Boundary Fragment Matching and Articulated Pose Under Occlusion

Nicholas R. Howe

Low-Budget Motion Capture

Constraints (informal/archived footage):

- Single camera
- No body markers

Consequent Challenges:

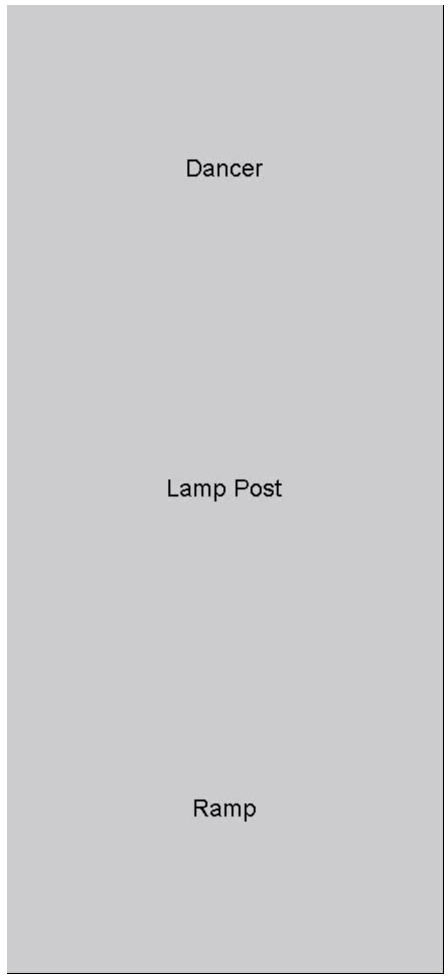
- 3D ambiguity
- **Occlusion (self & external)**

Static Backgrounds & Silhouettes

Favorable Circumstances:

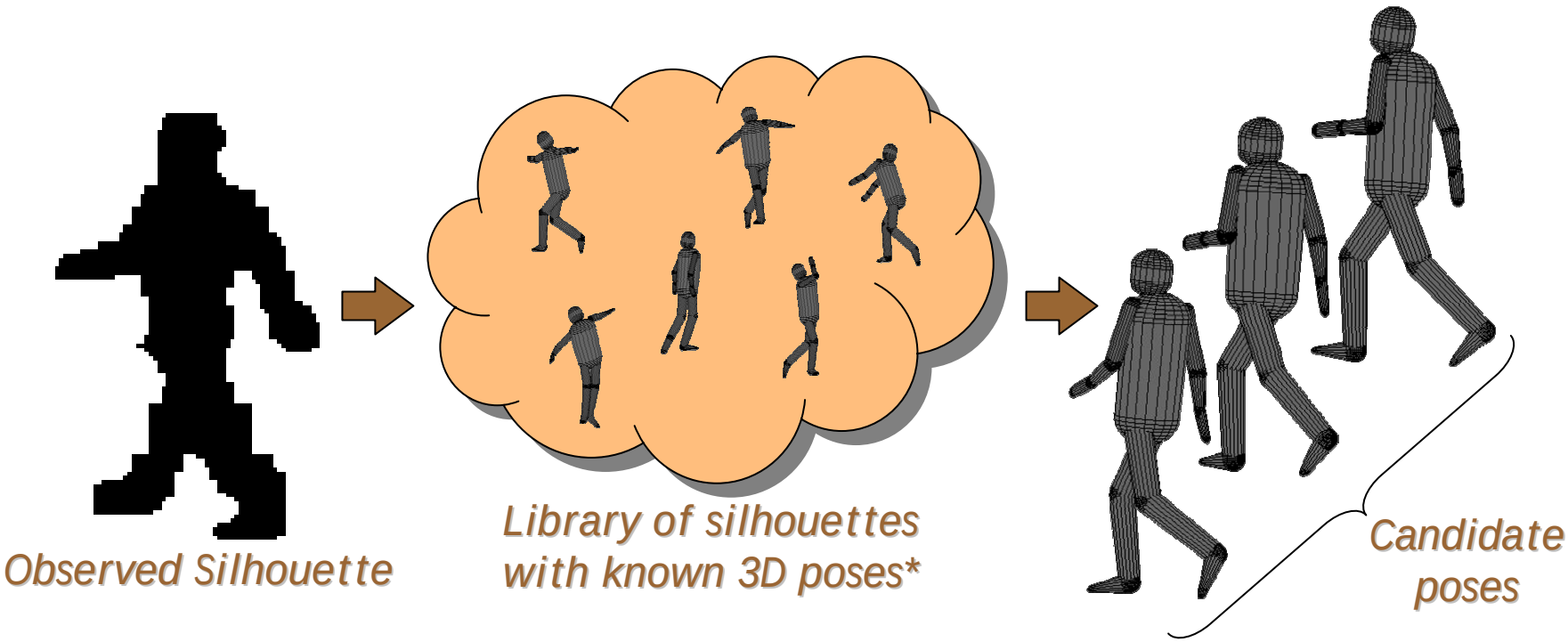
- Build statistical model of static background
- Outliers reveal subject silhouette (maybe!)

☞ *What if the silhouette is occluded?*



Motion Capture from Silhouette (1)

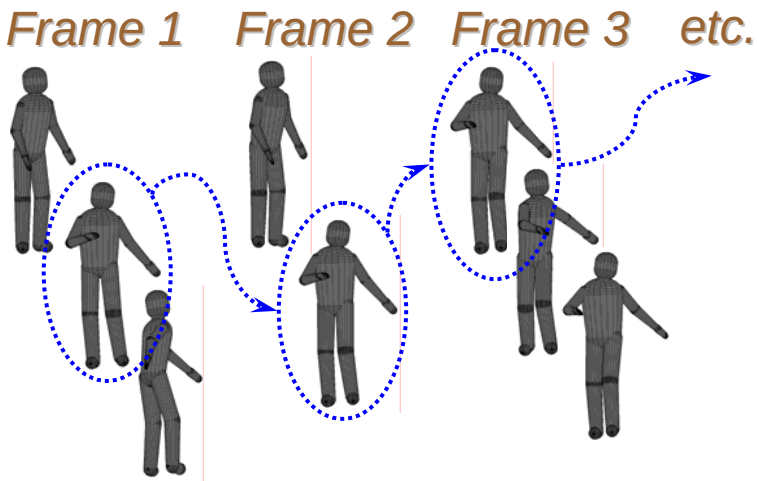
- Use silhouettes to retrieve known poses



*(Open question: how many library poses are enough for general motion?)

Motion Capture from Silhouette (2)

- Pose reconstruction = smoothest frame-to-frame sequence of candidate poses



- Temporal continuity resolves ambiguity
- Post-process: smooth & optimize

Some Related Work

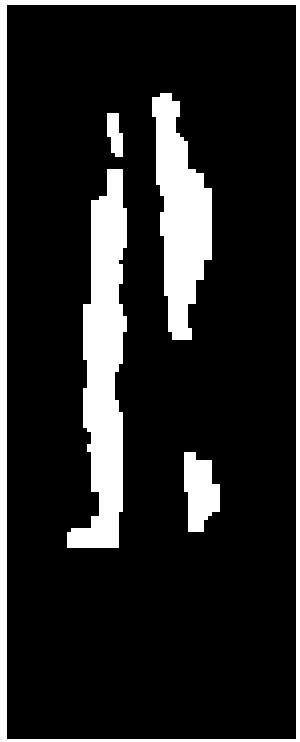
- Estimating Human Body Configuration Using Shape Context Matching
  Mori & Malik, ECCV 2002
- 3D Tracking = Classification+Interpolation
  Tomasi, Petrov, & Sastry, ICCV 2003
- Silhouette Lookup for Automatic Pose Tracking
  Howe, ANM 2004

- 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

What about OCCLUSION?

Occlusion Ruins Everything

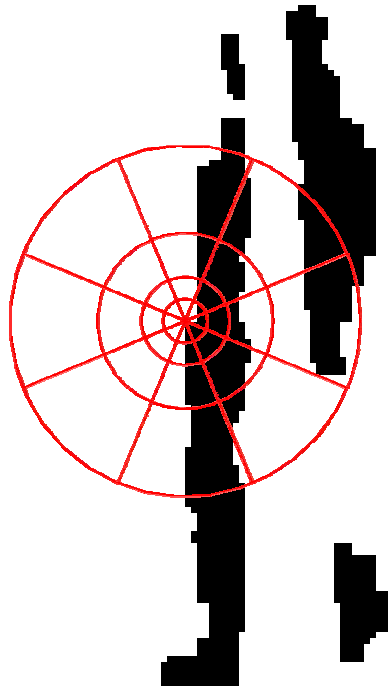
- Significant occlusion changes silhouettes
- Traditional methods cannot retrieve good candidates from library



☞ *Can we find a way to match partial silhouettes?*

Shape Context Is Not Local

- Shape context encodes appearance of a large area
  - Occluded areas can affect shape contexts over entire figure
  - Problematic for other shape methods also
- Occlusion creates false edges
  - Wrong shape is encoded
  - Spurious sample points increase confusion

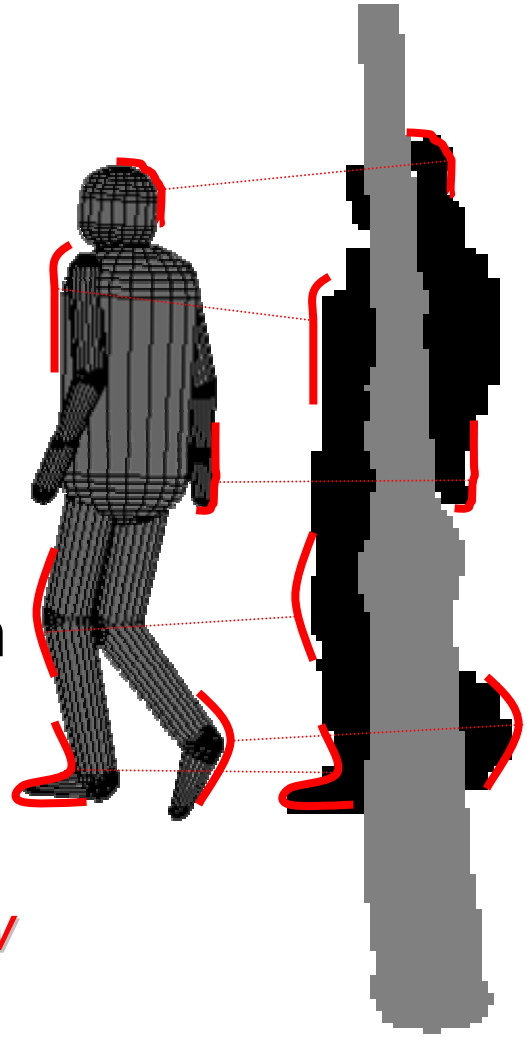


☞ *Must solve both problems to succeed*

Boundary Fragments

Boundary fragment = fixed-length section of silhouette boundary

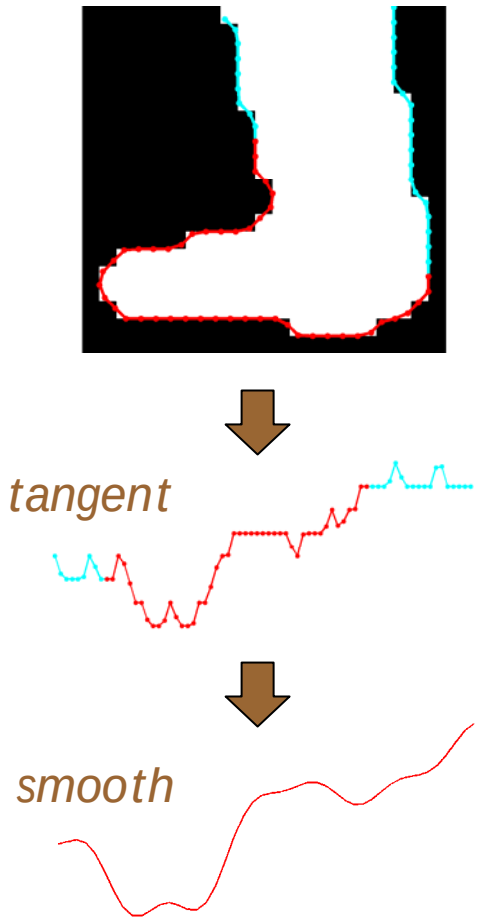
- **Distinctive**: Good match of boundary fragments means shapes are similar
- **Redundant**: Subset of fragments still matches to similar shapes
- **Local**: unaffected by remote occlusion
- **Efficient**: best fragment matching approximated via *EMD embedding* [5]



☞ *Assume for the moment that we know which parts of the boundary are "real"*

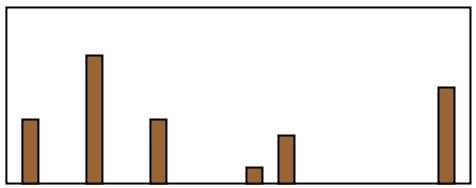
Handling Boundary Fragments

- Represent fragment by sampling tangent angle uniformly
- Normalize and smooth
- Embed in high-D space
- Shape ↔ sum embeddings of **all** boundary fragments of length L
- Occlusion: use visible fragments
  - Normalize based on fraction visible, or search for best normalization

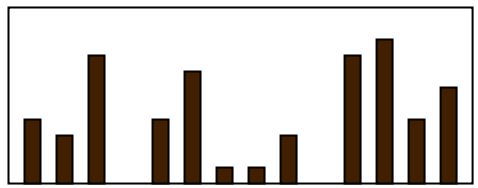


Partial Silhouette Embeddings

- Partial silhouettes will have some components missing compared with full
- Nonzero components should match closely



Embedding of partial silhouette

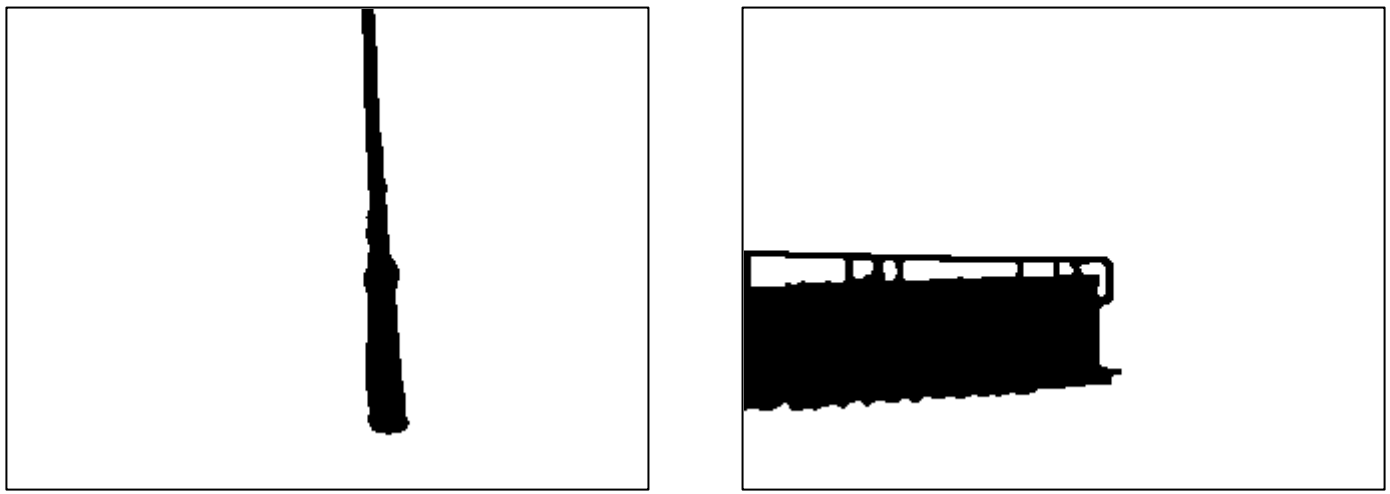


Embedding of full silhouette

☞ *Next: How to distinguish real silhouette edges from occlusion boundary?*

Occlusion Zones

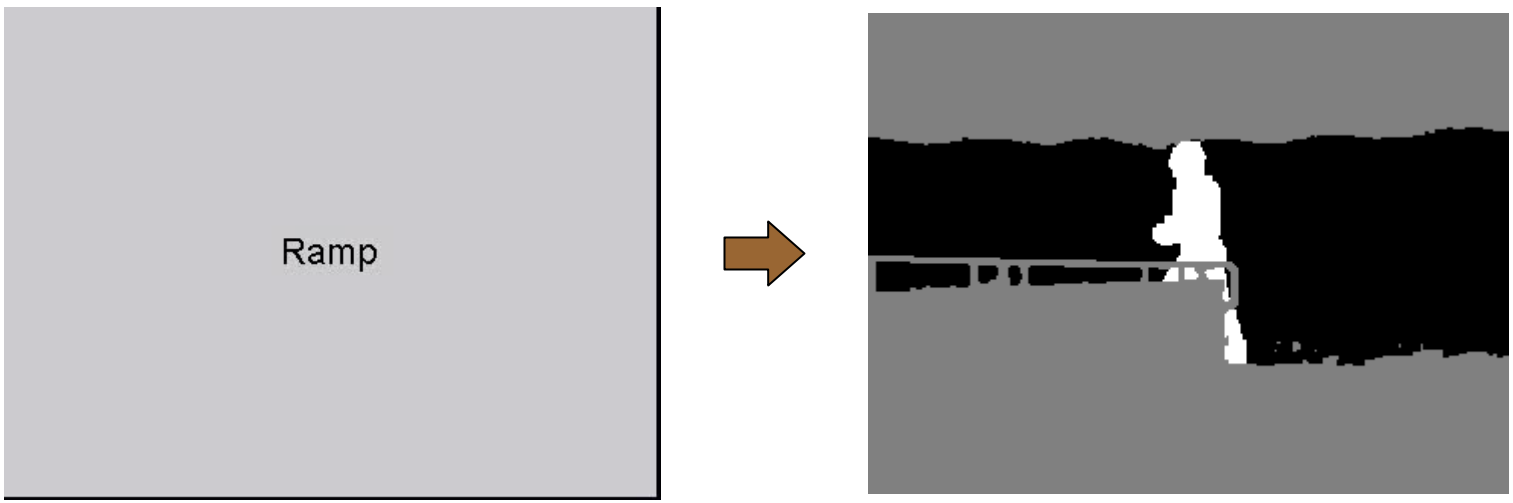
- Occlusion zones are areas where objects can occlude subject
- Silhouette edges near occlusion zone are probably spurious



Sample maps of occlusion zones created by hand

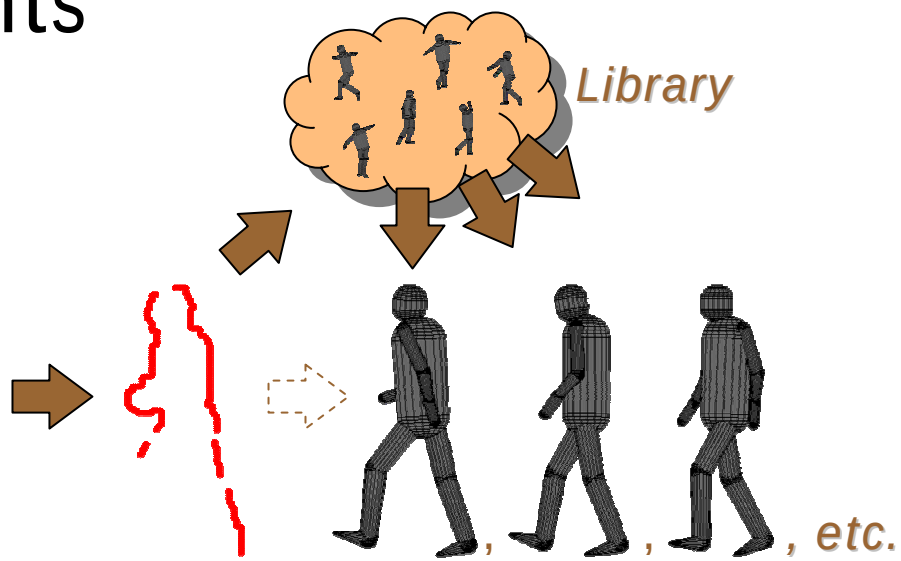
Finding Occlusion Zones

- Cumulative silhouettes avoid occlusion zones
- Some areas: no data available
- Better to use too large an occlusion zone than too little.



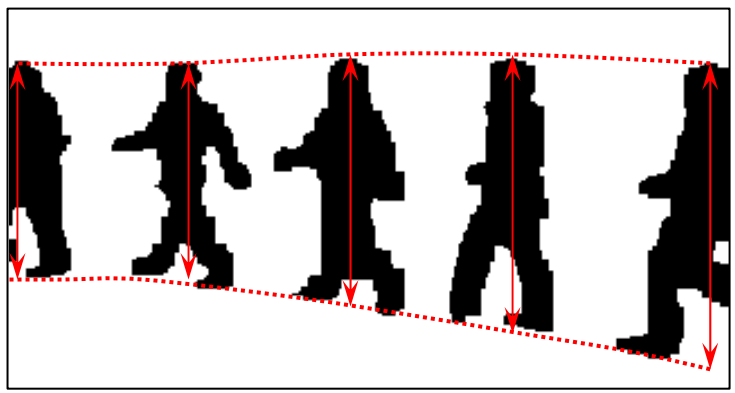
Pose Retrieval with Occlusion

- Use all boundary fragments not touching occlusion zones
- Form embedded vector using valid boundary fragments
- Match to library



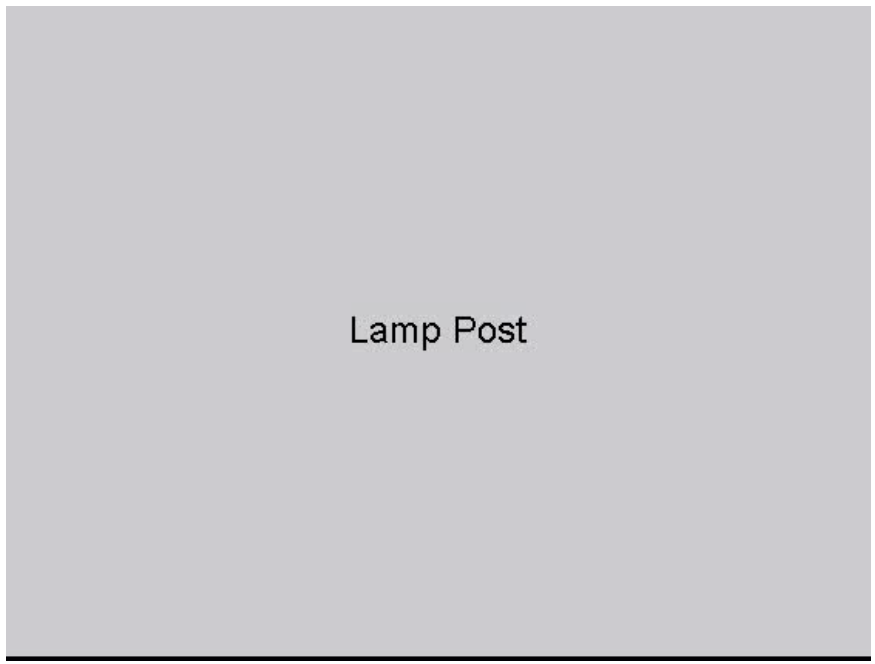
Scale Estimation

- Need figure scale to choose boundary fragment length
- Occlusion may obscure scale
- Assume: some frames are unoccluded
  - Interpolate/extrapolate to find scale of remainder



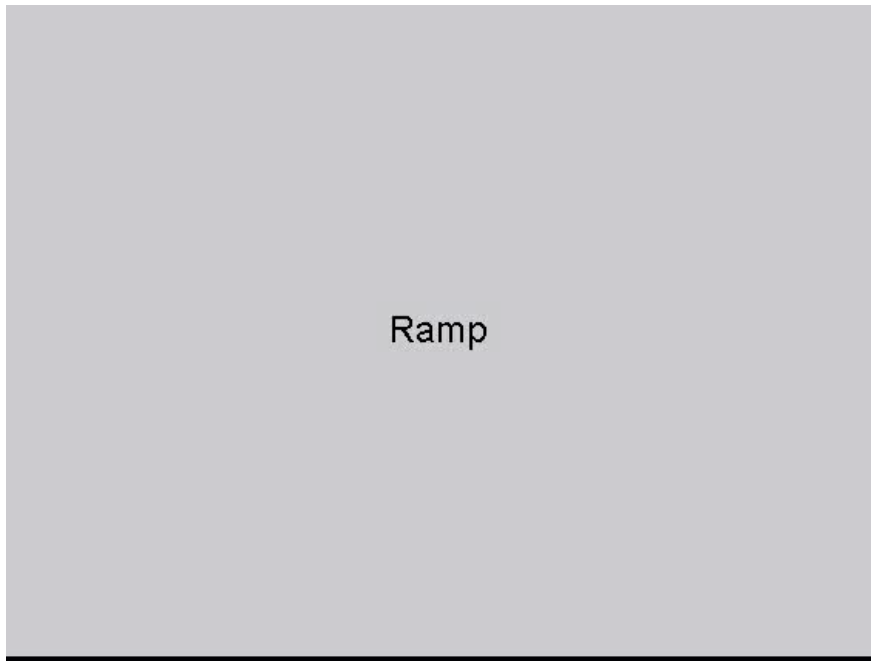
Sample Results

- Pole clip: lamp post & frame edges occlude body



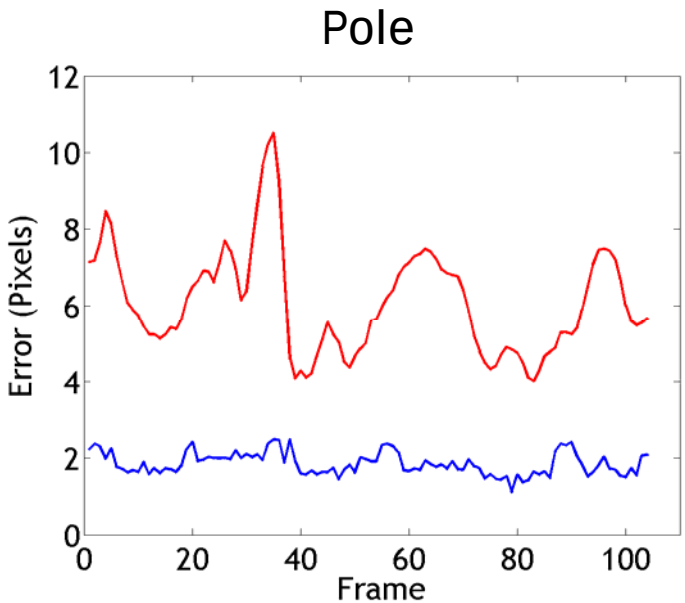
Sample Results (2)

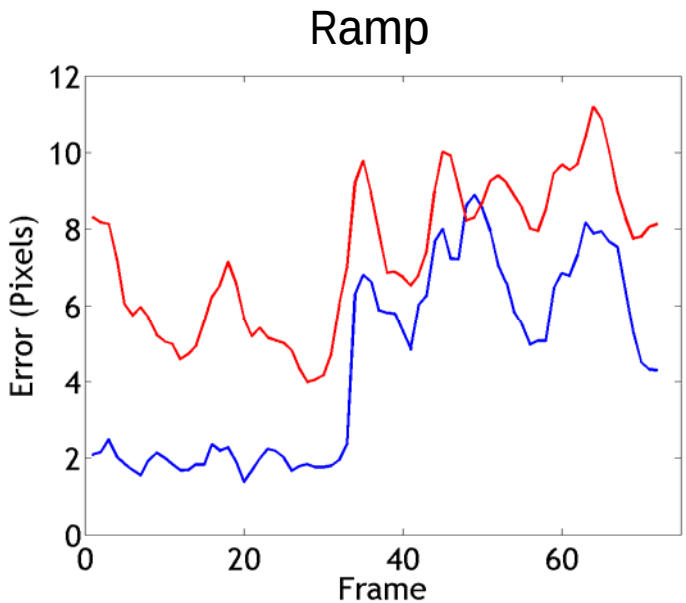
- Ramp clip: low wall occludes legs



Evaluation

- How accurate are the results?

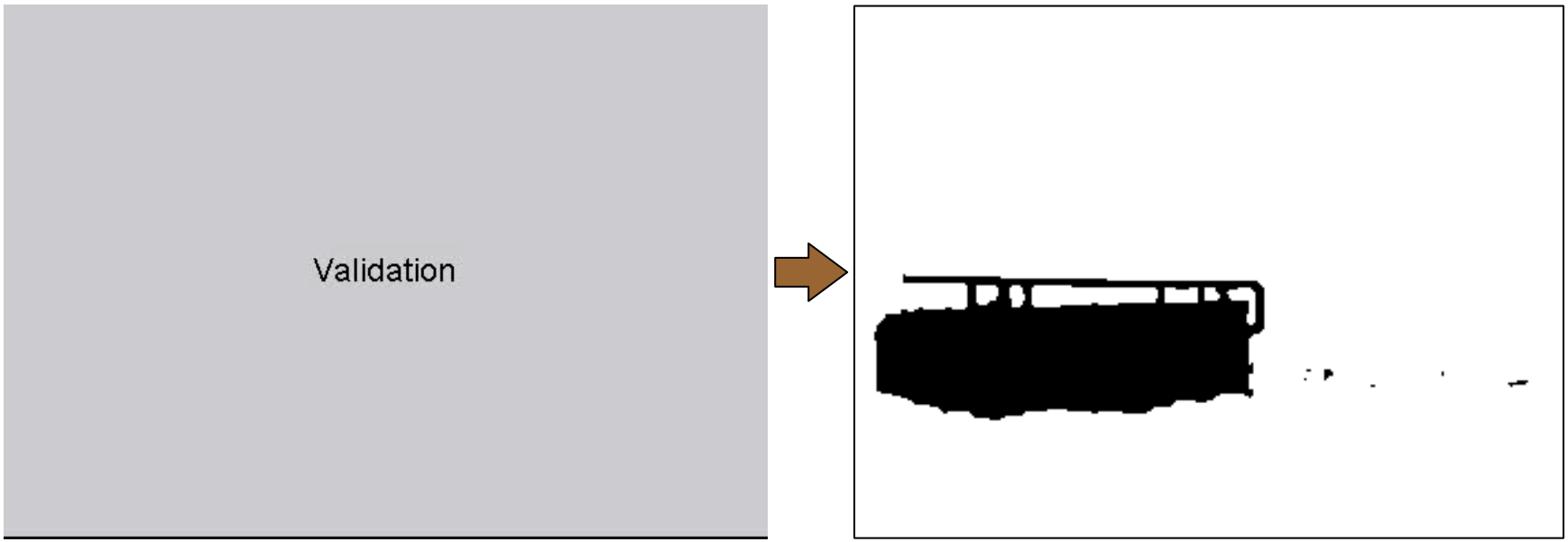




Blue = human error; Red = reconstructed error

Validated Occlusion Maps

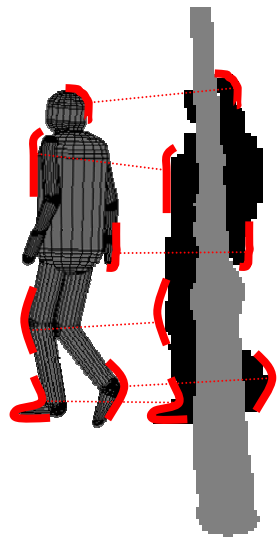
- Reconstructed pose trace can validate areas of occlusion map



☞ *Refine occlusion maps with experience*

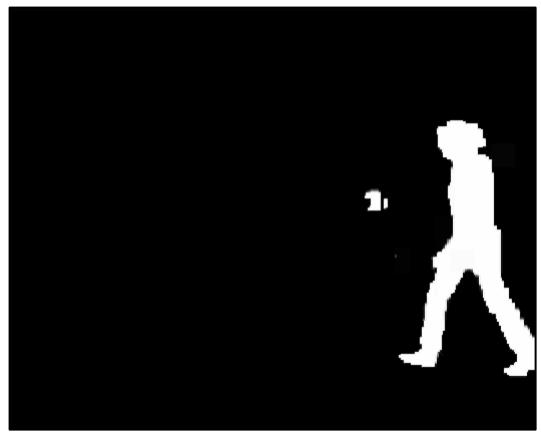
Conclusion

- 3D articulated reconstruction of 1-camera video with external occlusion
  - Silhouette lookup framework
  - Uses boundary fragment matches
  - Requires visible boundary >~33%
- Occlusion maps
  - Initial estimate using integrated silhouette
  - Validated maps reveal true occluding objects



Silhouette Extraction

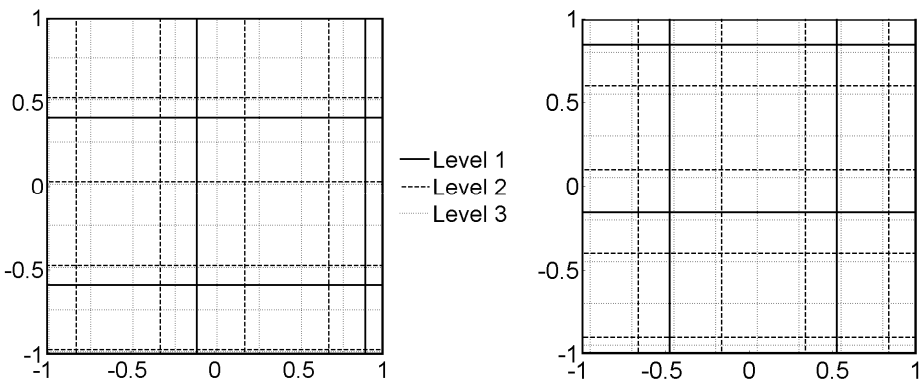
- Many candidate approaches.
  - Moving & fixed camera



- This work:
  - Static camera
  - Graph-based segmentation

EMD Embedding

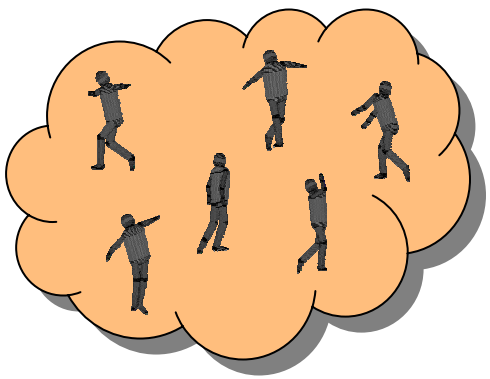
- Each boundary fragment is 15-vector
- Ranges divided in pyramidal scheme
  - Coarse-to-fine
  - Random shift
- Weight assigned to non-empty bins



← Examples of random pyramidal bin divisions

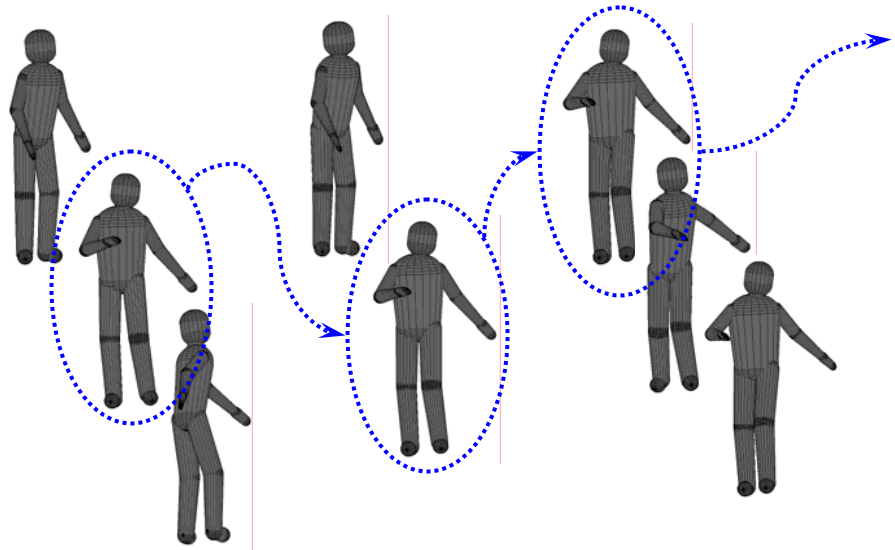
Motion Capture from Silhouette

Observed Silhouettes



*Library of silhouettes with known 3D poses**

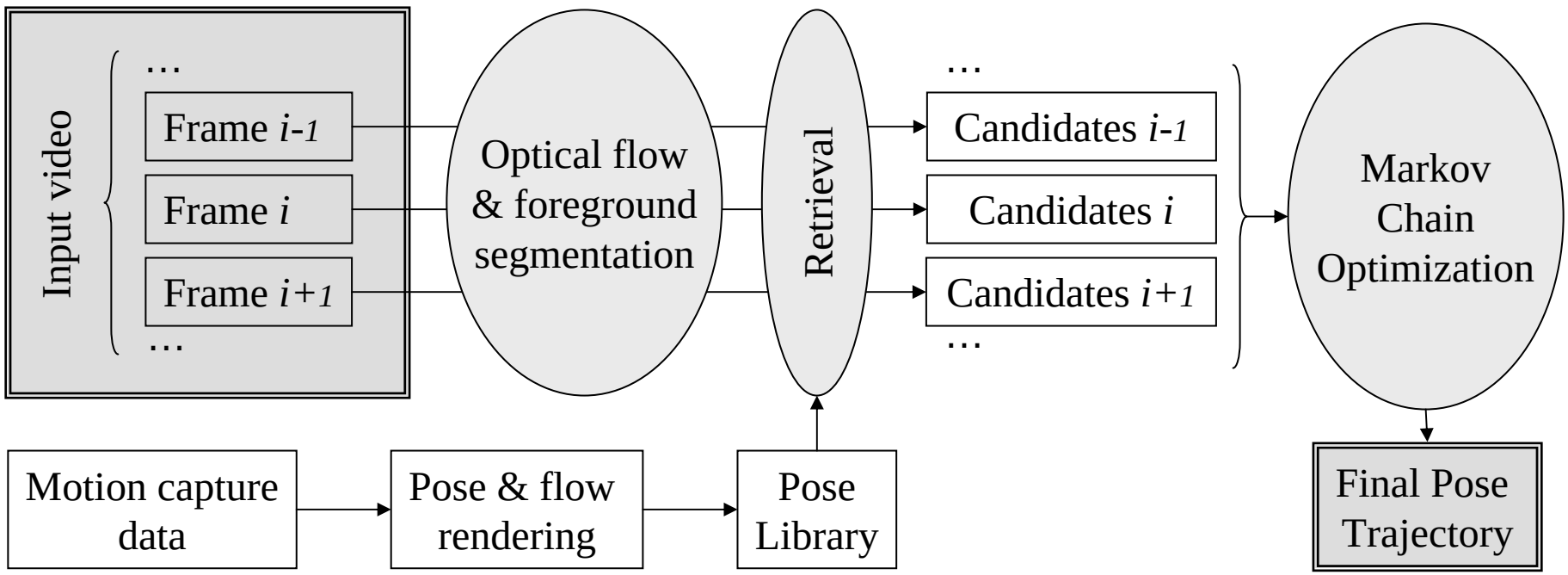
Frame 1 *Frame 2* *Frame 3* *etc.*



1. Each video frame selects **candidate poses** with similar silhouettes from the library
2. Pose reconstruction = smoothest frame-to-frame sequence of candidate poses

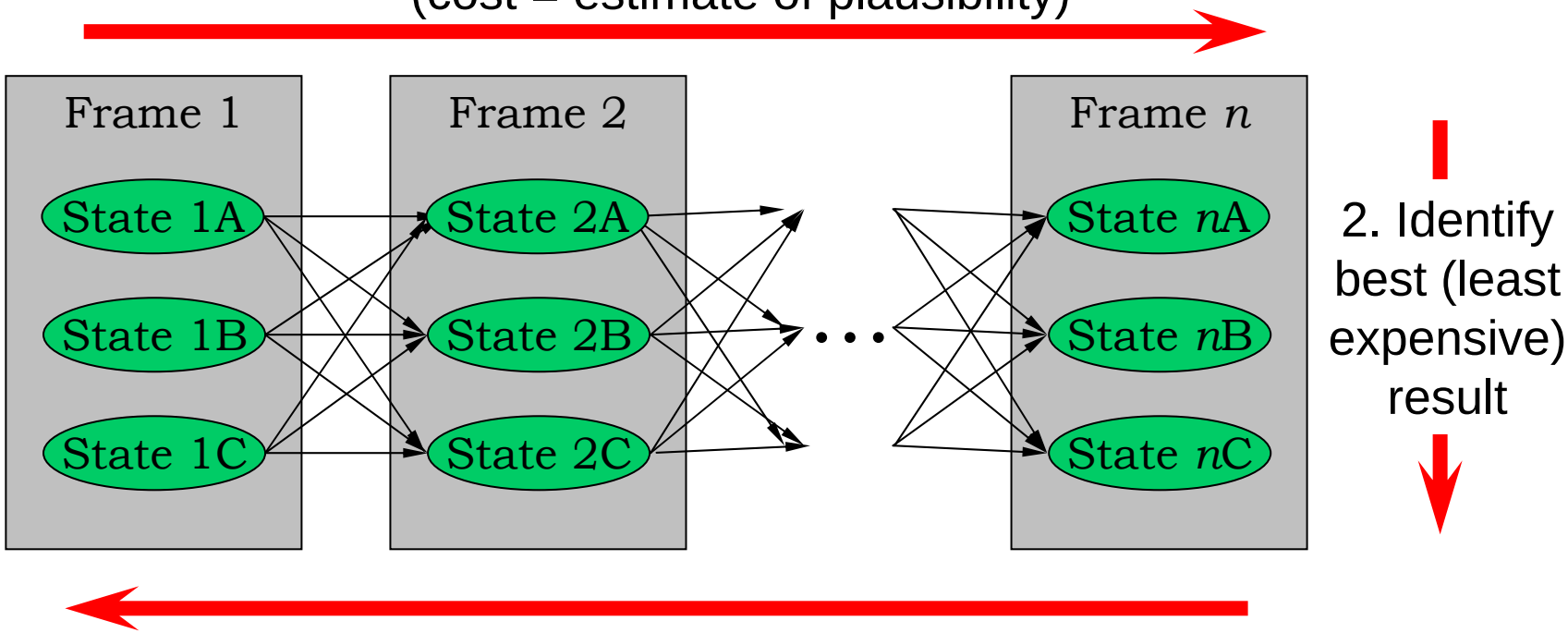
**(Open question: how many library poses are enough for general motion?)*

Flow Chart of Motion Capture



Markov Chain Minimization

1. Compute least expense to reach each state from previous frame (cost = estimate of plausibility)



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