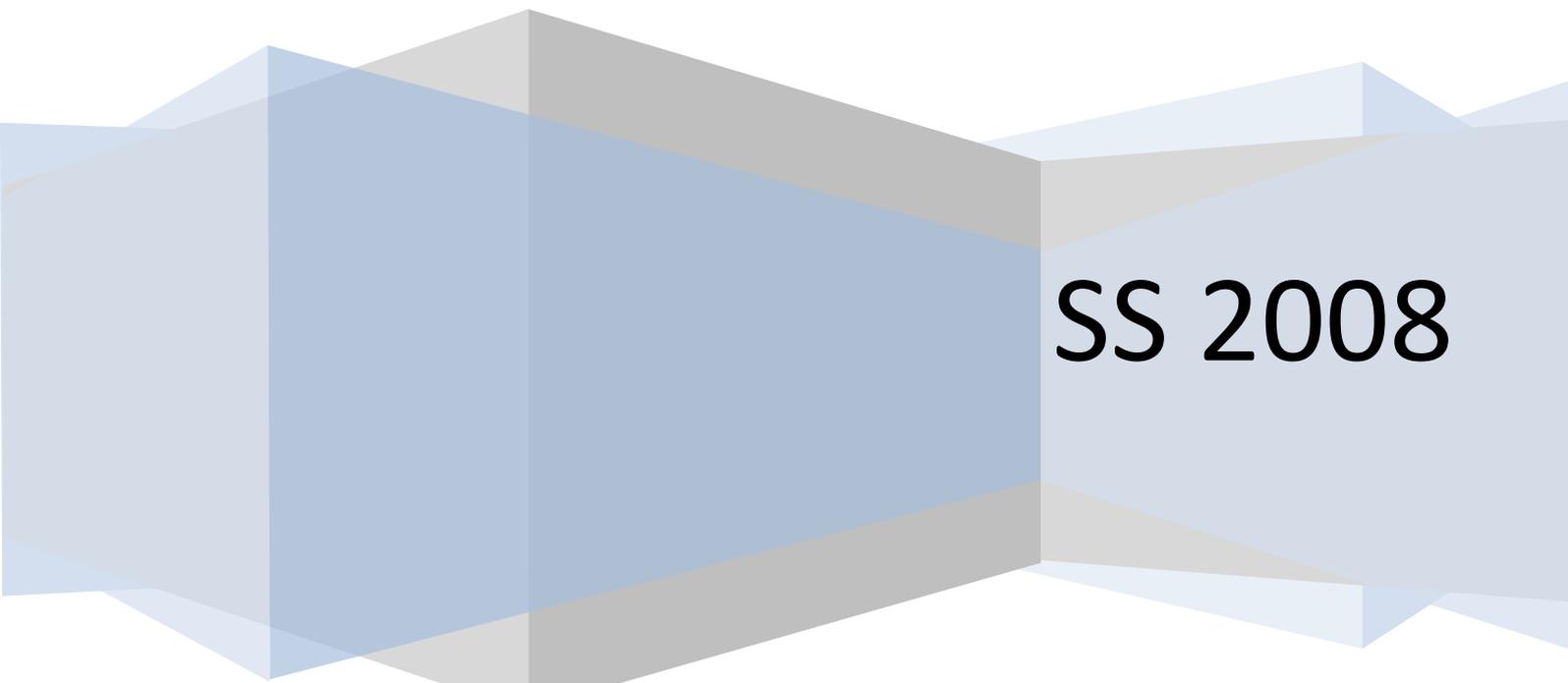


Visual Analysis of Human Motion

Zusammenfassung der Folien aus der Vorlesung

Nora Meiller



SS 2008

Ich übernehme keine Haftung für etwaige Fehler ;)

Bezüglich Copyright für die Bilder und den Text: siehe VO Folien

http://www.ims.tuwien.ac.at/teaching/human_motion/ss08/vo/vo_material.php

Inhaltsverzeichnis

1. Einheit: Bewegungsvisualisierung	3
Darstellung von Bewegung.....	3
Beurteilungskriterien für die Bewegungsdarstellung.....	5
Motion History Images (MHI).....	5
Comics aus Videos	5
Tanz	6
Sport	8
Bewegungsstudie	8
2. Einheit: Gastvortrag ArsenalResearch.....	9
1. Videoaufnahme / Analyse im Westbahnhof Wien	9
2. Aspekte der Personenzählung	11
3. Globale Analyse (Crowd Analysis)	12
3. Einheit: Eyetracking	13
Parameter bei Analyse von Augenbewegungen	13
Anwendungsbereiche.....	13
Studie A: Analyse von Websites	14
Studie B: Erkennung von Fotomontagen	14
Studie C: Schipanoramakarten	15
4. Einheit: Bewegungsstile	16
Wozu dienen Style Machines?	16
Bewegungsdaten	16
Technischer Hintergrund.....	16
Umwandlung von Bewegungsdaten in HMMs.....	17
Umwandlung von Choreographien in HMMs.....	17
Training einer Style Machine.....	17
Style Machine Resultate.....	18
5. Einheit: Sportanwendungen.....	18
Schwimmen	18
Volleyball	19
Tennis	19
Fußball	19

1. Einheit: Bewegungsvisualisierung

Darstellung von Bewegung

in einem statischen Bild mittels 5 verschiedener Ansätze:

- Asymmetrie
- Stroboskopische Effekte
- Vorwärtsneigung und affine Verzerrung
- Bewegungsunschärfe
- Bewegungslinien und Vektoren

a) Asymmetrie:

Skulptur: Zwei Ansichten der Bronzestatue Discobolos (Der Diskuswerfer)
Durch die Technik des „contrapposto“ (Hüften und Beine zeigen in eine andere Richtung als Schulter und Arme) wird der Eindruck von Bewegung erzeugt.

b) Stroboskopischer Effekt:

Die Bewegung des Objekts wird durch eine Folge von zeitdiskreten statischen Bildern dargestellt („Mehrfachaufnahmen“). Stroboskopische Effekte wurden in der Kunst bereits lange vor dem Aufkommen der Fotografie eingesetzt (Vetruvian Man (~1494) von Leonardo da Vinci).

- Wissenschaftliche Illustration: Ein weiteres Beispiel für diese Technik zeigen die Illustrationen aus Descartes' „Meditationen“. Der Zweck der Abbildungen war die Darstellung der Pendelbewegungen einer Schleife bzw. eines Stabes. Die bewegten Objekte sind in 3 verschiedenen Positionen gezeigt, wobei die Hand (idealisiert) als unbewegt dargestellt wird.
- Prähistorische Darstellung: Stroboskopische Effekte bereits in steinzeitlichen Darstellungen? Die Darstellung eines Bisons mit 2 rechten Vorderbeinen. Stand dahinter die Absicht Bewegung darzustellen? (Alter der Abbildung: ca. 32.000 Jahre)
- Mehrfachfotografien: Pionierarbeiten in der Fotografie durch E. Muybridge und E.-J. Marey im späten 19. Jahrhundert. Die neu aufkommenden Serienfotografien wurden durch technische Entwicklungen für schnellen Film und schnelle Beleuchtung stimuliert. Die Aufnahmen gaben neuartige Einblicke in Bewegungsabläufe bei Menschen und Tieren.
- Bewegungsstudien an Tieren: Durch die fotografischen Aufnahmen von Muybridge gelang erstmals der Nachweis, dass ein galoppierendes Pferd zeitweise keinen Bodenkontakt hat.
- Menschliche Bewegung: Stroboskopische Studie menschlicher Bewegung von E.-J. Marey (franz. Mediziner und Naturwissenschaftler, 19. Jh.). Aufgenommen wurde die Laufbewegung einer Person in schwarzem Anzug mit weißen (Streifen-) Markierungen entlang Armen, Beinen, Kopf und Schulter.
- Weiterreichende Einflüsse: Die neuartigen fotografischen Darstellungen beeinflussten weiterführende Entwicklungen in den Naturwissenschaften und der Kunst.
 - Medizin und Orthopädie (z.B. Ganganalysen für Rehabilitation)
 - Erforschung menschlicher Bewegung im Sport (Biomechanik, Trainingslehre, ...) einschließlich Aufbereitung der erfassten Bewegungen für den Zuseher
 - Zoologie

- Kunst nach 1900

Beispiel 1 (*Dynamik eines Hundes an der Leine* (1912) von G. Balla): Stroboskopische Effekte sind erkennbar an den „zahlreichen“ Hundepfoten sowie den Mehrfachdarstellungen der Beine des Begleiters und der Hundeleine. Die Richtung der Bewegung ist jedoch (wie bei den meisten stroboskopischen Darstellungen) nicht eindeutig erkennbar, sondern wird nur durch die Erfahrung des Betrachters hineininterpretiert.

Beispiel 2 (*Woman Picking Flowers* (1910-1911) von F. Kupka): Überlagerung von mehreren Darstellungen einer Frau, die sich vom Sessel erhebt, einige Schritte geht und sich vorbeugt. Im Laufe der Bewegung ändern sich die Farben von orange über rot und blau nach grün.

- Anwendungsbereich Sport: Eiskunstlauf, Weitsprung, Reitsport, Schisprung und Fußball

c) Vorwärtsneigung und affine Verzerrung:

Beispiel für Vorwärtsneigung: Der Eindruck von Geschwindigkeit wird vor allem durch die Vorwärtsneigung des Oberkörpers vermittelt. Zusätzliche Indikatoren der Bewegung sind die Lage von Hut und Schal sowie die Beinhaltung. Vorteile dieser Darstellungsart sind, dass das Objekt klar erkennbar ist und auch die Richtung der Bewegung ersichtlich ist.

Beispiel für affine Verzerrung: Durch entsprechende technologische „Tricks“ bei der Aufnahme wird eine affine Verzerrung (Vorwärtsneigung des Autos, Rückwärtsneigung der Personen) erzeugt. Ein Problem der affinen Verzerrung bei wissenschaftlichen Darstellungen ist die Verfälschung der Objektformen und -eigenschaften (Verzerrung von Objekten, die in Wirklichkeit starr sind).

d) Bewegungsunschärfe:

Bewegungsunschärfe in der Fotografie: Die scharfe Aufnahme des Gesichtes steht in Kontrast zur verschwommenen Darstellung der bewegten Arme und Bluse sowie der bewegten Äste im Hintergrund durch entsprechend lange Belichtungszeit. Zwei offensichtliche Probleme dieser Art der Bewegungskunst sind die zunehmende Unschärfe der Objekte sowie das Fehlen von Information über die Richtung der Bewegung.

e) Bewegungslinien und Vektoren:

Bewegungslinien und Vektoren sind die häufigsten Darstellungsformen von Bewegung in wissenschaftlichen und künstlerischen Darstellungen.

Vorteile:

- Das bewegte Objekt bleibt klar sichtbar (z.B. keine Bewegungsunschärfe).
- Die Richtung der Bewegung kann durch Vektoren und Bewegungslinien angezeigt werden (keine Deformation des Objekts nötig).

Bewegungslinien (motion lines) werden in der Literatur auch häufig als *action lines*, *speed lines* oder *zip-ribbons* bezeichnet.

Vektoren werden üblicherweise durch Pfeile gezeichnet, aber solange der Ausgangspunkt und der Endpunkt der Bewegung unterscheidbar sind, können auch „normale“ Linien verwendet werden. Durch Richtung und Länge der Vektoren werden die Richtung der Bewegung und die Geschwindigkeit angegeben.

Beispiel Fotografie: Bei entsprechend langer Aufnahmezeit (mehrere Minuten) bewegte sich eine Person durch die nächtliche Straße und schwang die Lichtquelle. Die Person selbst ist durch die Bewegungsunschärfe im Bild nicht mehr sichtbar.

Beispiel Comics: Als Bewegungslinien werden häufig Geraden oder Teile von Kreisen bzw. Ellipsen verwendet.

Beurteilungskriterien für die Bewegungsdarstellung

- ✓ Wie gut wird der Eindruck von Bewegung dem Betrachter vermittelt? (*evocativeness*)
- ✓ Wie klar ist das Objekt dargestellt/erkennbar? (*clarity of object*)
- ✓ Kann man die Richtung der Bewegung erkennen? (*direction of motion*)
- ✓ Ist dazu Vorwissen beim Betrachter nötig?
- ✓ Genauigkeit (*precision*)
- ✓ Kann man das Ausmaß der Bewegung ablesen?
- ✓ Ist das vorrangige Anwendungsziel eine präzise wissenschaftliche Darstellung oder die Erzeugung eines ästhetischen Effekts?

• Anwendung der Kriterien auf die vorgestellten Methoden

Methode	<i>Evocative-ness</i>	<i>Clarity of object</i>	<i>Direction of motion</i>	<i>Precision</i>
Asymmetrie	o.k.	o.k.	¬, o.k.	x
Stroboskopischer Effekt	o.k.	o.k.	x	o.k.
Vorwärtsneigung & affine Verzerrung	o.k.	o.k.	o.k.	x
Bewegungsunschärfe	o.k.	x	x	x
Bewegungslinien & Vektoren	o.k.	o.k.	o.k.	o.k.

Motion History Images (MHI)

Videoaufnahme einer Bewegungssequenz (ausgewählte Frames) \Rightarrow Den einzelnen Frames zugeordnete Motion History Bilder: Der „Blick in die Vergangenheit“ der Bewegungsfolge wird durch die Grauwertkodierung ermöglicht.

Parametervariation: Durch Wahl der entsprechenden Parameter kann man die „Merkzeit“ des Systems variieren.

Comics aus Videos

(für Bilder siehe Folien 28 – 39)

Ausgangsbildsequenz:

Der Ausgangsdatensatz ist ein mittels 2 räumlich versetzter Kameras aufgenommenes Stereovideo. Die Abbildung zeigt jeden 10. Frame der von einer der Kameras aufgenommenen Bildfolge.

Extraktion der Objektkonturen:

Die oberen Abbildungen auf Folie 29 zeigen ein Stereo-Bildpaar des aufgenommenen Stereo-Videos. Die Abbildung links unten zeigt das aus den 2 oberen Abbildungen mittels Stereoanalyse berechnete Tiefenbild. Hellere Grauwerte bedeuten geringeren Abstand zur Kamera (= geringere Tiefe). Das Bild rechts unten zeigt das aus dem Tiefenbild abgeleitete Konturbild. Genauer gesagt, wurde das Konturbild aus den Diskontinuitäten (= Kanten) im Tiefenbild unter Verwendung geeigneter

Bildverarbeitungsalgorithmen berechnet. Das Konturbild ist ein wichtiges Zwischenprodukt zur nachfolgenden Erzeugung Comic-ähnlicher Darstellungen.

Tracking:

Das Bild links oben auf Folie 31 zeigt ein Ausgangsbild mit den darauf (automatisch) ausgewählten Tracking-Punkten. Die Auswahl der Tracking-Punkte und das nachfolgende Tracking erfolgen unter Verwendung des KLT (Kanade-Lucas-Tomasi) Feature-Trackers.

Das Ziel des Trackings ist, die Bewegung signifikanter Bildobjekte zu erkennen um daraus anschließend die Bewegungslinien zu berechnen.

Das Bild rechts oben auf Folie 31 zeigt den durch das automatische Tracking ermittelten Bewegungspfad eines selektierten Punktes.

Berechnung der Bewegungslinien:

In Imitation eines Comic-ähnlichen Stils wollen wir „einfache“ Bewegungslinien (Ausschnitte aus Gerade, Kreis, Ellipse, etc.) erzeugen. In den tatsächlichen Bewegungspfad eines Punktes wird auf Folie 31, Bild links unten, ein Kreis eingepasst. Um mehrfache Bewegungslinien zu generieren, kann in dem Beispiel auf Folie 31, Abbildung rechts unten, der Benutzer nun interaktiv weitere Punkte auf der Kontur auswählen, die Ausgangspunkt weiterer Bewegungslinien sein sollen. Für die weiteren Bewegungslinien wird dieselbe geometrische Darstellung wie für die ursprüngliche verwendet, um ein einheitliches Aussehen der Mehrfachlinien zu gewährleisten.

Diskussion der Ergebnisfolie:

Aus den Originalbildern (siehe obere Bildreihe auf Folie 34) wurde zunächst eine stilisierte Darstellung in Imitation einer handgezeichneten Skizze erzeugt (Bild links unten). Die automatische Erzeugung der Skizze erfolgte nach einem Algorithmus, der in [Markovic05] beschrieben wird. Der stilisierten Skizze wurden dann die aus der echten Bildfolge abgeleiteten Bewegungslinien in der Form mehrerer identischer Kreisbögen überlagert.

Die durch das Endergebnis (Bild rechts unten auf Folie 34) suggerierte Bewegung kann mit der tatsächlichen Bewegung innerhalb der Original-Bildfolge auf Folie 28 verglichen werden.

Diskussion der Mehrfachkonturen:

Die Beispiele auf den Folien 36 und 37 zeigen eine weitere Variante der Comic-ähnlichen Darstellung der Bewegung durch Kombination von Bewegungslinien mit Mehrfachkonturen (vgl. stroboskopischer Effekt). Die Konturen wurden aus dem Silhouettenbild von Folie 29 abgeleitet. Verschiedene Varianten (z.B. Teilkonturen) sind möglich.

Wurden die Konturlinien der ursprünglichen Videosequenz in regelmäßigen zeitlichen Abständen entnommen, geben sie die Geschwindigkeitsänderungen realitätsgetreu wieder.

Tanz

(für Bilder siehe Folien 40 – 52)

Menschliche Tanzbewegungen stellen besondere Anforderungen an die Bewegungsaufnahme, Analyse und Visualisierungsverfahren (z.B. anspruchsvoller als biomechanische Gangstudien) wegen ihrer Charakteristika:

- Schnelligkeit
- Feinheit/Präzision
- Ausdrucksstärke

Neue Kameratechnologien erlauben relativ kostengünstige Videoaufnahmen komplexer Ganzkörperbewegungen. Die Entwicklung von Verfahren zur automatischen Erfassung und Archivierung von Tanzbewegungen und Choreografien steckt noch in den Kinderschuhen.

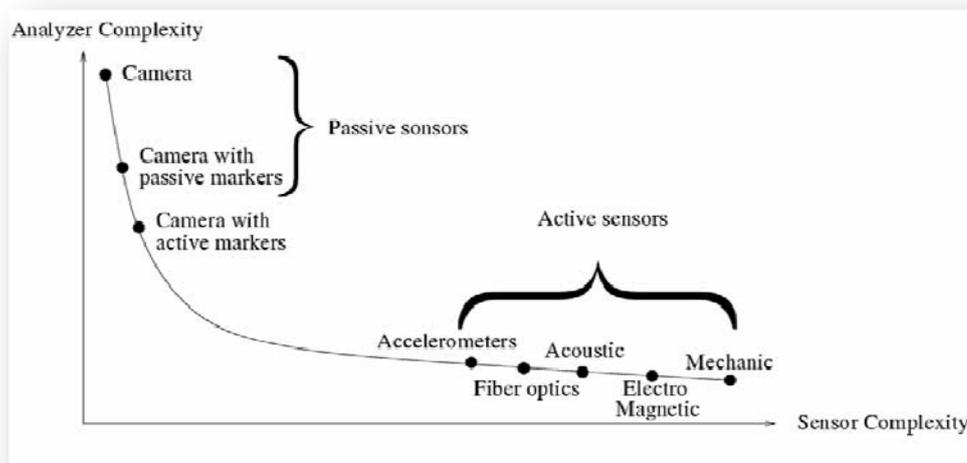
⇒ Tanznotation zur Kodierung und Visualisierung der Körperbewegung und Schrittfolgen

Eine spezielle Tanznotation: Labanotation

Die Labanotation – benannt nach dem österreichischen Tänzer und Choreographen Rudolph Laban (1879-1958) – wird von Choreographen und Tanz-Notatoren benutzt um menschliche Körperbewegung aufzuzeichnen. Sie basiert auf einer klar konstruierten Grammatik und Syntax für Bewegungen und liefert somit eine gute Grundlage für computergestützte Verarbeitung. Labansymbole beschreiben z.B.

- die Richtung einer Bewegung
- welcher Körperteil die Bewegung ausführt
- die Zeitdauer einer Bewegung

Sensoren für Bewegungsaufnahmen



- Künstlerische Visualisierung einer Tanzbewegung: „*Biped – A Dance With Virtual and Company Dancers*“
Unter der Leitung des Choreographen Merce Cunningham interagieren auf der Bühne echte mit virtuellen Tänzern. Die (virtuellen) Tanzbewegungen wurden mittels „Motion Capture“ Verfahren aufgenommen und danach stilisiert dargestellt.
Am Körper wurden Sensoren platziert. Die Bewegungsverfolgung (optisches Tracking) erfolgte durch ein System von 10 Kameras mit einer Aufnahmezeit von 60 Frames/sec für jeden Sensor. Die aufgenommenen 3D Punkte erlauben die nachfolgende Betrachtung und Visualisierung der Bewegung von einem frei wählbaren Blickpunkt.
„Motion Capture“ spielt auch eine wichtige Rolle im Hinblick auf die Archivierung von Bewegung.
Die aufgenommenen Bewegungspunkte bilden die Grundlage für die nachfolgende Visualisierung und Animation. Stilisierte Zeichnungen werden in die aufgenommenen Bewegungspunkte eingepasst.
Darstellung auf der Bühne: echte Tänzer interagieren mit den projizierten Animationen der davor aufgenommenen Bewegungssequenzen in ästhetischer Form.
- Visualisierung einer Tanzbewegung durch einen Roboter: „*Imitative Dance Performance by a Biped Humanoid Robot*“
Visualisierung eines traditionellen japanischen Tanzes durch einen Roboter nach Beobachtung eines menschlichen Vorbildes

Software:

Ein Beispiel einer kommerziellen Software zur Visualisierung & Animation im Bereich Tanz/Choreografie ist *Credo Interactive*.

Sport

(für Bilder siehe Folien 53 – 58)

- Tischtennis: Darstellung der berechneten Bewegung mittels Bewegungsvektoren
- Golf: Die Bildfolge auf Folie 54 zeigt das Ergebnis einer 3D Rekonstruktion aus der Videoaufnahme eines Golfschwungs. Bei der 3D Rekonstruktion werden gewisse Voraussetzungen über den Bewegungsablauf gemacht (modellbasiertes Verfahren). Aus den rekonstruierten 3D Koordinaten wird ein „Skelett“ berechnet um die Bewegung zu visualisieren ⇒ Überlagerung des Skeletts mit der ursprünglichen Videoaufnahme
Die Folien 56 und 57 zeigen eine Visualisierung der berechneten Bewegungssequenz mit Hilfe eines volumetrischen Körpermodells.
Folie 58 zeigt die Visualisierung des Bewegungspfad des Golfschlägers bei der Darstellung einiger Schlüsselpositionen der Bewegung.

Bewegungsstudie

(für Bilder siehe Folien 59 – 63)

Hochgeschwindigkeitsaufnahme und Animation:

Die Bewegungssequenz wurde mittels einer Hochgeschwindigkeitsvideokamera aufgenommen. Zur besseren Bewegungserkennung und -verfolgung wurden am Körper über der dunklen Kleidung helle Marker angebracht.

Markerbasierte Bewegungsverfolgung:

Zur automatischen Lokalisierung des Markers im nachfolgenden Frame wurde ein Korrelationsverfahren (Identifikation „ähnlicher“ Bildausschnitte) verwendet.

Visualisierung:

1. Die einfachste Visualisierung verbindet lediglich die Punkte, welche den Markerpositionen entsprechen, durch Linien miteinander. Der Kopf wird durch einen Kreis angedeutet. So entsteht ein tanzendes „Strichmännchen“. Es wurden dabei nicht alle Marker verwendet, sondern nur diejenigen, welche die zentralen Körperpositionen erfassen, wie Hals, Bauch, Schritt, die beiden Arme und Beine. Die Position des Kopfes wurde durch Mittelung der Markerpositionen unter den beiden Augen bestimmt.
2. Verbindung von (gezeichneten) Körpersegmenten entsprechend der aufgenommenen, zeitabhängigen Markerpositionen.
⇒ Artefakte: Hand wirkt flach, Unterarm zu stark rotiert, Unterschenkel haftet nicht an Oberschenkel

2. Einheit: Gastvortrag ArsenalResearch

Forschungsziele:

„Human Centered Mobility Technologies“

- **Gewinnung** quantitativer Einsichten in das Bewegungsverhalten von Personen in Infrastrukturen
- **Simulation** des Bewegungsverhaltens von Personen auf Basis der quantitativen Erhebung
- **Visualisierung** des Bewegungsverhaltens
- **Empfehlung** zur „besseren“ Lenkung und Leitung von Personen.



⇒ Einsichten in die praktischen Probleme bei der Videoanalyse in realen Infrastrukturen

1. Videoaufnahme / Analyse im Westbahnhof Wien

Fallstudie: Entwickeln eines Richtsystems basierend auf „landmarks“

langfristige Video-Beobachtungen werden gebraucht um Plätze zu identifizieren, wo Menschen häufig stehen bleiben (= decision points ... Entscheidungspunkte)

Personenspuren bekommt man durch einen People Tracker.

Kamerapositionierung und Videoaufnahme in einer komplexen Infrastruktur wo kein CCTV-System vorinstalliert ist.

⇒ statistische Analyse der gewonnenen Tracks

Wozu Aufnahmeplanung?

Bestimmung der Anzahl + Lage der Kameras!

Weitwinkelobjektiv	Teleobjektiv
+ guter Überblick (bis 90° Blickwinkel)	+ hohe Detailauflösung
- geringe Detailauflösung	- kleines Blickfeld

Kamerapositionierung:

- Strategische Platzierung der Kameras
- Ermittlung aller erforderlicher Brennweiten
- Begrenzte Installationszeit

- Non-permanente Installation begrenzt die Energieversorgung und die Platzierung der Übertragungskabel
 - 100m Begrenzung für die Distanz zwischen den Sensoren und den Aufnahmegegeräten führen zu einer Begrenzung zwischen nördlicher und südlicher Wand
- ⇒ Lösung: Kamerapositionierung und Blickfeldermittlung basierung auf einem 3D-CAD-Plan
 ⇒ 2D-Pläne von den ÖBB wurden als 3D-Modelle umgestaltet (ausgewählte Höhen wurden mit Laser-Abstandsmesser bestimmt). Weiters wurden 7 Kameras wurden virtuell positioniert

Probleme:

- Aktualisierungsrate:
 Zu langsam: Veränderungen in der Helligkeit werden nicht berücksichtigt
 Zu schnell: Stehende Personen werden zu Hintergrund ...
- Phänomen: Beleuchtung/Schattierung
 Spiegelungen können gleiche Form wie Menschen haben
 Schatten bewegen sich mit
- Phänomen: Unterschiedliche Lichtquellen
 Kamera bei Kunstlicht vs. Kamera bei Tageslicht
- Fokussierung ist von Wellenlänge abhängig!

Visuelle Personenverfolgung

Gewinnung Personentrajektorien auf der Grundebene (Füße), inkl. Personenzähler: virtueller „Draht“

Probleme: Bewegungsdaten für Grundriss?

- ⇒ Abbildung Bildpunkte → Koordinaten Grundriss?
- ⇒ Bildpunkte außerhalb der Grundebene → Fehler! (z.B. Rolltreppe)

Datenvorverarbeitung:

- Tracking-Daten für einen Tag wurden in Weltkoordinaten transformiert wobei das normalisierte DLT [Hartley, Zisserman] verwendet wurde.
- Deutliche Ausreißer wurden weggelassen
 - Spuren die außerhalb der Infrastruktur gefunden wurden
 - Tracklets mit fünf oder weniger Zuständen
- Spline-Glättung wird auf die übrigbleibenden Trajektorien angewendet um das Rauschlevel zu reduzieren

Datenerfassung & -analyse:

- Videobasierte Analyse der Personenbewegungen
- Ermittelte Stops: Stops werden definiert als Spur die einen Kreis mit einem Radius < 1m innerhalb 2-3 Sekunden nicht verlässt

2. Aspekte der Personenzählung

Genauigkeit der Zählinstrumente (Counting Devices)

Ähnliche Fallstudie: Evaluierung von Personenzähl-Produkten in London

⇒ Ground truth:

Manuelles Zählen mit Clickers (handheld manual counters)

Ein Mitarbeiter zählt alle Personen die in eine Richtung gehen (für hohe flow rates = Durchsatz).

Dieser muss jede Stunde eine kurze Pause machen um die Konzentration beizubehalten. Bei solchen manuellen Zählungen wird eine durchschnittliche Fehlerrate von 10% angenommen.

(Paralleles) Kamera-Filmmaterial der Szenen würde viel helfen!

- ⇒ Manuelles Zählen mit Video Annotation Software
- ⇒ 100 % Exaktheit möglich (Zeitlupe, wiederholbar, Mehrfachzählungen,...)

- Mall: Active Infrared Line Counts
- Events: Vision-Based + Active Infrared Line Counts
- Public Transport: Computer Vision-Based Line Counts

a) Infrared Sensor in Shopping Mall:

Statistisches Modell für den Zählfehler: $IR_COUNT_t = f(MAN_COUNT_t) + u_t$

Lineares Regressionsmodell

Restwerte hängen von der Flow Rate ab

Modell-Validierung (= Gültigkeit)

90 Zählungen Training Set

30 Zählungen Validation Set

→ Sensoren haben die Personenzählungen um ca. 60 % unterschätzt ⇒ Sensor war nicht korrekt konfiguriert

b) Event Scenario Near Stadium:

Aktives Infrarot übertrifft die sicht-basierten Methoden.

Die Genauigkeit für eine einminütige Zählperiode ist weit unter 90 %. Der Fehler hängt ab von der Personendichte.

Die Fehlerstreuung kann verwendet werden um Massen-Kontroll-Messungen (crowd control measures) zu designen.

c) Public Transport

Fußgängerzählung in genereller Kamerasicht basierend auf Head-Shoulder-Kandidaten

Potentieller Nutzen: CCTV Added Value

Individuelles Fußgänger-Tracking in Mengen jenseits derzeitiger Systeme

Zählansatz: Heuristik basierend auf mehreren Head-Shoulder-Kandidaten

Conclusio:

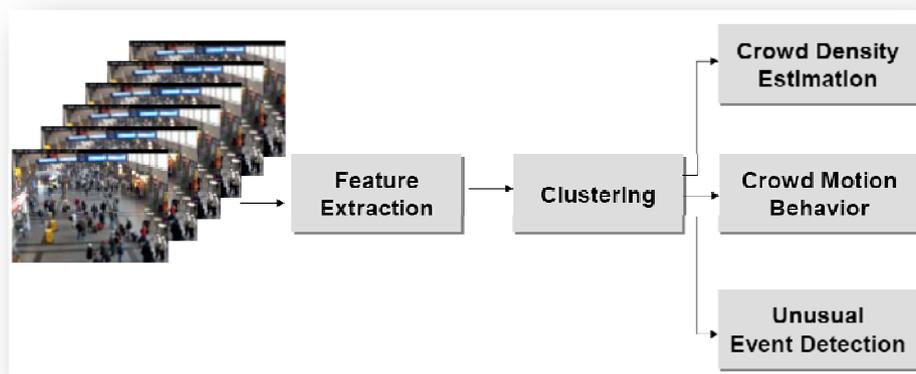
Statistische Modelle für Zählfehler können die Genauigkeit außerordentlich erhöhen/verbessern.

Zählfehler sind für gewöhnlich nicht homogen → der Fehler erhöht sich normalerweise bei höherer Personendichte

Ausblick: 3D processing (stereo) kann Beleuchtungsprobleme besser bewältigen.

3. Globale Analyse (Crowd Analysis)

Conceptual Framework:



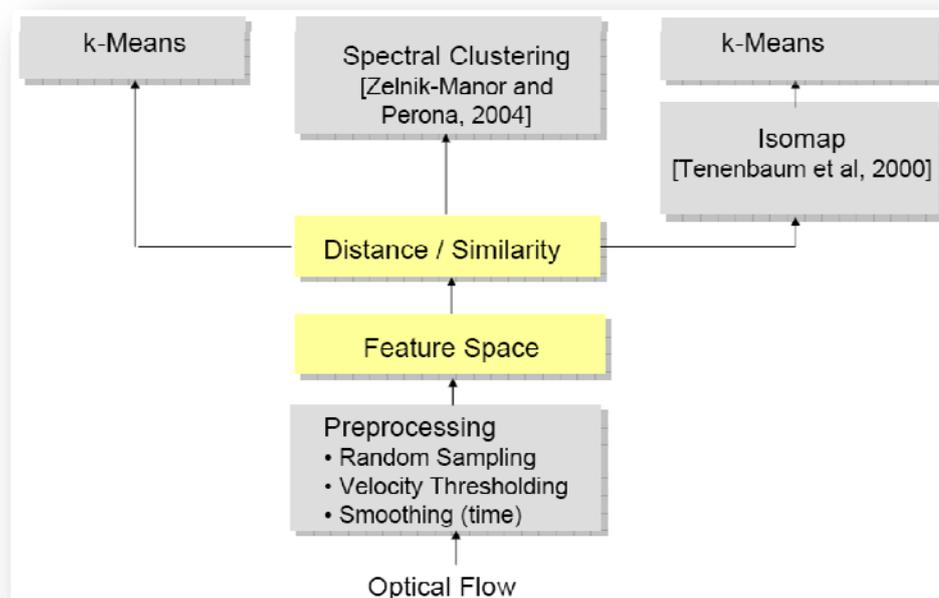
Optical Flow

Optical flow ist das Muster von sichtbarer Objekt-, Flächen- und Kantenbewegung in einer anschaulichen Szene, verursacht durch die relative Bewegung zwischen Kamera und der Szene.

Idee: Flow Vector Clustering

Verwendet zufällige Teilmenge von optischen Flow Vectors (N=1000) → rechnerische Gründe
Verwendet Multiple Samples für Cluster Performance Evaluation

Überblick der ausgewerteten Annäherungen:



Quantitative Evaluation:

- 4 einfache manuell segmentierte real 'ground truth' Data Sets
- viele glaubhafte Wege der Clusterbildung
- Anzahl der Clusters kann oft nicht eindeutig bestimmt/ermittelt werden

⇒ Fußgängerübergang, Westbahnhof, Rolltreppen die in verschiedene Richtungen fahren

3. Einheit: Eyetracking

Parameter bei Analyse von Augenbewegungen

- **Sakkaden:**
Sakkaden sind schnelle, sprunghafte Bewegungen um das Auge neu auszurichten und somit einen anderen Bereich eines Objektes scharf zu sehen. Eine Sakkade dauert von der Planung bis zur Ausführung ungefähr 150-200ms und kann bewusst oder unbewusst ausgelöst werden. Es kommt zu einer geringen Informationsaufnahme während der Dauer einer Sakkade.
- **Sakkadenlatenzen:**
Sakkadenlatenzen können auch als Reaktionszeiten bezeichnet werden, die vom Auftreten eines visuellen Reizes bis zur Auslösung einer Sakkade vergehen; sie dauern ungefähr 200 Millisekunden.
- **Fixationen:**
Während einer Fixation ist das Auge auf einen Bereich fixiert und somit ist hier die Informationsaufnahme möglich. Die Dauer hängt davon ab wie schwierig oder einfach die Verarbeitung des Betrachteten abläuft. Wenn der Bereich schwieriger zu verarbeiten ist oder das Gesehene das Interesse des Betrachters geweckt hat, wird die Fixation länger dauern, beziehungsweise wird es in diesem Bereich mehrere Fixationen geben, die eng beisammen liegen.
- **Scanpath:**
Der Scanpath beschreibt eine fortlaufende Sequenz von Sakkaden und Fixationen wie zum Beispiel Sakkade-Fixation-Sakkade. Der optimale Scanpath hängt von der Aufgabenstellung ab. Bei einer Suchaufgabe wäre zum Beispiel der beste Scanpath eine gerade Linie zum gewünschten Suchobjekt.

Mittels Eye-tracking Kameras werden folgende Informationen aufgezeichnet:

- **Fixationen und ihre Dauer/Anzahl:**
Hot Spots: Darstellung der Dauer einer Fixierung bzw. der Anzahl von Fixierungen mittels Farbkodierung
Die Gesamtheit der Hot Spots wird auch als Aufmerksamkeitslandschaft bezeichnet.
- **zeitliche Reihenfolge der Fixationen:**
Gaze Views zeigen den zurückgelegten Augenpfad der Versuchsperson an, indem die Fixationen in der zeitlichen Abfolge aufsteigend nummeriert werden.

Anwendungsbereiche

- **Medizin/Psychologie**
 - Sämtliche Augenbewegungen
 - Pupillengröße
 - Lidschlüsse
- **Leseforschung**
 - Fixationen (Anzahl, Dauer, Konstanz,...)
 - Sakkaden (Größe, Anzahl, Dauer, Frequenzcharakteristik,...)
- **Werbung/Design**
 - Gesamtverweildauer in einer Bildregion (area of interest)
 - Anzahl an Blickkontakten in einer Region
 - Messer der Zeit bis zum ersten Blickkontakt

- Sicherheitsbereich
- Bild- und Videoverarbeitung/Computergrafik
 - Unterscheidung zwischen Bildregionen, die den Blick des Betrachters besonders anziehen, und solchen, die kaum (oder nur periphär) wahrgenommen werden
 - Anwendungsbeispiele:
 - Computergrafik/Rendering: Adaptives Setzen eines level-of-detail
 - Videoverarbeitung/-kodierung: Unterstützung bei der automatischen Segmentierung von Bildinhalten, Anpassung an den Bildinhalt bei verlustbehafteter Kodierung
 - Verwendung von Eye-tracking als Eingabegerät: z.B. im medizinischen oder militärischen Bereich

Eye-tracking Gerät:

Als Beispiel: Tobii x50 Eye-tracking Gerät

Eigenschaften:

nicht in Display integriert → große Flexibilität

Kalibrierung ist nötig zur Abstimmung auf:

- Geometrie des Messaufbaus
- individuellen Benutzer

Die zugehörige Software liefert:

- Rohdaten
- Abgeleitete Produkte (z.B. Aufmerksamkeitslandschaften)

Studie A: Analyse von Websites

Ergebnisse: Gaze Plots, Hot Spots (Aufmerksamkeitslandschaft)

Studie B: Erkennung von Fotomontagen

Matting von Bildern/Videos:

Hochgenaue Segmentierung mit Zwischenstufen („Semitransparenzen“)

⇒ Wichtig für realistisches Einfügen vor neuem Hintergrund

Matting & Compositing:

Originalbild mit Input des Benutzers (Markierungen: was gehört zum Vorder-, was zum Hintergrund)

→ Matting → Überlagerung mit neuem Hintergrund

Artefakte können bei Bildkompositionen auftreten:

Durch nicht perfektes Trennen des Vordergrundobjekts vom Bildhintergrund im Bereich der feinen Haarstrukturen bleibt die Hintergrundfarbe in den Randbereichen des segmentierten Vordergrundobjekts „kleben“ und wird beim Einfügen vor einem neuen Bildhintergrund als Artefakt sichtbar.

Aufgabenstellung:

Testpersonen sollten für eine Reihe von vorgelegten Bildern entscheiden, ob es sich ihrer Meinung nach um ein Originalbild oder um eine Fotomontage handelt.

Frage: Auf welche Bildelemente konzentriert sich die Aufmerksamkeit der Betrachter, wenn sie versuchen zu entscheiden, ob eine Fotomontage vorliegt?

Ergebnisse: Aufmerksamkeitslandschaft und zeitlicher Blickverlauf

Studie C: Schipanoramakarten

Aufgabenstellung:

Testpersonen sollten mit Hilfe vorgelegter (eingescannter) Schipanoramakarten Orientierungsaufgaben lösen.

Ziel: Aus dem aufgenommenen Blickmuster sollen Rückschlüsse auf die Qualität der Karten (in welchen Bildbereichen hat der Betrachter Schwierigkeiten?) gezogen werden. Die Ergebnisse können zur Verbesserung des Kartendesigns herangezogen werden.

Testmaterial: Panoramakarten des Schigebiets Zell am See (für touristische Zwecke)

Ergebnisse:

Aufmerksamkeitslandschaft: Visualisierung der Anzahl der Fixationen (Summe über alle Testpersonen) bei der aufgabenorientierten Suche auf der Schipanoramakarte.

Zeitlicher Blickverlauf: Visualisierung der Fixationen (mit fortlaufender Nummerierung) einer Testperson bei der aufgabenorientierten Suche auf der Schipanoramakarte.

4. Einheit: Bewegungsstile

Wozu dienen Style Machines?

- Analyse von Bewegungen
 - Kategorisierung von Motion Capture Daten
 - Wie bewegt sich eine Person fort? (Gehen, Laufen, Marschieren,...)
 - Welche Sportart wird ausgeführt? (Hochsprung, Weitsprung,...)
- Verbesserung von Bewegungsabläufen bei Anfängern
 - Optimierung bei Leistungssport (z.B. Absprung bei Hochsprung)
 - Mehr Eleganz/Ausdruck bei Tanzchoreographien
- Erzeugung von neuen Bewegungen
 - Erzeugung von neuen Choreographien (Styles)
 - Variation von bestehenden Tanzchoreographien
 - Übertragung eines Stils (Ausdruck, Eleganz, Schnelligkeit,...) einer Choreographie auf einen anderen Bewegungsablauf
 - Automatische Generierung von Bewegungen von Charakteren in der Unterhaltungsindustrie
 - Zufällige Gehbewegungen für animierte Massenszenen im Kinofilm
 - Generierung von unterschiedlichen Bewegungsmustern von Charakteren in Computerspielen

Bewegungsdaten

Wie können die Bewegungsdaten aussehen?

Für diskrete Zeitpunkte (wie bei Frames in einem Video)
3D Positionen für jedes Gelenk

Wie können die Bewegungsdaten codiert werden?

Vektoren für jeden Zeitschritt

$t_0: (x_1, y_1, z_1, x_2, y_2, z_2, \dots)$

wobei x_i, y_i, z_i die 3D-Koordinaten
von Gelenk i beschreiben

$t_1: (x_1, y_1, z_1, x_2, y_2, z_2, \dots)$

Wie können die Bewegungsdaten visualisiert werden?

Stickfigures, welche die 3D-Positionen der Gelenke verbinden.

Stickfigures können für jeden Zeitschritt gezeichnet werden.

Zwischen signifikanten Gelenken können über die Zeit hinweg Verbindungen gezeichnet werden.

Technischer Hintergrund

Eine Style Machine dient zur Analyse und zur Erstellung von Bewegungsdaten. Dafür ist ein spezieller technischer Aufbau notwendig.

Eine Style Machine bezeichnet eine Familie von *Hidden Markov Models* (HMM).

Ein HMM ist wie eine Finite State Machine mit Übergangswahrscheinlichkeiten und Wahrscheinlichkeitsverteilungen für jeden Status.

Stil-Parameter definieren, welches HMM verwendet werden soll.

- Finite State Machine:
 - Festgelegte Zustände
 - Fest vorgeschriebene Übergänge gemäß Input-Werten

- Finite State Machine mit Übergangswahrscheinlichkeiten:
Feste Zustände
Übergänge sind nach Wahrscheinlichkeiten geregelt.
- HMM:
Fixe Zustände, die eine Wahrscheinlichkeitsverteilung aufweisen.
Übergänge, die nach Wahrscheinlichkeiten geregelt sind.

Umwandlung von Bewegungsdaten in HMMs

Abbildung von signifikanten Haltungen/Posen (= signifikante Zeitpunkte) in der Choreographie auf die Zustände der HMM.

Festlegung der Übergangswahrscheinlichkeiten zwischen den Zuständen.

Festlegung der Wahrscheinlichkeitsverteilungen pro Zustand.

Diese kann verschiedenste Formen haben.

Umwandlung von Choreographien in HMMs

- Unsupervised Learning
- Trennung von Stil und Struktur
 - Gehen, Laufen, Schlendern, Marschieren, Hinken,... weisen selbe Struktur auf:
Fortbewegung auf 2 Beinen
 - Der Stil ist aber zwischen allen diesen Fortbewegungen signifikant anders.

Probleme:

Overfitting: Das gelernte HMM ist zu sehr an die gelernten Daten angepasst und kann keine neuen Stile erstellen.

HMM zu simpel: Das gelernte HMM ist zu einfach, um den komplexen Bewegungsablauf genau genug zu erfassen.

Training einer Style Machine

Motion Capture Daten einer Bewegungsfamilie werden verwendet um ein generisches Modell zu lernen.

Beispiele für Bewegungsfamilien:

Verschiedene Stile sich auf 2 Beinen fortzubewegen (laufen, spazieren, hinken,...).

Bewegungsablauf beim Hochsprung gemessen von verschiedenen Personen.

- Bewegungsablauf von 3 verschiedenen Datensätzen vom n-dimensionalen Raum auf 2D abgebildet.
- Jeder Punkt entspricht den Werten in einem Zeitschritt
- Ermittelte Zustände durch Training → Overfitting
- Das ermittelte HMM ist zu kompliziert.
- Optimale Struktur für ein HMM für Fortbewegung auf 2 Beinen
- Durch Festlegung des Parametervektors v wird das spezifische HMM und somit der Gehstil festgelegt.

Style Machine Resultate

Parametervektor v wurde festgelegt

Beispiele: Gehen und Laufen

Parametervektor v verändert dabei die Wahrscheinlichkeitsverteilungen des HMMs.

5. Einheit: Sportanwendungen

Schwimmen

Aufgabenstellung:

Videogestützte Beobachtung von Schwimmern im Training.

Möglichst automatische Auswertung und Evaluierung von Trainingsdaten.

Beispiele für Fragestellungen:

Welche Strecke hat ein bestimmter Schwimmer während einer Trainingseinheit zurückgelegt?

Zeitmessung bei den einzelnen Längen

Ziel: Computergestützte Optimierung des Trainings

Aufnahmekonfiguration:

Aufnahme des Schwimmbereichs durch 3 Kameras am Beckenrand

Bildmaterial:

Blick auf das Schwimmbecken aus der Sicht der 3 Kameras

- ⇒ Linsenverzerrungen sind deutlich erkennbar
- ⇒ Starke Spiegelungen im Wasser stellen besondere Herausforderungen an die Bildverarbeitungsalgorithmen
- ⇒ Weiters sind Unterschiede in der Farbkalibrierung der einzelnen Kameras sichtbar

Vorverarbeitung:

Entzerrung mittels perspektivischer Transformation

Im entzerrten Bild sind auch Bildbereiche enthalten, die von der zugehörigen Kamera nicht erfasst wurden.

Bildverarbeitung:

Erkennung eines Schwimmers

Verfolgung mittels Tracking

„Tracking-Spuren“:

- Horizontale Achse: Frames der Videosequenz
 - Vertikale Achse: Position des Schwimmers
- ⇒ Resultate werden verfeinert

Herausforderungen für die Bildauswertung:

Rastender Schwimmer, Schwimmer unter Wasser, Reflexionen, Fehlidentifikation

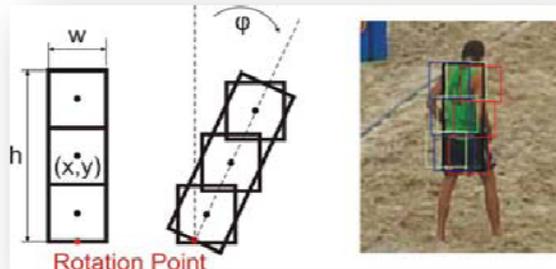
Volleyball

Aufgabenstellung:

Videogestütztes Tracking von Volleyballspielern.
Verwendung einer einzigen Kamera.

Herausforderungen:

Überlappungen/Verdeckungen von Spielern
Spezielle Bewegungselemente: Sprung, Bagger,...
Bei Beachvolleyball: Farbähnlichkeit zwischen Haut und Hintergrund
Modellierung und Erfassung der Rotation



Körperdarstellung mittels Quadraten die verschoben übereinandergesetzt werden können. Damit man auch geneigte/gebückte Personen oder Personen die horizontal dargestellt sind (am Boden liegend) erfassen kann.

Tennis

Visualisierungsbeispiele / Ziele:

Aufbereitung für den Zuseher
Spieldauswertung für taktische Analysen (Training)

Fußball

Systembeschreibung:

Die Bildaufnahme erfolgt durch mehrere statische Videokameras mit überlappenden Blickfeldern.
Segmentierung der Spieler und Tracking.
Kombination der Mehrfachbeobachtungen zur robusteren Bestimmung der Positionen und Geschwindigkeiten der Spieler und um Mehrdeutigkeiten bei Verdeckungen aufzulösen.

Spielersegmentierung:

Verwendung eines zusätzlich aufgenommenen (leeren) Hintergrundbildes
Unterdrückung der Schatten am Rasen

Visualisierungen:

Trajektorien der Spieler
Tracking in Mehrfachansichten