

Informatische Aspekte der Bewegungsanalyse

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

Einführung

Die Analyse menschlicher Bewegungen im Sport, in der Medizin und in der Biometrie basierte bislang häufig auf der halbautomatischen Auswertung von Fotoreihen oder Videomaterial. Vorangetrieben durch Entwicklungen im Bereich der Computeranimation sind in jüngster Zeit computergestützte Aufnahmetechniken zur dreidimensionalen Erfassung von Bewegungen entstanden, die unter dem Oberbegriff *Motion Capturing* (MoCap) zusammengefasst werden.

Dieser Beitrag skizziert einige Grundtechniken zur Bewegungsanalyse basierend auf MoCap-Daten. Insbesondere geht es um die fundamentalen Fragen, wann zwei MoCap-Aufnahmen als *ähnlich* anzusehen sind und wie ähnliche Bewegungen auf automatische Weise in zeitliche Korrespondenz gebracht werden können. Als mögliche Herangehensweise an diese Problemstellungen wird das generische Konzept des *Dynamic Time Warping* vorgestellt.

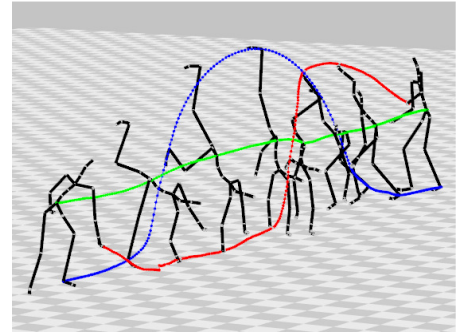


Abbildung 1: Einzelposen aus einer MoCap-Aufnahme eines Radschlags. Die 3D-Trajektorien einiger Gelenke sind eingezeichnet.

Optisches Motion Capturing und Datenformate

Bevor wir uns dem Problem der Ähnlichkeitsbewertung für Bewegungen widmen, soll es zunächst um die Aufnahmetechnik und das Datenformat der zugrundeliegenden MoCap-Daten gehen. Beim weitverbreiteten *optischen* Motion Capturing wird eine Anordnung von mehreren

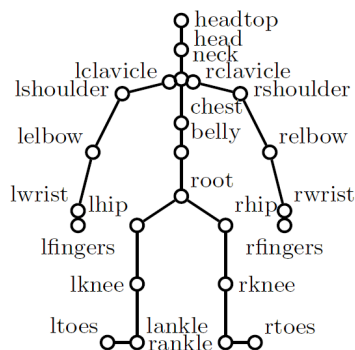


Abbildung 2: Skelettmodell.

Highgeschwindigkeitskameras verwendet, um durch wiederholte Positionsbestimmung die dreidimensionalen Trajektorien von reflektierenden Markern an ausgewählten Punkten des menschlichen Körpers zu erfassen. Der resultierende MoCap-Datenstrom ist eine fortlaufende Folge von *Frames*, die für jeden Aufnahmezeitpunkt die räumlichen Positionen der ausgewählten Körperpunkte mit einer Auflösung im Millimeterbereich beschreiben, siehe Abb. 1.

Das gebräuchlichste Modell zur Beschreibung menschlicher Bewegungen ist das der kinematischen Kette, einem vereinfachten Abbild des menschlichen Skelettes, siehe Abb. 2. Wie auch das menschliche Skelett besteht eine kinematische Kette aus verschiedenen Typen

von beweglichen Gelenken und diese verbindenden Segmenten, welche den Knochen entsprechen. Mittels inverser Kinematik kann jeder Frame eines MoCap-Datenstroms in eine Pose der zugrundegelegten kinematischen Kette umgerechnet werden. Eine Pose besteht aus einem Parametersatz, der den Zustand der kinematischen Kette zu einem festen Zeitpunkt beschreiben. Hierzu werden meistens 3D-Ortskoordinaten oder Gelenkwinkel verwendet. Durch Verfolgung solcher Parameterwerte über die Zeit erhält man sogenannte *Trajektorien*, die z.B. den Winkelverlauf von Gelenken oder den 3D-Ortskoordinatenverlauf von ausgezeichneten Körperpunkten beschreiben (siehe Abb. 1 und 3).

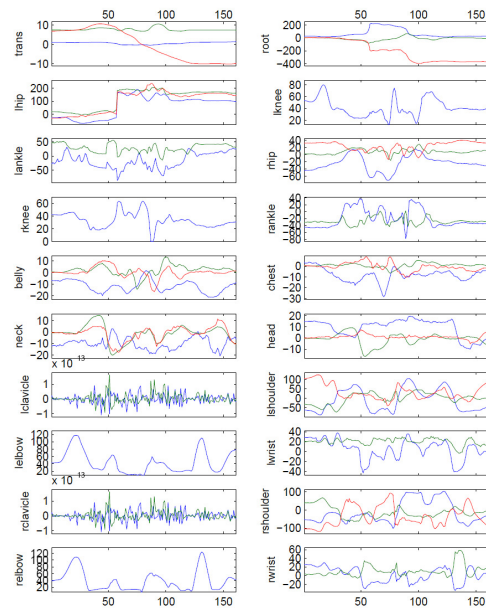


Abbildung 3: Gelenkwinkeltrajektorien für den Radschlag aus Abb. 1. Die horizontalen Achsen repräsentieren die Zeit, die vertikalen Achsen die Eulerwinkel in Grad.

Aspekte der Ähnlichkeit

Beim Vergleich zweier Bewegungen gibt es Aspekte, die typischerweise keinen Einfluss auf die Ähnlichkeitsbewertung haben sollten. So entspricht es meist der menschlichen Wahrnehmung, globale räumliche Transformationen wie Translationen der Gesamtbewegungen (räumliche Position der Bewegung), räumliche Skalierungen (Größe des Menschen) oder Rotationen um die Hochachse (Richtung der Bewegung) beim Vergleich unberücksichtigt zu lassen, siehe Abb. 4. In einem gewissen Rahmen sind oft auch globale zeitliche Transformationen (Geschwindigkeit der Bewegung) irrelevant.

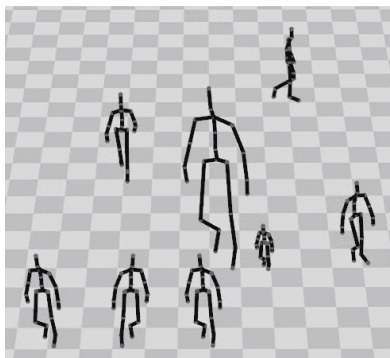


Abbildung 4: Transformationen einer Gehbewegung

Schwieriger wird es für automatische Ähnlichkeitsbewertungen, wenn Bewegungen auch dann noch als ähnlich angesehen werden sollen, wenn sie zwar derselben Bewegungsklasse entstammen, aber in verschiedenen Stilen ausgeführt werden. In Abb. 5 sieht man z. B. neun Posen aus verschiedenen Gehbewegungen, die den Stilen „Marschieren“, „Schleichen“, „Hum-

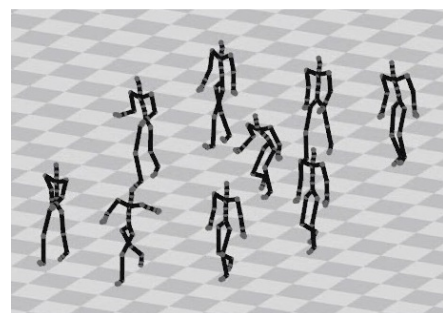


Abbildung 5: Verschiedene Stile beim Gehen.

peln“, „Schlurfen“, „trauriges Gehen“ usw. entsprechen. Die Übergänge zwischen solchen Bewegungsstilen sind fließend, so dass es selbst für den Menschen häufig schwierig ist, Bewegungen hinsichtlich ihrer Ähnlichkeit zu klassifizieren.

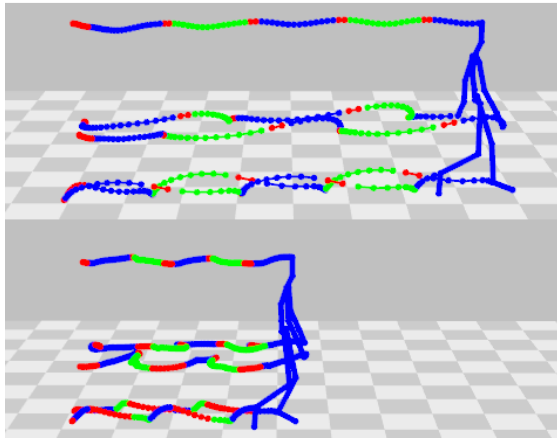


Abbildung 6: Schnelles, dynamisches Gehen (oben) gegenüber langsamem Gehen einer alten Frau (unten). Korrespondierende Phasen der Gehzyklen sind in gleichen Trajektorienfarben dargestellt.

Die Technik zur Synchronisation von Bewegungen, die sich durch lokale raum-zeitliche Deformationen voneinander unterscheiden, vorgestellt.

Neben globalen Transformationen und Variationen des Bewegungsstils können *lokale* raum-zeitliche Deformationen zu erheblichen Unterschieden zwischen Bewegungen führen. So unterscheiden sich z.B. die in Abb. 6 gezeigten Gehbewegungen außer in der globalen Fortbewegungsgeschwindigkeit auch hinsichtlich zeitlicher Phasenbeziehungen innerhalb der Gehzyklen sowie hinsichtlich räumlicher Abweichungen wie z.B. Pendelbewegungen der Arme.

Im Folgenden wird eine generische Tech-

Lokale Kostenmaße

Wie oben erwähnt kann ein MoCap-Datenstrom als Folge von Einzelposen der zugrundeliegenden kinematischen Kette aufgefasst werden. Es liegt daher nahe, beim Vergleich von Bewegungen eine Bottom-Up-Sicht einzunehmen und *lokale* Ähnlichkeitsbewertungen zwischen Posen in geeigneter Weise zu *globalen* Ähnlichkeitsbewertungen zwischen Gesamtbewegungen zusammenzufügen. Wir gehen im Folgenden auf zwei solche lokalen Ähnlichkeitsmaße ein. Bei Maßen, die die *Unähnlichkeit* von Bewegungen messen, spricht man auch von *Kostenmaßen*.

Kovar und Gleicher [1] beschreiben jede Pose durch die Familie der Ortskoordinaten der Skelettpunktmenge, welche mit J bezeichnet sei. Der Vergleich der Posen zu den Zeitpunkten

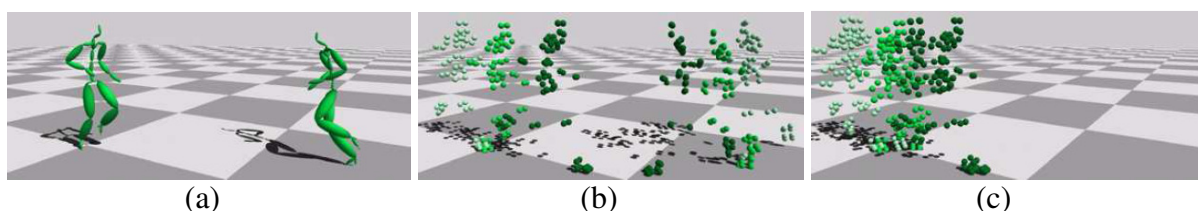


Abbildung 7: Zum lokalen Kostenmaß nach Kovar und Gleicher: (a) zwei Posen; (b) Umwandlung in Punktwolken unter Berücksichtigung eines zeitlichen Kontextes; (c) Ausrichtung der beiden Punktwolken durch Rotation und Translation.

m und n erfolgt im Kontext einer festen Anzahl 2ρ von vorhergehenden und nachfolgenden Posen, und die resultierenden Punktwolken werden mittels einer Transformation T so verschoben (Translationsparameter x_0, z_0) und gedreht (Rotationsparameter θ), dass die gewichtete Summe der euklidischen Abstände einander entsprechender Punkte minimiert wird. Dies sind die lokalen Kosten $d^p(m, n)$, oder auch der *Abstand* zwischen der m -ten und der n -ten Pose:

$$d^p(m, n) = \min_{\theta, x_0, z_0} \sum_{t \in [-\rho; \rho]} \sum_{j \in J} w_j \|p_{m+t}^j - T_{\theta, x_0, z_0} p_{n+t}^j\|^2$$

Hierbei bezeichnet p_k^j den Ortskoordinatenvektor des Skelettpunkts $j \in J$ zum Zeitpunkt k . Die Gewichte $w_j \in \mathbf{R}$ für $j \in J$ ermöglichen eine unterschiedlich starke Berücksichtigung der verschiedenen Körperteile.

Ein anderes Kostenmaß basiert auf einer Gelenkwinkeldarstellung von Posen. Mittels sogenannter Einheitsquaternionen q lassen sich die Rotationen an den Gelenken beschreiben. Man bildet nun analog zum vorherigen Kostenmaß eine gewichtete Summe der Beträge von Skalarprodukten solcher Einheitsquaternionen, wobei q_k^j die Quaternionendarstellung der Gelenkwinkel am Gelenk $j \in J$ zum Zeitpunkt k bezeichnet:

$$d^q(m, n) = 1 - \sum_{j \in J} w_j |\langle q_m^j, q_n^j \rangle|$$

Beide vorgestellten Kostenmaße sind invariant gegenüber globalen Translationen und Rotationen. Weiterhin ist d^q invariant gegenüber räumlichen Skalierungen. Im Folgenden wird exemplarisch auf das Kostenmaß d^q zurückgegriffen.

Kostenmatrizen

Die lokalen Kosten fasst man systematisch zu sogenannten Kostenmatrizen zusammen. Die Kostenmatrix für zwei MoCap-Bewegungen bestehend aus M bzw. N Posen ist eine Matrix mit M Zeilen und N Spalten, wobei der Eintrag in Zeile m und Spalte n die lokalen Kosten $d^p(m, n)$ beschreibt. Zur Veranschaulichung dieses Konzepts betrachten wir zunächst den Spezialfall, dass eine Bewegung mit sich selbst verglichen wird. Abb. 8 (a) zeigt die

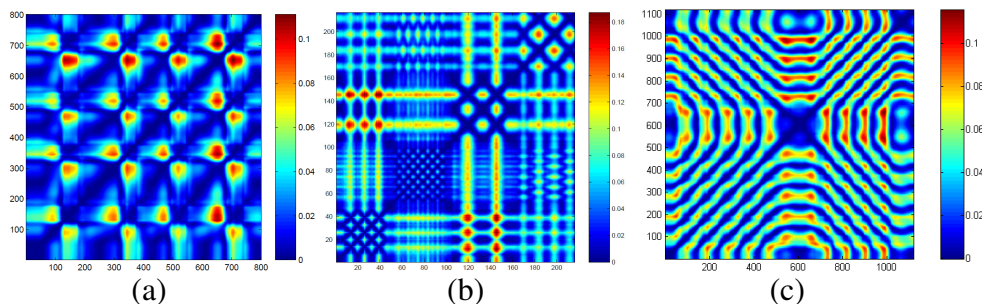


Abbildung 8: Kostenmatrizen für den Vergleich diverser Bewegungen mit sich selbst; hellere bzw. rötlichere Farben stehen für höhere Kosten. (a) Hüpfbewegung bestehend aus vier Sprüngen; (b) Gymnastiksequenz; (c) Armkreisen, fünf mal vorwärts, vier mal rückwärts.

Kostenmatrix einer Sequenz von vier aufeinanderfolgenden Hüpfbewegungen. In diesem Fall ist die Kostenmatrix quadratisch und symmetrisch bezüglich der Hauptdiagonalen (diese verläuft von links unten nach rechts oben), entlang derer nur Nullkosten auftreten, da hier die Unähnlichkeit jeder Pose zu sich selbst gemessen wird. Zusätzlich erkennt man verschiedene Nebendiagonalen niedriger Kosten. Diese zeigen an, dass in der Bewegung manche Teile wiederholt vorkommen. So besagt z. B. die Nebendiagonale, die an der vertikalen Achse bei Pose 400 ansetzt, dass die zwei ersten Sprünge (Posen 1-400) ähnlich zu den zwei letzten Sprüngen (Posen 400-800) sind. Man beachte, dass die beiden Dreiecke oberhalb und unterhalb der Hauptdiagonalen aufgrund der Symmetrie des lokalen Kostenmaßes dieselbe Information enthalten.

Als weiteres Beispiel soll eine Gymnastiksequenz dienen, siehe Abb. 8 (b): zwei Hampelmänner, zehn Schritte auf der Stelle, zwei Kniebeugen, vier Wiederholungen von „Knie berührt Ellenbogen“. Auch hier spiegelt sich die Repetitivität der einzelnen Übungen in der Nebendiagonalenstruktur wider. Weiterhin fallen die zwei Streifen hoher Kosten auf, die sich bei den Posen 120 und 145 durch die gesamte Matrix ziehen. Diese Streifen rühren daher, dass die beiden Kniebeugen sehr unähnlich zu sämtlichen anderen Bewegungen sind.

Abb. 8 (c) zeigt die Kostenmatrix für fünfmaliges Armkreisen vorwärts (Posen 1-500) gefolgt von viermaligem Armkreisen rückwärts (Posen 700-1000), jeweils synchron mit beiden Armen ausgeführt. Parallel zur Hauptdiagonalen entstehen auch hier Nebendiagonalen niedriger Kosten als Ausdruck der Bewegungswiederholungen. Zusätzlich ergeben sich auffällige Strukturen in der oberen linken bzw. unteren rechten Ecke der Matrix – man erkennt Bereiche niedriger Kosten, die *senkrecht* zur Hauptdiagonalen verlaufen. Betrachten wir z.B. die Diagonale, die bei Pose 1050 an der vertikalen Achse ansetzt und nach rechts unten bis Pose 450 auf der horizontalen Achse verläuft. Diese Diagonale lässt sich wie folgt interpretieren: Denkt man sich die ersten vier Wiederholungen des Vorwärts-Armkreisens rückwärts abgespielt, dann sind sie ähnlich zu den vier Wiederholungen des Rückwärts-Armkreisens.

Dynamic Time Warping

Berechnet man die Kostenmatrix für zwei verschiedene Bewegungen, wie etwa in Abb. 9 für zwei verschieden schnelle Gehbewegungen, lassen sich analoge Informationen ablesen. Die schnellere Gehbewegung bezeichnen wir hier informell als die *vertikale* Bewegung, weil ihre Zeitachse in der Matrix vertikal verläuft. Entsprechend nennen wir die langsamere Gehbewegung die *horizontale* Bewegung. Pfade von der linken unteren Ecke in die rechte obere Ecke der Kostenmatrix entsprechen hier Zuordnungen von Posen zwischen den Bewegungen: verläuft solch ein Pfad in einem gewissen Bereich in horizontaler Richtung, entspricht das einer

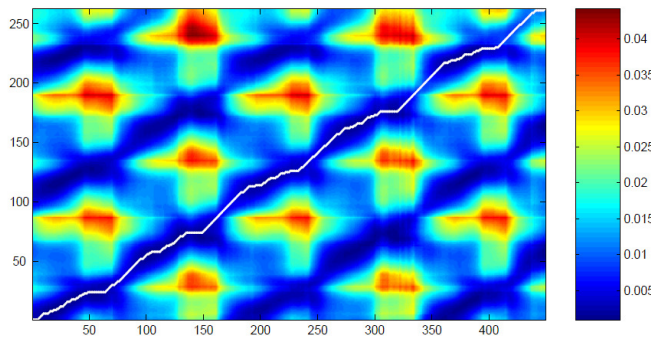


Abbildung 9: Kostenmatrix mit weiß markiertem kostengünstigstem Pfad. Es wird eine schnelle Gehbewegung (vertikale Achse) mit einer langsamen Gehbewegung (horizontale Achse) verglichen; jeweils fünf Schritte.

Zuordnung mehrerer Posen der horizontalen Bewegung zu einer Pose der vertikalen Bewegung, die Zeit bleibt also für die vertikale Bewegung stehen. Umgekehrt entsprechen vertikale Pfadabschnitte einem Stehenbleiben der Zeit für die horizontale Bewegung. Bei einem diagonalen Verlauf läuft in beiden Bewegungen die Zeit weiter. Von besonderem Interesse sind *kostenoptimale* Pfade in der Kostenmatrix, das sind solche Pfade, entlang

derer die Summe der Kosten ein Minimum annimmt. Solche Pfade codieren eine optimale zeitliche Verzerrung, um die beiden Bewegungen bezüglich des gewählten lokalen Kostenmaßes zu *synchronisieren*. Die räumlichen Verzerrungen zwischen den Bewegungen werden dabei durch das lokale Kostenmaß erfasst. Im Beispiel von Abb. 9 (a) sieht man, dass der eingezeichnete Pfad eher horizontal als vertikal ausgerichtet ist. Die Zeit verläuft also für die vertikale Bewegung langsamer, um sie an die langsamere horizontale Bewegung anzupassen.

Dynamic Time Warping (DTW) ist eine Technik zur automatischen Berechnung solcher kostenoptimaler Pfade. Der Hauptnachteil dieser Technik ist die hohe Berechnungskomplexität: Zur Synchronisation zweier Bewegungen mit M bzw. N Frames werden größenordnungsmäßig $M \cdot N$ Rechenoperationen benötigt, was für große M und N in der Praxis zu viel ist. Ein völlig anderer, alternativer Ansatz ohne derartige Einschränkungen findet sich in [2].

Literatur

- [1] Kovar, L. und Gleicher, M.: *Automated extraction and parameterization of motions in large data sets*. ACM Transactions on Graphics 23(3):559-568, 2004.
- [2] Müller, M., Röder, T. und Clausen, M.: *Efficient content-based retrieval of motion capture data*. ACM Transactions on Graphics 24(3): 677-685, 2005.