Evaluating Lookup-Based Monocular Human Pose Tracking on the HumanEva Test Data

Nicholas R. Howe
Orientation

• Goal of work: Develop HumanEva results
  – Algorithms not necessarily state of the art
  – Useful as baseline

3D pose recovery

Single camera
Lookup-Based Motion Capture (1)

• Use silhouettes to retrieve known poses
Lookup-Based Motion Capture (2)

- Use second-order hidden Markov model to select pose sequence with low energy

Optimize for:
- Agreement with frame observations
- Agreement with flow observations
- Small inertial changes between frames
Flow Chart of Motion Capture

Input video

- Frame \(i-1\)
- Frame \(i\)
- Frame \(i+1\)

\[\ldots\]

Optical flow & foreground segmentation

- Frame \(i-1\)
- Frame \(i\)
- Frame \(i+1\)

\[\ldots\]

Retrieval

- Candidates \(i-1\)
- Candidates \(i\)
- Candidates \(i+1\)

\[\ldots\]

Markov Chain Optimization

- Final Pose Trajectory

Motion capture data

Pose & flow rendering

Pose Library
Some Related Work

• Estimating Human Body Configuration Using Shape Context Matching
  Mori & Malik, ECCV 2002
• 3D Tracking = Classification+Interpolation
  Tomasi, Petrov, & Sastry, ICCV 2003
• Silhouette Lookup for Automatic Pose Tracking
  Howe, ANM 2004

---

• 3D Articulated Models and Multi-View Tracking with Silhouettes
  Delamarre & Faugeras, ICCV 1999
• Temporal Integration of Multiple Silhouette-based Body-part Hypotheses
  Kwatra, Bobick, & Johnson, CVPR 2001
Caveats

• Scalability of pose library is a concern
  – May limit technique to specific applications
  – Walking library: 805 poses
  – Boxing library = several thousand poses
  – Some work exists on sublinear retrieval

• Lookup employs background subtraction
  – Good segmentation result is often achievable
  – BS not required in principle for lookup-based methods
    • Others have demonstrated edge-based techniques
Overview

• Feature from video
  – Background subtraction
  – Optical flow
• Lookup techniques
• Markov chaining
• Results
Background Subtraction

- Graph cut formulation uses edge data
  - Segmentation tends to follow edges
- HSV color space with shadow correction
- Robust estimation of background
- Failures mostly due to poor contrast

Typical good result

Typical bad result
Optical Flow

- Optical flow from Krause method
- Mask by foreground & compute moments
Pose Lookup

- Candidate pool combines multiple queries:
  - Half-chamfer distance (A)
  - Turning angle distance (B)
  - Flow moment distance (C)

Most results are constrained to lie near the previous frame’s candidates.

A few open retrievals are also included.

Combinations:
- A+B+C
- A+B
- C only
Frame “Stitching”

• First-order Markov chain sufficient for “smoothness”
• Second-order chain is needed for conservation of momentum
• Flow match & momentum conservation intended to prevent “shuffle-step” errors
Results

• Results available for:
  S3_Walking_1_(BW2): Mean error 11 pixels
  S3_Walking_1_(C2): Mean error 14 pixels
  S2_Walking_1_(BW2): Mean error 13 pixels*
  S3_Walking_1_(C1): Mean error 18 pixels*
  *Affected by error in background subtraction

• Boxing available soon (hopefully!)

• Observations:
  – Left-right inversion problems
  – Error is highest at extremities
S3_Walking_1_(BW2)

- Mean error after swaps is 11 pixels.

* apparent error in mocap for head
Optimization

• Match may be improved by optimization on pose parameters

• One frame at a time
  – Improve chamfer match with silhouette
  – Improve smoothness: use quadratic fit to parameters over 11-frame window

• Error improves to 10 pixels after one round
S2_Walking_1_(BW2)

- Mean error after swaps is 13 pixels.
S3_Walking_1_(C2)

- Mean error after swaps is 14 pixels.

* apparent error in mocap for head
S3_Walking_1_(C1)

- Mean error after swaps is 18 pixels.
Conclusions

Few results, but some trends are clear:

• Pixel accuracies in the teens

• Limitations of silhouettes evident
  – Left-right ambiguity is still an unsolved problem for this method
  – Arm locations drift when obscured by torso

• One set of results for HumanEva
Coordination Between Frames

• Need to pick from top matches at each frame.
  – Want **good image match** at all frames
  – Want **small change** between frames
  ⇒ Markov chain minimization!

• Best local choices minimize global error
Markov Chain Minimization

1. Compute cheapest path to each state from previous states (cost = estimate of plausibility)

2. Identify best (least expensive) result

3. Backtrack, picking out path that gave best result.
Silhouette Comparison

Turning angle
(Captures morphology)

Chamfer distance
(Captures overlap)

• Combine using Belkin technique
  (score = sum of individual ranks)