielbewegung werden analog

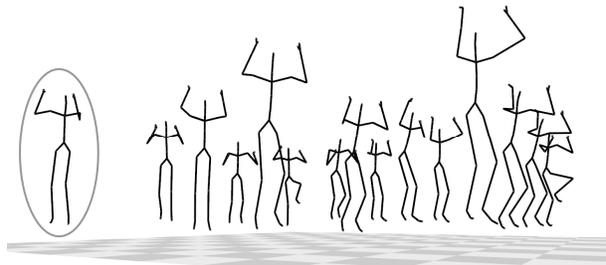


Abbildung 6: 15 Treffer bei der Suche nach einem Strecksprung (Anfrage ist eingekreist).

zum Indexierungsschritt wichtige Bewegungscharakteristika extrahiert. Anschließend wird der vorab erstellte Index mit einem geeigneten Suchverfahren nach diesen Charakteristika durchsucht. Treffer im Index korrespondieren zu Treffern in der ursprünglichen Datenkollektion und werden als Suchergebnis zurückgeliefert.

Entscheidend für die Qualität der Suchergebnisse ist die Beschaffenheit der Bewegungscharakteristika, die bei der Indexierung und bei der Anfragespezifikation extrahiert werden. Es zeigt sich, dass die oben eingeführten „Fingerabdrücke“ sehr gut für die inhaltsbasierte Suche geeignet sind. Zum Beispiel wurde in Abbildung 6 nach Strecksprüngen gesucht, wozu ein MoCap-Fragment von ca. 2 Sekunden als Anfrage diente. Die resultierenden Treffer enthielten neben Strecksprüngen auch noch weitere, ähnliche Bewegungen. Allerdings ließen sich die Treffer auf automatische Weise derart sortieren, dass die 15 besten Treffer auch wirklich alle in der Datenkollektion enthaltenen Strecksprünge waren. Typische Suchzeiten für derartige Anfragen betragen nur wenige zehntel Sekunden.

Literatur

- [1] Ren, L., Patrick, A., Efros, A. A., Hodgins, J. und Rehg, J. M.: *A data-driven approach to quantifying natural human motion*. ACM Transactions on Graphics 24(3): 1090-1097, 2005.
- [2] Liu, C. K., Hertzmann, A. und Popović, Z.: *Learning physics-based motion style with nonlinear inverse optimisation*. ACM Transactions on Graphics 24(3): 1071-1081, 2005.

- [3] Hsu, E., Pulli, K. und Popović, J.: *Style translation for human motion*. ACM Transactions on Graphics 24(3): 1082-1089, 2005.
- [4] Kovar, L. und Gleicher, M.: *Automated extraction and parameterization of motions in large data sets*. ACM Transactions on Graphics 23(3):559-568, 2004.
- [5] Müller, M., Röder, T. und Clausen, M.: *Efficient content-based retrieval of motion capture data*. ACM Transactions on Graphics 24(3): 677-685, 2005.