Single Camera Gesture Analysis

The holy grail of motion capture

Josep R. Casas

Fraunhofer IAIS, Sankt Augustin

10.01.2014

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Departament de Teoria del Senyal
i Comunicacions

Image Processing Group (GPI)

Summary

1. Introduction
   - HMA: HBM and features
2. Pose inference and tracking
3. Action and gesture recognition
4. Dynamics
   - Action recognition with Nonlinear Dynamical Systems (López, 2012)
5. Conclusion

Human motion analysis (HMA)

- HMA framework
  - Motion tracking and estimation
  - Action and behavior recognition
  - Segmentation of human motion
- Challenges
  - Viewing perspectives (2D!)
  - Scene clutter
  - Imprecise semantics of actions and human motion
- Computer vision approaches
  - HBM/markerless motion capture
  - Low level features

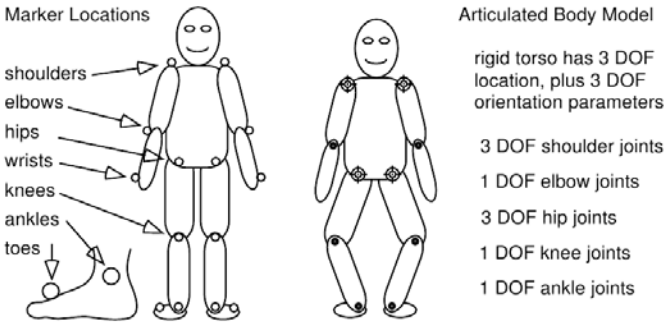
Markerless motion capture

- Articulated Human Body Models (HBM)
  - Use real-world magnitudes: e.g. 3D positions, angles…

  → challenging to estimate from 2D images

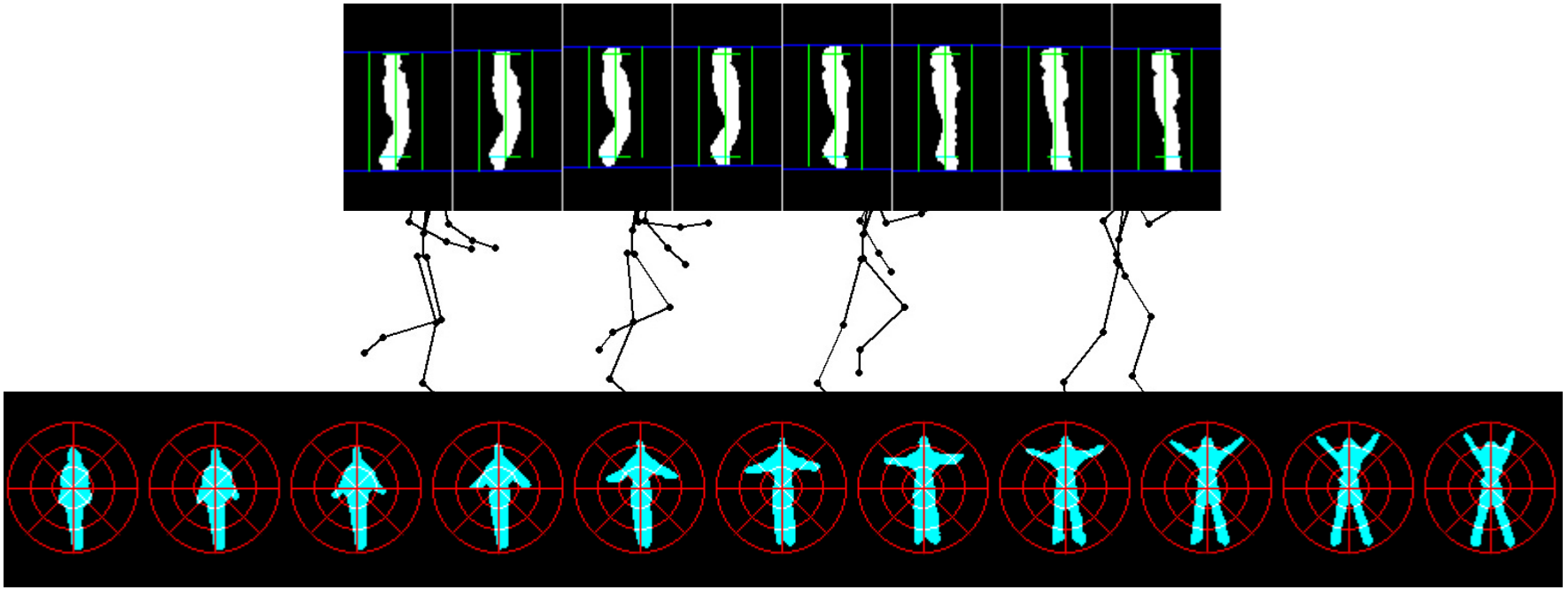
  → may offer richer models of dynamics and behavior

Just by looking at a set of dots in motion, we can identify what's going on (Johanson 1973)



(Campbell and Bobick 1995)

HBM → Low level features



Low-level features for HMA

(Siskind and Morris 1996)

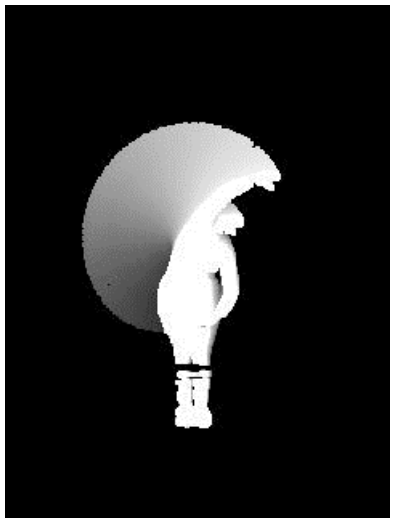
Human event perception does not presuppose object recognition. Visual event recognition performed by a visual pathway separated from object recognition. Experiments on simple gestures (pick up, put down, push, pull, drop and throw) demonstrated that visual events can be classified based only on motion profiles

- Features
  - Interest points, skin patches, silhouettes, filter banks, moments, trajectories, motion…
  - → surprisingly effective in very specific/constrained tasks
  - → difficult on generic scenarios (conditioned by appearance and viewpoint)

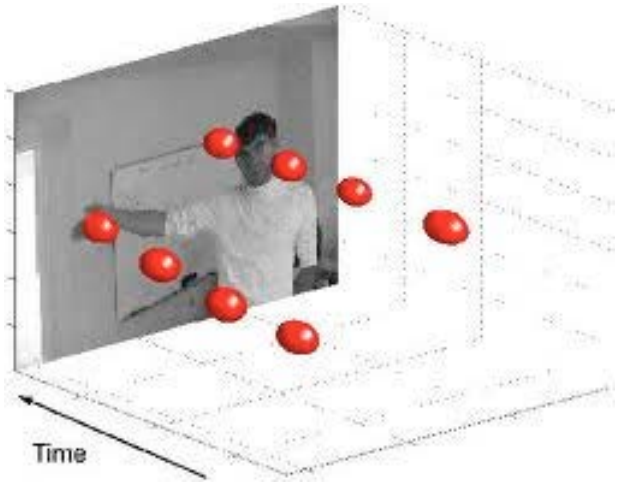
'Holy grail'

Capturing motion from a single camera for HMA in generic scenarios…

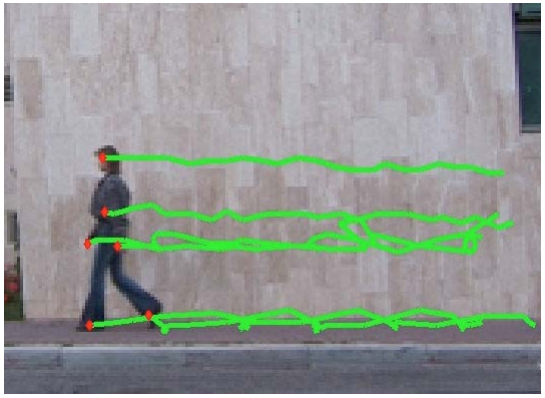
Low-level features for HMA



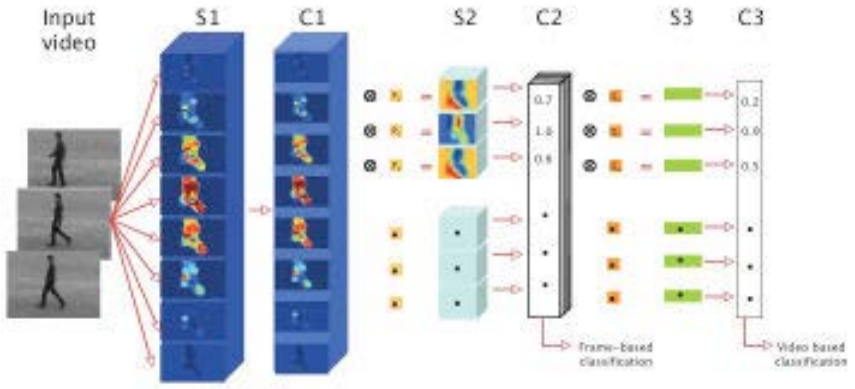
Motion templates



Interest points



Trajectories



Filter banks

Pose inference

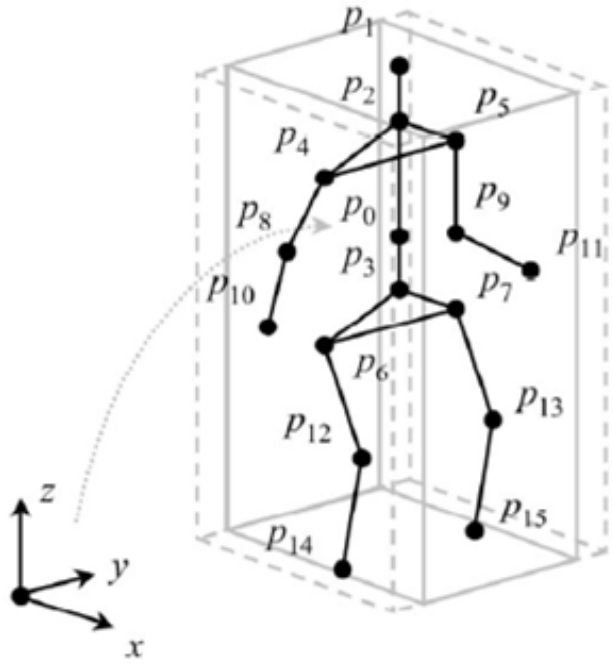
- Ill-posed problem
  - spatial ambiguities, mainly caused by self-occlusions
  - variability in individuals‘ appearance
  - background clutter
- Approaches
  - Exploiting a HBM (analysis-by-synthesis)
  - Without a HBM: pose estimation, motion classification

Pose inference with HBM

- Human Body Models
  Dimensionality reduction
  - Example-based

    direct mapping between feature space and 3D pose space
  - Model-based

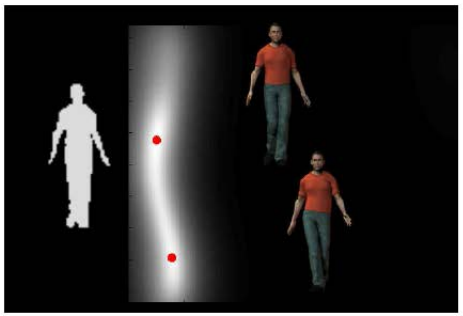
    use kinematic constraints imposed by an articulated HBM
    - Inverse kinematics
    - ML approaches
      NN/motion graphs (Kovar et al 2002)
      LVMs (Grochor et al 2004)
      Dynamics (Urtasun 2008)
    - Space-time constraints: physics (arm connected to shoulder, elbow angle [-π,0])
    - Body parts
      (Singh 2010, Bergtholdt 2010)

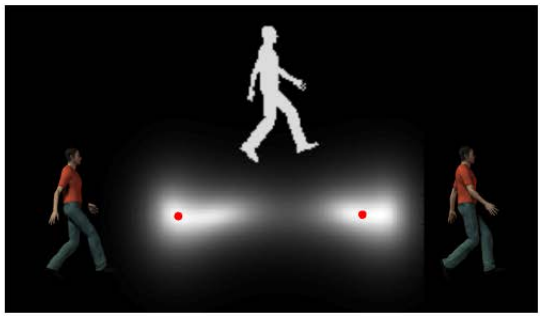


Pose inference without HBM

- Pose estimation from images mapping observations to pose

Problem cannot be solved directly as a regression task, since the mapping is multimodal (an image observation can represent more than one pose)





- Inverse kinematics
- Discriminative models: NN, regression, structured-output
- Generative models: Kalmann, particle filters…
- Likelihood models
- Priors: shape models, motion models, joint limits, physics

Human Motion Classification

- Gesture recognition
  - Traditional gestures

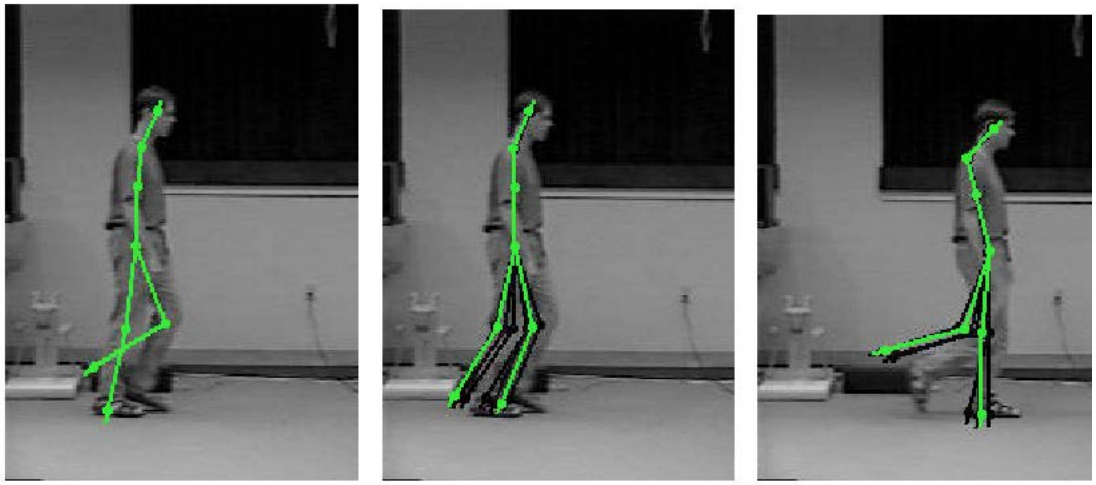
  - Gestures for communication (sign language)

- Activity recognition
  - Simple activities (walk, run, jog…)
  - Sports movements and actions
- Gait analysis

  …for surveillance (ID), orthopedics, style

Dynamics

- Dynamical systems to model human motion



(Pavlovic and Rehg 2000)

- Action recognition with Nonlinear Dynamical Systems (López, 2012)

Nonlinear Dynamical Systems

- Related work

  Human Motion as Time Series

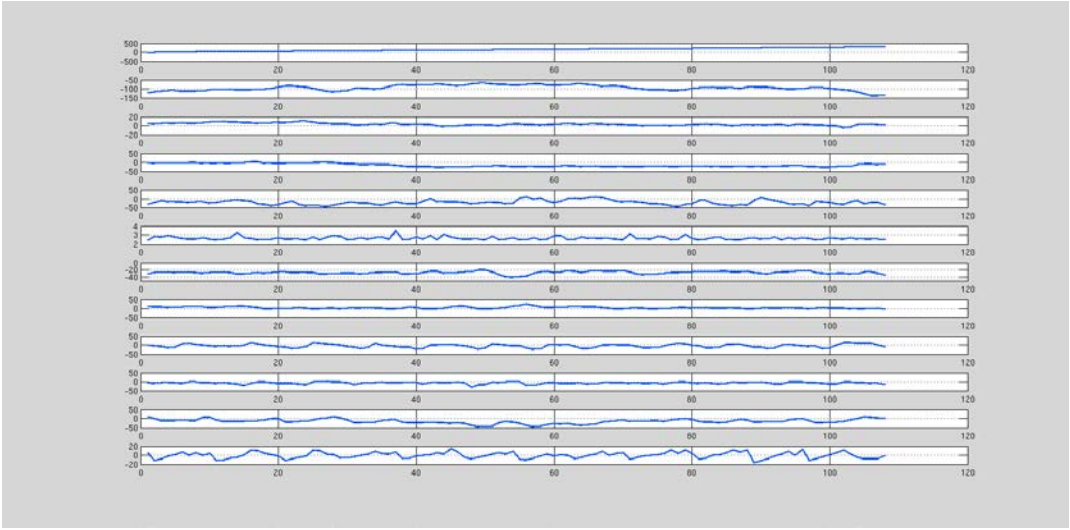
  - [Lv et al. 2005] model actions weighted channels consisting in temporal evolution of 3D joint coordinates
  - [Raptis et al. 2010] propose spike-train driven dynamical models for human motion

  Theory of Nonlinear Dynamical Systems

  - [Ali et al. 2007] Chaotic Invariants for Human Action Recognition
  - [Basharat and Shah, 2009] Time-delay embeddings for motion synthesis

Human Motion as Time Series

- Time-delay embeddings
- Trajectory matching in phase spaces (dynamic programming)
- Random Forests in phase space (for delay vectors)
  - Discriminative partitions
  - Efficient handling of high-dimensional spaces



Results

- Future light dataset (5 actions, 158 sequences, strong stylistic variations)

Method	Accuracy
Spike Train Driven Dynamical Models [Raptis 2010]	86.07%
Chaotic Invariants [Ali 2007]	90.20%
Flexible dictionaries + BoW [Raptis 2008]	97.40%
Flexible dictionaries + String Matching [Raptis 2008]	98.03%
Trajectory Matching ($m = 5$, SVM Approach)	88.76%
Trajectory Matching ($m = 5$, Weighted Approach)	91.09%
Random Forest in Phase Space ($m = 5$)	96.18%

A. López-Méndez and J. Casas, *Model-Based Recognition of Human Actions by Trajectory Matching in Phase Spaces*, Image and Vision Computing, 2012
López Méndez, *Articulated Models for Human Motion Analysis*,
PhD Universitat Politècnica de Catalunya (UPC), December 2012

Conclusion

Potential strategies for AUVIS

- Novel frameworks for action recognition based on time-delay embeddings can be applied to feature-based gesture tracking/recognition
  - Characterization of the geometrical and topological structure of the embedded phase-space by means of trajectory matching
  - Random forest approach deals with reconstructed phase spaces of multivariate time series (López 2011), outperforming state-of-the art approaches using time-delay embeddings
- (alternative) Using Poses for temporal clustering of human motion
  - Weakly supervised temporal clustering
  - Information Theoretic Metric learning from poses

A. López-Méndez, J. Gall, J. Casas, and L. van Gool, "Metric Learning from Poses for Temporal Clustering of Human Motion", BMVC 2012