Semantically Consistent Human Motion Segmentation

Michal Balážia
Jan Sedmidubský
Pavel Zezula

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

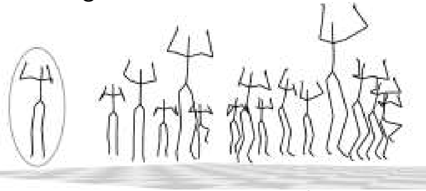
March 24, 2014

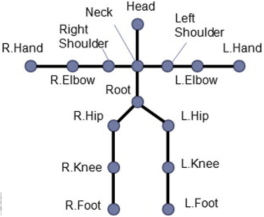
Presentation Outline

1. 3D Human Motion Capture and Representation
2. Key poses
3. Existing approaches to key-pose detection
4. Our purpose of motion segmentation
5. How our method works
6. Results

3D Human Motion Capture and Representation

- Employment
  - Motion simulation and exposition - avatar graphics
  - Content-based retrieval in motion databases
  - Biomechanical analysis - detection of gait disorders, rehabilitation
  - Generating new motion instances

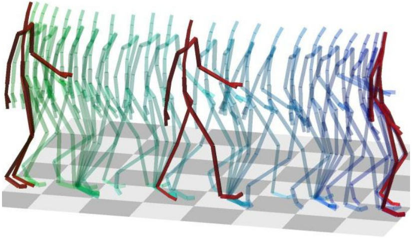




- 3D data capturable via Microsoft Kinect or on-body sensors
- Human motion in the form of a sequence of body poses
- Pose characterised by 3D coordinates of selected body points and time
- Various extractable features
  - Joint angles, angular velocity, acceleration
  - Joint relative distances
  -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

Key poses

- Not all poses are important for us
- We detect key poses - the poses of interest
- Key pose selection according to the purpose

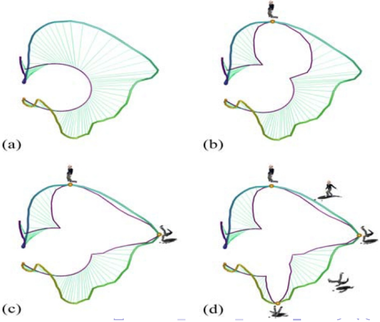


- Purposes
  - Summarization and compression of motion data
  - Motion retrieval
  - Motion sequence segmentation
- Relevant approaches to key-pose detection - Assa, Gong, Xiao

Assa

Assa - based on curve averaging

- Pose described skeletal joints and their associated aspects: (1) positions, (2) angles, (3) velocity, (4) angular velocity
- f_t^a - value of aspect a at time t
- High-dimensional curve f_t^a $(4\times\#\text{joints})$ is reduced by RMDS algorithm to a curve $C(t)$ of 5-8 dimensions
- Point t in $C(t)$ is projected onto a predefined average curve $\overline{C}(t)$
- $r_t = |C(t) - \overline{C}(t)|$ - distance at time t
- Iterative key-pose detection algorithm:
  1. Find pose index t of maximum r_t
  2. Mark the t-th pose as key pose
  3. Modify $\overline{C}(t)$ to touch $C(t)$



Gong

Gong - based on local-motion energy extremes

- Human model consists of 12 limb bones
- Motion features in form of limb bone axis rotations
- $\theta_t^{l,a}$ - angle between limb bone l and axis a at time t
- 36-dimensional pose P_t:

$$[cos(\theta_t^{1,x}), cos(\theta_t^{1,y}), cos(\theta_t^{1,z}), \ldots, cos(\theta_t^{12,x}), cos(\theta_t^{12,y}), cos(\theta_t^{12,z})]$$

- Local-motion energy in pose at time t

$$E_t = |P_t - P_{t-1}|^2$$

- Key poses are the poses P_t of locally extremal E_t

Xiao

Xiao - based on curve simplification

- Human model consists of 8 limb bones
- Motion features in form of limb bone angles
- θ_t^l - angle between limb bone l and central bone at time t
- 8-dimensional pose P_t:

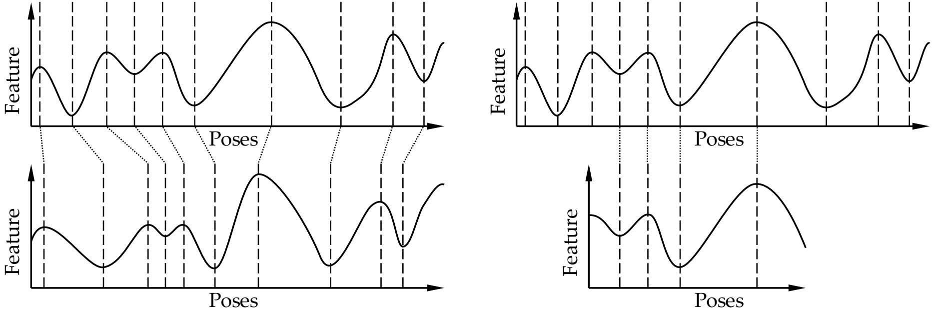
$$[\theta_t^1, \ldots, \theta_t^8]$$

- Key-pose candidates are the poses where any θ_t^l is locally extremal
- Curve is constructed to enclose the key-pose candidates
- LCS (similar to SCS) is applied to this curve to refine key poses

Our Purpose of Motion Segmentation

Our purpose: motion retrieval

- Pose-level indexing - inefficient
- Segment-level indexing
  - Both motion database and query are segmented
  - Segments can be separately analysed, indexed, clustered, or compared
  - Query is searched in the database as a sequence of motion segments
  - Semantic consistency - key poses at semantically equivalent phases



(a) (b)

How Our Method Works

Our - based on changes in body point trajectories

- Database: HDM05 Motion Capture Database - University of Bonn
- Two major steps:
  1. Feature extraction
  2. Pose energy curve construction and analysis

1. Feature extraction

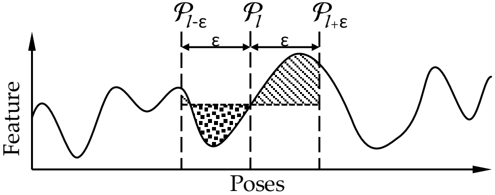
- Relative distances of selected body point pairs:
  1. left hand and right hand
  2. left hand and left foot
  3. right hand and right foot
  4. left knee and right knee
  5. left foot and right foot
- Motion - sequence of poses $\mathcal{M} = (\mathcal{P}_1, \ldots, \mathcal{P}_n)$
- Pose - configuration of features in time $\mathcal{P}_t = (f_t^1, \ldots, f_t^5)$
- f_t^p - relative distance of p-th body point pair at time t

How Our Method Works

2. Pose energy curve construction and analysis

- Energy is determined by pose neighbourhood of radius ϵ
- Feature energy of each feature f_t^p at pose P_t is calculated as

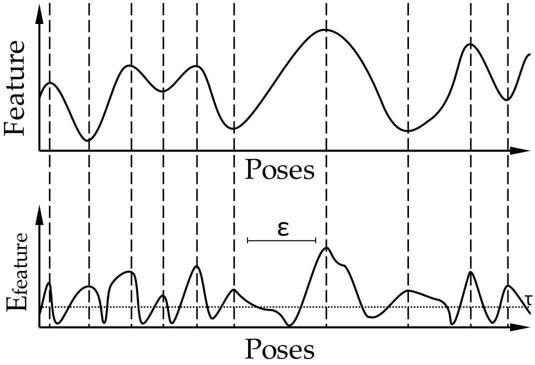
$$E_{feature}(f_t^p) = \sum_{t'=t-\epsilon}^{t+\epsilon} \left(f_t^p - f_{t'}^p\right)$$



- Pose energy of each pose P_t is calculated as

$$E_{pose}(P_t) = \sum_{p=1}^{5} |E_{feature}(f_t^p)|$$

- Pose-energy curve is constructed by concatenating the pose-energy values through the processed motion in chronological order



(c)

- Local maxima greater than a τ threshold of this curve are identified as key poses

Results

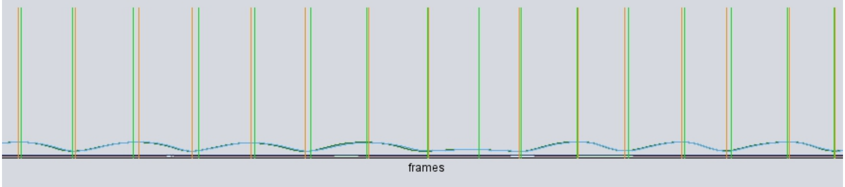
We have:

- Extracted body point distances as motion features
- Manually defined ground truth over 11 motion sequences
- Implemented and tested all methods - our, Assa, Gong, Xiao

Evaluation results:

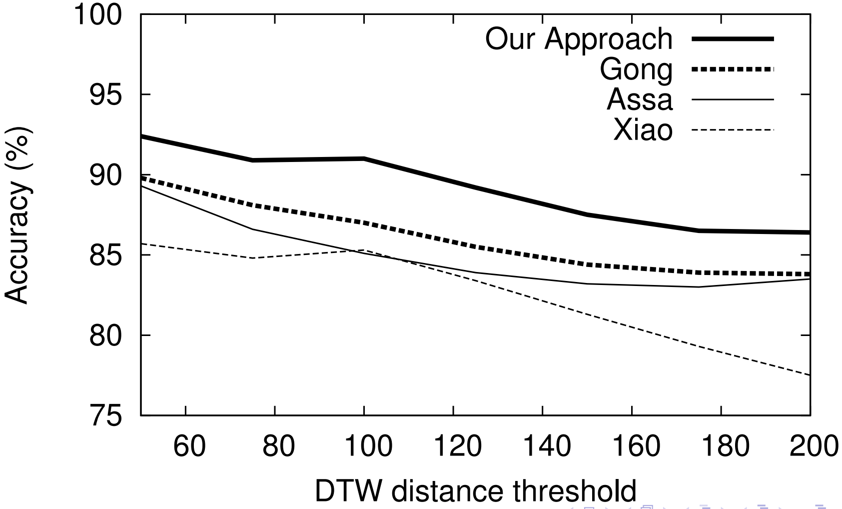
- Evaluation against ground truth: hit within 10-pose neighbourhood

	Recall (%)	Precision (%)	F-measure (%)	Ratio (%)
Our	52.2	46.5	49.2	112.1
Assa	35.2	39.2	37.1	89.7
Gong	92.1	4.6	8.7	2040.8
Xiao	30.7	45.3	36.6	67.7



Results

- Evaluation of semantic consistency of all methods



Questions?

Thank you for attention.

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

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.