Institut für Anthropomatik und Robotik (IAR)

Prof. Dr.-Ing. T. Asfour

Karlsruher Institut für Technologie, Fakultät für Informatik

Motions in Man and Machine (3M-Seminar)
Sommersemester 2014

Themenvorschlag 1:

Segmentation of Human Motion

Betreuer:

Dipl.-Inform. Christian Mandery

Juni 2014

Kurzfassung: *Die Beobachtung des Menschen kann einen erheblichen Beitrag dazu leisten, die Leistungsfähigkeit eines Roboters hinsichtlich der Interaktion mit dem Menschen und dessen Umwelt zu erhöhen. Ein wichtiger Schritt um das Verständnis und die Modellierung von beobachteten Aktionswissen zu erleichtern ist die Segmentierung von Bewegungsdaten. Insbesondere bei langen Aktionsfolgen, welche in längeren Bewegungssequenzen mit hoher Variabilität niedergeleget werden müssen, ist eine zweckmäßige Unterteilung von Bewegungsdaten eine notwendige Voraussetzung um die Verarbeitung zur Analyse, Erkennung und dem Erlernen von Aktionen für Roboter zu ermöglichen. In diesem Seminarthema sollen Grundlagen sowie aktuelle Methoden zur Bewegungsaufnahme und -segmentierung vorgestellt werden.*

Inhaltsverzeichnis

1 Einleitung **3**

2 Bewegungsaufnahme **3**

3 Bewegungsrepräsentation **5**
3.1 Komprimierte Bewegungsrepräsentation 7

4 Segmentierung von Bewegungsdaten **9**

5 Seminar-Aufgaben **11**
5.1 Aufnahme von Bewegungsdaten 11
Aufgabe 1 11
5.2 Theorie-Teil 11
Aufgabe 2 11
Aufgabe 3 11
Aufgabe 4 12
5.3 Praxis-Teil 12
Aufgabe 6 12

Aufgabe 7 12
Aufgabe 8 12
Aufgabe 9 12

1 Einleitung

Menschen erlernen neue Fähigkeiten und Bewegungen durch die Beobachtung von Demonstrationen, welche durch Experten vorgeführt werden. Die Fähigkeit, neues Wissen aus der Bebachtung zu extrahieren und in geeigneter Weise abzulegen, kann dazu beitragen, das Lernen bei Robotern zu beschleunigen. Dies ist insbesondere bei solchen Systemen, die in menschzentrierten Umgebungen zu agieren haben, der Fall. Um aufgabenspezifische Aktionen und deren Folgen auf die Umgebung und anderen Agenten verstehen zu können, sind essentielle Merkmale aus den menschlichen Demonstrationen zu extrahieren und in einer generischen Form zu repräsentieren, die zugänglich für Roboter ist und eine zukünftige Adaption und Reproduktion erlaubt.

Für die Repräsentation von Bewegungen wird in der Literatur häufig auf das Konzept der Bewegungsprimitive zurückgegriffen, welches wiederum durch das menschliche Vorbild inspiriert ist. Es wird davon ausgegangen, dass der Mensch über ein umfangreiches Bewegungrepertoire verfügt, welches sich aus Grundbausteinen zusammensetzt, wobei jeder Baustein eine einfache parametrisierbare Bewegung repräsentiert. Für die Synthese komplexer Bewegungen werden Sequenzen von Bewegungsprimitven generiert, angepasst und ausgeführt. Um dieses Repertoire zu erweitern, werden hierfür Bewegungsabläufe in Segmente heruntergebrochen, um diese verarbeiten zu können. Um das gleiche Prinzip des menschlichen Bewegungslernen auf einen Roboter übertragen zu können, ist hierfür die Darstellung komplexer Bewegungen in Segmenten unerlässlich. Die gestückelte Aufnahme einer Bewegungen ist oft zu aufwendig und für einige Aktionen nicht zu realisieren, z.B. beim Jonglieren. Längere Aufnahmen haben zudem den Vorteil, dass Bewegungen in einem natürlichen Ablauf aufgenommen werden können. Um den Aufwand der mit der manuelle Sichtung und Segmentierung dieser Daten verbunden ist zu verringern, wurden Verfahren entwickelt, welche eine automatisierte Analyse und Segmentierung von Bewegungsdaten ermöglichen.

2 Bewegungsaufnahme

Die Aufnahme menschlicher Bewegungen kann mit unterschiedlichen Sensortechniken erfolgen. Die gängigsten Systeme verwenden visuelle Sensoren, um die Bewegungen des gesamten menschlichen Körpers sowie der Umwelt erfassen zu können. Grundsätzlich lassen sind visuelle Systeme darin unterscheiden, ob Marker verwendet werden oder nicht. Markerlose Systeme verwenden in der Regel einfache, preiswerte Sensoren, wie Farbbildkameras oder Laserscanner, die zur Erfassung der sensorischen Information der gesamten Szene genügen. Mittels probabilistischer Verfahren und Heuristiken aus dem Vorwissen über das menschliche Erscheinungsbild werden entsprechende Bildmerkmale wie Kontur- und Farbmerkmale extrahiert, aus denen sich die Bewegungen des beobachteten Menschen rekonstruieren lassen. Der Vorteil dieser Verfahren liegt in dem geringen Hardware- und Installationsaufwand, welche sich im Preis und der Portabilität solcher Systeme niederschlägt. Ein großer Nachteil dieser System ist jedoch die Ungenauigkeit und die niedriger Bildrate, mit der Bewegungen erfasst werden können. Verfahren zur

markerlosen Erfassung der menschlichen Bewegung lassen sich in ansichtsbasierte und modellbasierte Verfahren einteilen. Ansichtsbasierte Verfahren zeichnen sich dadurch aus, dass keinerlei Vorwissen über die Kinematik des menschlichen Körpers notwendig ist, wohingegen modellbasierte Verfahren ihre Schätzungen über die Bewegungen des Körpers auf diesem Wissen fundieren. Aus diesem Grund erreichen modellbasierte Verfahren in den meisten eine höhere Performanz sowohl bezüglich der Genauigkeit als auch der Laufzeit.

(Für den interessierten Leser: In [4] wird eine Methode vorgestellt, welche auf Basis von Suchalgorithmen und den den extrahierten Konturmerkmalen eine entsprechende Schätzung für ein Menschmodell ermittelt. Es sind hier jedoch Ansichten von unterschiedlichen Positionen notwendig, um in 3D-Raum eine exakte Lösung zu finden. In [6] wird mittels dem Iterative-Closest-Point-Verfahren ein Modell bestehend aus starren 3D-Körpern an eine 3D-Punktwolke angepasst. Für dieses Verfahren werden zusätzliche Sensoren mit guter Tiefenauflösung benötigt wie z.B. der Swissranger. Verfahren, welche lediglich auf den Ansichten von unidirektionalen Kameras basieren, werden in [1] und [2] präsentiert. [1] versucht aus monokularen Ansichten mit Regression mittels Relevance-Vektoren die Gelenkwinkel eines Menschmodells zu schätzen. Nachteilig wirkt sich die fehlende Tiefeninformation aus, welche zu hohen Fehlern in der Schätzung führen. Ausgehend von einem Stereokamerasystem wird in [2] ein Verfahren vorgestellt, welches den Partikelfilter-Algorithmus anwendet um eine Menschmodell mit Kanten- und Tiefenmerkmalen abzugleichen.)

Markerbasiertes Systeme hingegen benötigen die eindeutige Markierung relevanter Regionen des menschlichen Körpers. Die angebrachten Marker sind auf das System abgestimmt und erlauben somit eine präzise Ortung im Raum mit geringem Rechenaufwand. Aufgrund der Genauigkeit und der Möglichkeit, Bewegungen mit hoher Bildrate zu erfassen, werden markerbasierte Systeme häufig im Animationsbereich eingesetzt. Ein bekanntes, kommerzielles System wurde von der Firma VICON entwickelt und hergestellt. Die Funktionsprinzipien basieren auf der Verwendung reflektierender Marker (in Kugelform). Um diese Marker registrieren zu können, werden Kameras eingesetzt, die ein Infrarotsignal emittieren, welches von den Marker reflektiert und den Marker auf diese Weise für die Registrierung durch die Kamera kenntlich macht. Die Verwendung mehrerer Kameras ermöglicht die Berechnung der Markerposition im 3D-Raum mittels Triangulation. Aufgrund des geringen Rechenaufwandes ermöglicht dieses Verfahren die Aufnahme von Bewegungsdaten mit bis zu 200Hz, wobei die Ortungsgenauigkeit für Marker unter optimalen Bedingungen im Submillimeterbereich liegen kann.

Für die Aufnahme des menschlichen Körpers sind ca. 30-50 Marker notwendig, wobei jeder Marker entsprechend der Position auf dem Körper mit einem Label zu kennzeichnen ist. Um die Robustheit der Erfassung zu erhöhen und die Verwechslungsgefahr der Marker zu minimieren, wird ein Skelettmodell verwendet, welches mit den zu erwartenden Markerpositionen auf dem Körper annotiert wird. Dieses Modell wird in jedem Bild bzw. in jede registrierte Markerkonfiguration eingepasst um mögliche Mehrdeutigkeiten aufzulösen.

3 Bewegungsrepräsentation

Erfasste Bewegungen können in Form von Trajektorien relevanter Punkte im kartesischen Raum beschrieben werden. Punkte, die dabei von besonderem Interesse sind, sind vor allem die Endeffektoren wie die Hände und die Füße, welche mit der Umgebung interagieren. Unter Verwendung eines kinematischen Modelles können Positionen im kartesischen Raum in Gelenkwinkelposen überführt werden. In welcher Form die Repräsentation gewählt wird (ob im kartesischen oder im Gelenkwinkelraum) hängt meistens von der Art der Bewegung ab. Zielgerichtete Bewegungen, welche besipielsweise eine Interaktion mit Objekten vorsehen, werden häufig im kartesischen Raum dargestellt. Auf diese Weise lässt sich eine Bewegung bzw. Aktion in einen bestimmten Kontext aus Aufgabe und Umgebung einfacher einbetten. Bewegungen hingegen, die weniger Präzision erfordern wie Gesten und Tanzbewegungen, werden oft im Gelenkwinkelraum dargestellt, um die Natürlichkeit dieser Bewegungen zu erhalten. Je nach Darstellungsform der Bewegungen kann die Segmentierung und Verarbeitung von Daten, welche die gleiche Bewegung repräsentieren, zu unterschiedlichen Ergebnissen führen.

Beide Repräsentation lassen sich ineinander überführen. Die Berechnung der Endeffektorposition aus einer Gelenkwinkelkonfiguration lässt sich mit einem gegebenen kinematischen Modell des Roboters bzw. des Menschen mittels der direkten Kinematik lösen (welches sich mittels Verkettung von Transformationsmatrizen ohne Weiteres lösen lassen sollte). Das umgekehrte Problem, also die Berechnung von Gelenkpositionen aus der kartesischen Endeffektorposition, ist etwas komplexer und ist beschrieben durch das Problem der inversen Kinematik, welches mit unterschiedlichen Methoden gelöst werden kann. Diese lassen sich unterteilen in numerische, algebraische und geometrische Methoden:

Numerische Methoden Gängigste Vorgehensweise bei der numerischen Lösung der inversen Kinematik ist die Annäherung der gewünschten Bewegungsänderung mittels Linearisierung. Für eine Bewegungsfunktion $f : \mathbb{R}^M \to \mathbb{R}^N$:

$$p(t) = f(q(t)) \tag{1}$$

mit welcher die Position p des Endeffektors für eine Gelenkwinkelkonfiguration q berechnet werden kann, wird die Jacobi-Matrix wie folgt aufgestellt:

$$J(q) = \begin{pmatrix} \frac{\vartheta f_1}{q_1} & . & . & . & \frac{\vartheta f_1}{q_N} \\ . & . & & & . \\ . & & . & & . \\ . & & & . & . \\ \frac{\vartheta f_M}{q_1} & . & . & . & \frac{\vartheta f_M}{q_N} \end{pmatrix}. \tag{2}$$

Diese enthält sämtliche partiellen Ableitungen von f und beschreibt den Zusammenhang zwischen den Gelenkgeschwindigkeiten und Geschwindigkeit im kartesischen Raum. Die kartesische Bewegung $\dot{p}(t)$ kann dann anhand der Gelenkwinkelgeschwindigkeiten $\dot{q}(t)$ wie folgt bestimmt werden:

$$\dot{x}(t) = J(q(t))\dot{q}(t). \tag{3}$$

Analog kann mittels der inversen Jacobi-Matrix $\dot{q}(t)$ aus $\dot{x}(t)$ bestimmt werden:

$$\dot{p}(t) = J^{-1}(q(t))\dot{q}(t). \tag{4}$$

Falls $J(q(t))$ nicht invertierbar ist, ist die Inverse mittels der Pseudoinversen $J^{+}(q(t))$ zu approximieren, für die Folgendes gilt:

$$J^{+}(q(t)) = \left(J^{T}(q(t))J(q(t))\right)^{-1} J^{T}(q(t)). \tag{5}$$

Die Ableitungen $\dot{q}(t)$ und $\dot{p}(t)$ werden mit Hilfe der Differenzenquotienten $\delta q(t)$ und $\delta x(t)$ für einen Zeitschritt δt angenähert. Das Problem wird auf diese Weise linearisiert. $\delta q(t)$ bzw. $\delta p(t)$ kann auch als Abweichung zur Sollposition $q(t)$ bzw. $p(t)$ aufgefasst werden. Durch iterative Berechnung der Jacobi-Matrix für jede Gelenkwinkelstellung und Lösen der Gleichung

$$\delta p(t) = J^{-1}(q(t))\delta q(t). \tag{6}$$

wird der Endeffektorposition an die gewünschte Position p angenähert.

Algebraische Methoden Um das inverse kinematische Problem auf algebraische Weise zu lösen, ist ein Gleichungssystem aufzustellen, welches die Beziehung zwischen Gelenkwinkelkonfigurationen und Endeffektorpositionen beschreibt. Hierfür sind die Endeffektorposition $T_{TCP} \in \mathbb{R}^{4x4}$ und die Transformation $T \in \mathbb{R}^{4x4}$ von der Basis bis zum Endeffektor des kinematischen Modells gleichzusetzen. Ein Koeffizientenvergleich zwischen den beiden Matrizen liefert 16 Gleichungen, wovon 4 trivial sind. Auflösen nach den unbekannten Gelenkwinkelwerten löst das inverse Kinematikproblem. Oft können aber nicht alle Gelenkwinkel aus den 12 Gleichungen bestimmt. Um die Chance zu erhöhen das Gleichungssystem auflösen zu können, verwenden Verfahren auf algebraischer Basis Kenntnisse über die Transformationskette, um Gleichungen zu erzeugen, welche lediglich eine Unbekannte enthalten und somit einfach gelöst werden können. Diese Gleichungen resultieren durch sukzessives Invertieren der einzelnen Transformationen, welche in T enthalten sind.

Geometrische Methoden Bei einem einfachen kinematischen Modell können auch trigonometrische Eigenschaften ausgenutzt werden, um Gelenkwinkelkonfigurationen zu bestimmten Endeffektorpositionen im kartesischen Raum auszurechnen. Hierzu werden trigonometrische Funktionen und Sinus-/Kosinussätze unter Angabe der Längen der einzelnen Modellsegmente verwendet. Für einen Arm mit zwei Segmenten s_1 und s_2 mit den Längen l_1 und l_2 sind die Gelenkwinkelwerte q_1 und q_2 für zwei Gelenke. Für q_2 gilt Folgendes:

$${p_x}^2 + {p_y}^2 = l_1^2 + l_2^2 - 2l_1l_2cos(q_2), \tag{7}$$

während für q_1 folgende Gleichung gilt:

$$l_2^2 = {p_x}^2 + {p_y}^2 + l_1^2 - 2l_1\sqrt{{p_x}^2 + {p_y}^2}cos(q_1). \tag{8}$$

Durch Auflösen der beiden Gleichungen 7 und 8 wird eine Lösung für die inverse Kinematik ermittelt. Für größere kinematische Ketten sind geometrische Verfahren ungeeignet, da die aufzulösenden Gleichungen an Komplexität zunehmen und nicht eindeutig zu lösen sind.

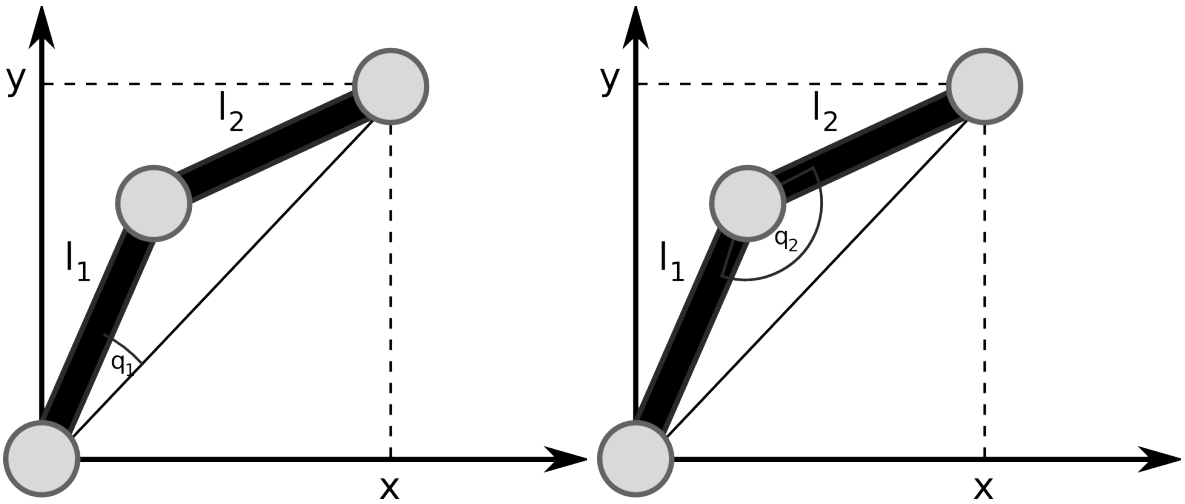

Abbildung 1: Ein Zwei-Gelenke-Arm mit zwei Segmenten der Länge l_1 und l_2. q_1 beschreibt die Gelenkposition vom Basisgelenk, q_2 die Gelenkposition vom Zwischengelenk. $p = (x, y)$ bezeichnet eine kartesische Koordinate im 2D-Raum.

3.1 Komprimierte Bewegungsrepräsentation

Eine direkte Nutzung der aufgenommenen Bewegungsdaten ist unpraktisch und aufwendig. Insbesondere bei höherer Abtastrate und großer Anzahl der Marker bzw. Bewegungsfreiheitsgrade ist der Bedarf an Speicher und Bandbreite sehr groß. Um die relevante Information kompakt und schnell zugänglich zu machen, ist eine Kompression notwendig. Dafür nutzt man die räumliche und zeitliche Kohärenz der Bewegungsdaten aus. Die Kompression lässt sich meist durch eine Abbildung aus dem hochdimensionalen Raum $\mathbb{R}^N$ in einen niederdimensionalen Merkmalsraum $\mathbb{R}^P$ beschreiben. Eines der bekanntesten Verfahren ist die Hauptkomponentenanalyse (engl. Principal Component Analysis), häufig als PCA abgekürzt. Ausgehend von einem Datensatz X mit L Datenpunkten $x_l = (x_1^1, \ldots, x_1^N)^T$

$$X = \begin{pmatrix} x_1^1 & . & . & . & x_L^1 \\ . & . & & & . \\ . & & . & & . \\ . & & & . & . \\ x_1^N & . & . & . & x_L^N \end{pmatrix} \tag{9}$$

werden Basisvektoren $V = \{v_1, \ldots, v_P\}$ gesucht, die einen Vektorraum aufspannen, in welchen x_l durch $\bar{x}_l \in \mathbb{R}^P$ repräsentiert wird. Um den Verlust durch die Projektion in den niederdimensionalen Merkmalsraum zu minimieren, werden die Basisvektoren ausgewählt, welche die Varianz der Merkmale maximieren und die Korrelation zwischen den unterschiedlichen Merkmale minimieren. Die Varianz eines Merkmals i beschreibt

die gemittelte Abweichung vom Mittelwert des Merkmals und wird wie folgt beschrieben:

$$var(x^n) = \frac{\sum_{i=1}^{L} (x_i^n - \hat{x}^n)(x_i^n - \hat{x}^n)}{(L-1)}. \tag{10}$$

Der Mittelwert wird mit folgender Gleichung berechnet:

$$\hat{x}^n = \frac{\sum_{i=1}^{L} x_i^n}{L}. \tag{11}$$

Mittels der Kovarianz lässt sich untersuchen, wie sich die Varianzen zweier Merkmale i und j zueinander verhalten. Die Kovarianz zwischen i und j kann wie folgt bestimmt werden:

$$cov(x^m, x^n) = \frac{\sum_{i=1}^{L} (x_i^m - \hat{x}^m)(x_i^n - \hat{x}^n)}{(L-1)}. \tag{12}$$

Für die Datenmatrix X mit N Merkmalen kann basierend auf Gleichung 12 die Kovarianzmatrix $C \in \mathbb{R}^{NxN}$ aufgestellt werden:

$$C = \begin{pmatrix} cov(x^1, x^1) & . & . & . & cov(x^N, x^1) \\ . & . & & & . \\ . & & . & & . \\ . & & & . & . \\ cov(x^1, x^N) & . & . & . & cov(x^N, x^N) \end{pmatrix} \tag{13}$$

Die Hauptkomponenten werden mittels Berechnung der Eigenvektoren und -werte der Kovarianzmatrix bestimmt. Dieses entspricht einer Diagonalisierung der Kovarianzmatrix und hat eine Dekorrelation der Daten zur Folge. Für einen Eigenwert λ gilt:

$$det\,(C - I\lambda) = 0. \tag{14}$$

Für den Eigenvektor v, welcher mit λ korrespondiert, gilt folgende Gleichung:

$$Cv = \lambda v. \tag{15}$$

Für die Definition des Merkmalsraumes werden aus der Menge aller Eigenvektoren Basisvektoren selektiert. Der Eigenvektor mit dem größte Eigenwert entspricht dem Basisvektor, welcher die größte Varianz beschreibt. Um die Hauptachsen zu bestimmen sind daher die Eigenvektoren in absteigender Reihenfolge zu sortieren, also v_1 mit dem Eigenvektor mit dem größten Eigenwert, anschließend v_2 mit dem zweitgrößten Eigenwert usw. Um die Abbildung des Datenpunktes x_l auf einen Punkt $\bar{x}_l$ des Merkmalsraumes, der durch die ersten P Hauptachsen aufgespannt, zu bestimmen, ist folgende Gleichung zu lösen:

$$\bar{x}_l = (x_l - \hat{x})V \tag{16}$$

Analog kann das ursprüngliche Datum x_l mit Hilfe der Gleichung:

$$\tilde{X}_l = V^T \bar{x}_l + \hat{x} \tag{17}$$

rekonstruiert werden, wobei $\tilde{X}$ lediglich eine Annäherung darstellt.

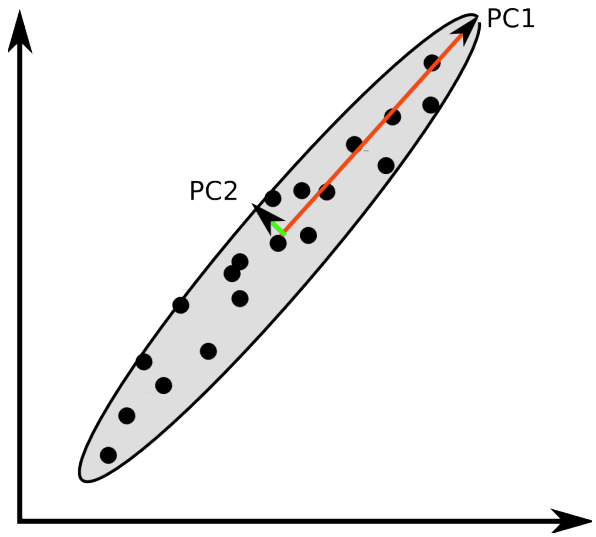

Abbildung 2: Hauptkomponentenanalyse mit zwei Hauptachsen (Principal Components) PC1 und PC2.

4 Segmentierung von Bewegungsdaten

Die Segmentierung der Bewegungsdaten kann sowohl räumlich als auch zeitlich erfolgen. Unter der räumlichen Segmentierung wird die Organisation der Bewegungsdaten in Bewegungsklassen verstanden. Hierfür werden ähnliche Bewegungsegmente zusammengefasst, was z.B. mittels Clustering-Verfahren erreicht werden kann. Die zeitliche Segmentierung hingegen beschreibt die Unterteilung einer längeren Bewegungssequenz in Bewegungssegmente, wobei jedes Segment eine Aktion bzw. eine Grundbewegung darstellt und sich die gesamte Sequenz aus diesen Aktionen zusammensetzt. Die grobe Vorgehensweise besteht aus einem Abtasten der Sequenz und der Identifikation markanter Posen, welche auf eine Ende bzw. Start einer Aktion hindeuten.

Verfahren A In [5] werden Bewegungssequenzen anhand der Gelenkwinkelgeschwindigkeiten segmentiert. Basierend auf der Annahme, dass jede neue Aktion aus einer Ruheposition gestartet wird, wird eine Bewegungssequenz an Zeitpunkten unterteilt, zu denen die Gelenkwinkelgeschwindigkeiten minimal werden. Um im Gelenkwinkelkonfigurationsraum sämtliche N Freiheitsgrade zu berücksichtigen, wird hier bei ein Maß $z(t)$ definiert, welches die Bewegungsveränderungen $\dot{x}_t \in \mathbb{R}^N$ zu Zeitpunkt t wie folgt bescheibt:

$$z(t) = \sum_{i=1}^{N} (\dot{x}_t^i)^2. \tag{18}$$

Mit Hilfe eines empirisch ermittelten Schwellwertes α wird die Bewegung an einem Zeitpunkt t_s segmentiert, falls gilt:

$$z(t_s) < \alpha. \tag{19}$$

Für Bewegungen im kartesischen Raum kann das Verfahren gleichermaßen angewendet werden. Zu beachten ist hierbei, dass für Positionen im 3D-Raum $N = 3$ gilt und für Posen (Position und Orientierung) entsprechend $N = 6$ gilt.

Verfahren B Wie in [3] gezeigt wird, kann die Hauptkomponentenanalyse dazu verwendet werden, Bewegungsdaten zu segmentieren. Dabei wird die Annahme getroffen, dass eine einfache Bewegung sich durch einen Merkmalsvektor mit minimaler Dimension darstellen lässt. Die Zugehörigkeit eines Datenpunktes zu einem Zeitpunkt t zu einem Bewegungssegment wird anhand des Projektionsfehlers bestimmt. Um aus einer Bewegungssequenz $X = \{x_1, \ldots, x_T\}$ ein Bewegungssegment $X_S = \{x_{T_1}, \ldots, x_{T_n}\}$ zu bestimmen, ist zunächst das Zentrum des Bewegungssegment $\hat{x}S \in \mathbb{R}^N$ entsprechend Gleichung 11 zu berechnen. Für X_S ist nun die Hauptkomponentenanalyse durchzuführen, sodass für eine vordefinierte Dimension P die Projektion $\bar{x}_t$ für jeden Datenpunkt x_t in X_S ermittelt werden kann. Für alle Punkte ist nun zu prüfen, ob der jeweilige Projektionsfehler unterhalb eines erlaubten Schwellwertes β liegt. Mit der Rückprojektion $\tilde{x}_t$ kann dies wie folgt formalisiert werden:

$$e_t = \|x_t - \tilde{x}_t\| < \beta. \tag{20}$$

Ist dieses für einen Datenpunkt $x_{T_{n+1}}$ nicht der Fall, wird ein neues Segment definiert. Um die gesamte Sequenz zu verarbeiten, kann nun iterativ vorgegangen werden, indem beginnend von einem Zeitpunkt T_1 in jeder Iteration ein Datenpunkt in der Hauptkomponentenanalyse mitberücksichtigt wird (1. Iteration $x_S = \{x_{T_1}, x_{T_2}\}$, 2. Iteration $x_S = \{x_{T_1}, x_{T_2}, x_{T_3}\}$, ...). Bei Überschreitung des Schwellwertes zum Zeitpunkt T_{n+1} ist dann $T_0 = T_{n+1}$ zu setzen.

5 Seminar-Aufgaben

Die Aufgabe bei diesem Seminar-Thema besteht darin, selbst aufgenommene Bewegungssequenzen nach den beiden vorgestellten Verfahren zu segmentieren. Hierfür sind die Daten, die mit dem VICON-System aufgenommen wurden, zu verarbeiten und im kartesischen sowie im Gelenkwinkelraum darzustellen. Die Segmentierungspunkte sollen dann ermittelt und visualisiert werden.

Zur Vorbereitung sind die Unterlagen durchzuarbeiten. Bei Aufgabe 1 erfolgt die Aufnahme von Bewegungen mit dem VICON-System des Lehrstuhls, Aufgaben 2-4 erfordern eine *schriftlich* Ausarbeitung.

5.1 Aufnahme von Bewegungsdaten

Aufgabe 1

Zu Beginn des Seminars sollen verschiedene Bewegungsaufnahmen mit dem VICON-Motion-Capturing-System aufgenommen werden.

Arbeiten Sie sich (mit Hilfestellung) in die Funktionsweise des VICON-Systems und die Bedienung der Software „Vicon Nexus“ ein und legen Sie einen oder mehrere Probanden für die Aufnahmen fest sowie die Bewegungsabläufe, welche Sie aufnehmen möchten. Sie sollten unterschiedliche Bewegungsabläufe wie z.B. Laufen, Greifen (stehend am Tisch), Kniebeugen usw. für ihre Segmentierungsversuche nachfolgenden aufnehmen.

5.2 Theorie-Teil

Aufgabe 2

Berechnen Sie für den in Abbildung 1 dargestellten Arm die Gelenkwinkelkonfiguration für die 3 kartesischen 2D-Koordinaten:

$$(10, 10), (25, 10), (5, 20) \tag{21}$$

entsprechend der geometrischen Vorgehensweise. Die Segmente haben die Längen $l_1 = 20$ und $l_2 = 10$.

Aufgabe 3

Gegeben sei die Datenmatrix:

$$X = \begin{pmatrix} 3 & 0 & 1 \\ -4 & 1 & 2 \\ -6 & 0 & 2 \end{pmatrix} \tag{22}$$

Führen Sie die Hauptkomponentenanalyse für X durch und projizieren Sie die Datenpunkte in einen 2D-Merkmalsraum ($P = 2$).

Aufgabe 4

Schreiben Sie den Algorithmus für die zwei vorgestellten Verfahren in Pseudocode-Form auf.

5.3 Praxis-Teil

Zur Programmierung der Aufgabe wird Ihnen ein Framework zur Verfügung gestellt, welches die Funktionalität bezüglich dem Einlesen von Bewegungsdaten und der Darstellung dieser Daten realisiert. Desweiteren ist es erforderlich, aus Ihren aufgenommenen Bewegungsdaten die Position von je drei für die Segmentierung relevanten Markern[1] zu extrahieren, was mit Hilfe des MMM-Frameworks geschehen kann. Weitere Informationen hierzu erhalten Sie von Ihrem Betreuer zu Beginn des Praxis-Teils.

Aufgabe 6

Segmentieren Sie die aufgenommenen Bewegungssequenzen „von Hand“. Legen Sie hierzu Frame-Nummern geeigneter Segmentierungspunkte fest.

Aufgabe 7

Berechnen Sie unter Verwendung der geometrischen Methode die Gelenkwinkeltrajektorie zu den gegebenen Bewegungssequenzen im kartesischen Raum und legen Sie diese in eine Datei ab.

Aufgabe 8

Segmentieren Sie die Bewegungsequenzen, repräsentiert im Gelenkwinkelraum und im kartesischen Raum, mit Hilfe der Methode A und visualisieren Sie die Segmentierungspunkte.

Angewendet in welchem Raum funktioniert die Segmentierung besser?

Aufgabe 9

Segmentieren Sie die Bewegungsequenzen, repräsentiert im Gelenkwinkelraum und im kartesischen Raum, mit Hilfe der Methode B und visualisieren Sie die Segmentierungspunkte.

Angewendet in welchem Raum funktioniert die Segmentierung besser?

Literatur

[1] A. Aggarwal and B. Triggs. Recovering 3d human pose from monocular images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(1):44–58, 2006.

[1] Beim Bein z.B. Fuß, Knie und Hüfte, beim Arm Hand, Ellenbogen und Schulter.

[2] P. Azad, T. Asfour, and R. Dillmann. Robust Real-time Stereo-based Markerless Human Motion Capture. In *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, pages 700–707, Daejeon, Korea, December 2008.

[3] Jernej Barbic, Jia-Yu Pan, Christos Faloutsos, Jessica K Hodgins, and Nancy Pollard. Segmenting motion capture data into distinct behaviors. In *In Proceedings of Graphics Interface 2004*, pages 185 – 194, May 2004.

[4] D. Gavrila and L. Davis. 3-D Model-based tracking of humans in action: a multi-view approach. In *Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 73–80, San Francisco, USA, 1996.

[5] Odest Chadwicke Jenkins and Maja J. Mataric. Automated derivation of behavior vocabularies for autonomous humanoid motion. In *Proceedings of the second international joint conference on Autonomous agents and multiagent systems*, pages 225–232, New York, NY, USA, 2003.

[6] S. Knoop, S. Vacek, and R. Dillmann. Fusion of 2D and 3D sensor data for articulated body tracking. *Robotics and Autonomous Systems*, 57(3):321–329, 2009.