Key-Frame Extraction for 3D Human Motion Sequence Segmentation

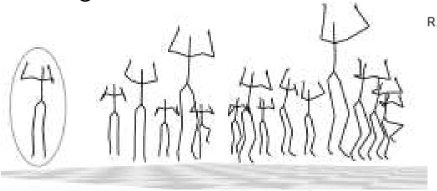
Michal Balážia

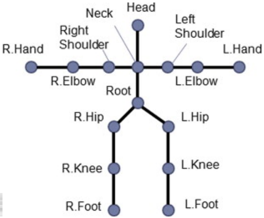
Laboratory of Data Intensive Systems and Applications
Faculty of Informatics
Masaryk University

October 7, 2013

3D Human Motion Capture

- Employment
  - motion simulation and exposition - avatar graphics
  - content-based retrieval in motion database
  - biomechanical analysis - gait disorders detection, rehabilitation
  - generating new motion instances

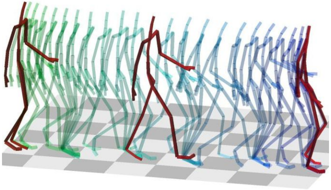




- 3D data capturable via Microsoft Kinect or on-body sensors
- Human motion in the form of a sequence of body poses
- Pose characterised by 3D coordinates of selected body points and time
- Various extractable features
  - joint angles, their velocity or acceleration
  - body points' distances
  - relational features

Key-Frames

- Key-frames are time frames of motion sequence extracted according to specific sampling strategy.

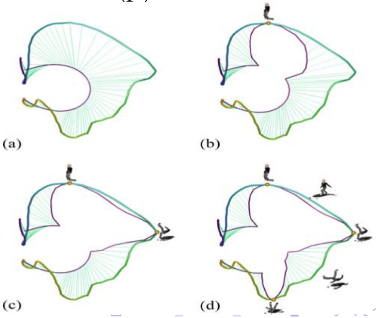


- Purposes
  - compression of motion data
  - motion retrieval
  - motion sequence segmentation
- Two base key-frame sampling strategies
  - uniform - each two consecutive key-frames of equal time difference
  - adaptive - respect to motion sequence (extremal poses, turnover, etc.)
- Approaches to key-frame extraction - Assa, Műller, Gong, Xiao, Liu

Assa

Assa - curve averaging

- Pose consists of skeletal joints and their associated aspects: (1) positions, (2) angles, (3) velocity, (4) angular velocity
- x_a^f - value of aspect a in frame f
- High-dimensional curve x_a^f ($4\times\#$joints) is reduced by RMDS algorithm to a curve $C(f)$ of 5-8 dimensions
- Point p in $C(p)$ is projected onto average curve $\overline{C}(p)$
- $r_p = |C(p) - \overline{C}(p)|$ - distance at point p
- Iterative key-frame extractor algorithm:
  1. add p_i of maximum r_{p_i} to key-frames
  2. modify $\overline{C}(p)$ to touch $C(p)$



#Műller

Műller - genetic learning

- $F = (F_1, \ldots, F_f)$ - set of f relational features
- F-segment of data stream D represented by matrix $M_F[D]$
- Example: $F = (F_{LeftKneeBent}, F_{RightKneeBent})$, $K = 5$ F-segments

$$M_F[D] = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 \end{pmatrix}$$

- error tolerant $(0/1 \rightarrow *)$ motion class patterns $X \in \{0, 1, *\}^{f \times K}$
- $V_k \subseteq \{0, 1\}^f$ - subset of alternative feature vectors = fuzzy set
- $T^{+/-}$ - set of positive/negative training F-motions
- Individual described by element of T^+ and submatrix of $M_F[D]$
- Mutations: change element of T^+, change row or column of $M_F[D]$
- Optimization in terms of recall and performance of fuzzy query $V(X) = (V_1, \ldots, V_K)$

Gong, Xiao, Liu

Gong - local-motion energy extremes

- θ_l^a - angle between limb l and axis a
- $\psi = [cos(\theta_1^x), cos(\theta_1^y), cos(\theta_1^z), \ldots, cos(\theta_{12}^x), cos(\theta_{12}^y), cos(\theta_{12}^z)]$ - pose
- $E_i = |\psi_i - \psi_{i-1}|^2$ - energy in i-th frame of pose ψ_i
- Key-frames are frames i of extremal E_i

Xiao - angle extremes

- $\theta_i^{(l)}$ - angle between limb bone l and central bone in frame i
- Key-frames are frames i such that $\exists l \in \{1, \ldots, 8\} : \theta_i^{(l)}$ is extremal

Liu - cluster centroids

- r_{lxi} - rotational parameter of lhip/rhip/chest in axis x/y/z in frame i
- $F_i = [r_{lxi}, r_{lyi}, r_{lzi}, r_{rxi}, r_{ryi}, r_{rzi}, r_{cxi}, r_{cyi}, r_{czi}]$ - frame
- σ_i - i-th cluster of frames clustered by weighted Euclidean distance
- Key-frames are frames i closest to centroid of σ_i

The Goal

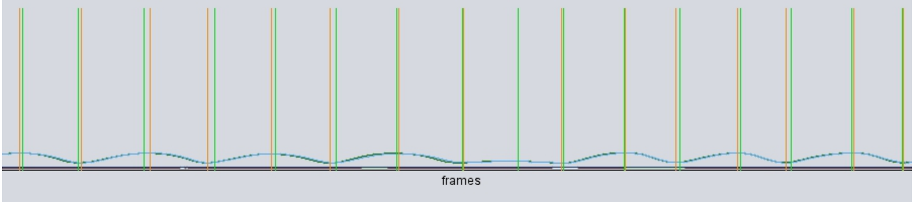
Key-Frame Extraction for 3D Human Motion Sequence Segmentation

- Soon-to-be paper
- Database: CMU Motion Capture Database
- Extracted features: distance signals of selected pairs of body points, angle signals of selected joints
- Key-frame sampling strategy: adaptive, local extremes with additional criteria (neighbourhood, weight)
- Evaluation against ground truth: recall, precision, F-measure, ratio between ground truth and algorithmically selected key-frames
- Evaluation against current approaches: evaluation against identical ground truth in all aspects, evaluation of flexibility and performance

Current State and Future Vision

We have:

- extracted distance and angle signals as motion features
- defined ground truth over 5 motion sequences
- implemented an algorithm that extracts key-frames according to given sampling strategy
- recall and precision of 50-70% (hit within 20-frames neighbourhood)



We continue with:

- evaluation against current approaches
- hierarchical motion segmenter

Questions?

Thank you for attention.

J. Assa, Y. Caspi, and D. Cohen-Or, Action Synopsis: Pose Selection and Illustration, ACM SIGGRAPH 2005.

M. Műller, B. Demuth, and B. Rosenhahn, An Evolutionary Approach for Learning Motion Class Patterns, DAGM 2008.

W. Gong, A.D. Bagdanov, F.X. Roca, J. Gonzàlez, Automatic Key Pose Selection for 3D Human Action Recognition, AMDO 2010.

J. Xiao, Y. Zhuang, T. Yang, and F. Wu, An Efficient Keyframe Extraction from Motion Capture Data, CGI 2006.

F. Liu, Y. Zhuang, F. Wu, and Y. Pan, 3D Motion Retrieval With Motion Index Tree, CVIU 2003.