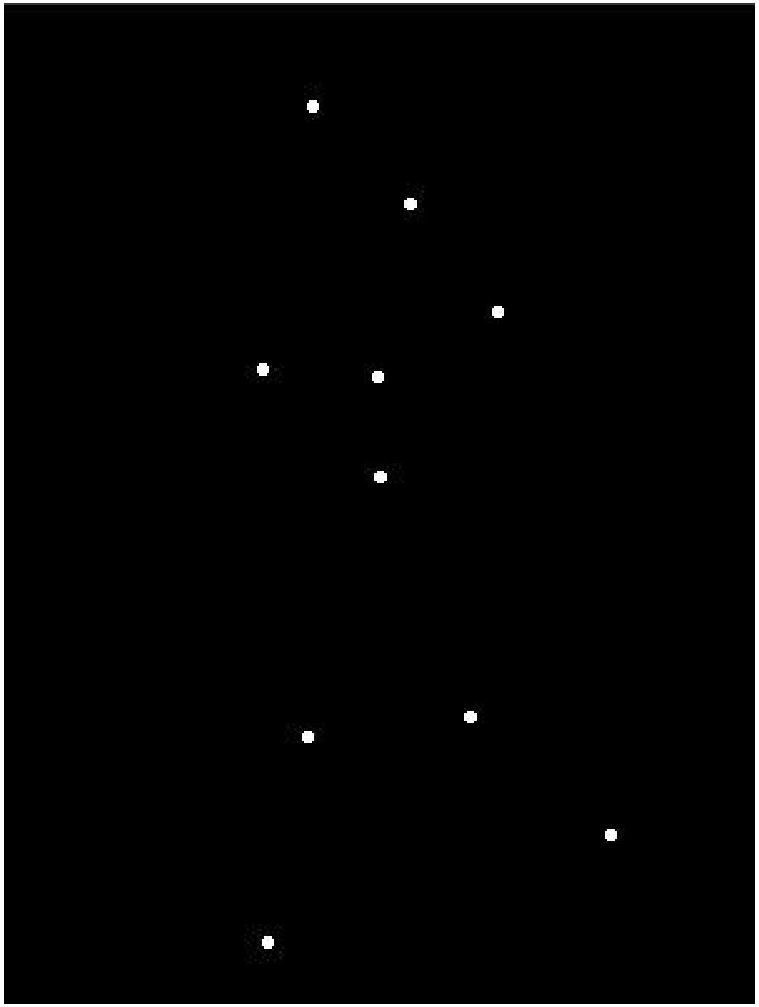
Learning Articulated Skeletons From Motion

Danny Tarlow
University of Toronto, Machine Learning

with
David Ross and Richard Zemel
(and Brendan Frey)

August 6, 2007

Point Light Displays



- It's easy for humans to recognize biological motion, and structure
- Other domains:
  - motion capture
  - animation
  - computer vision

Summary

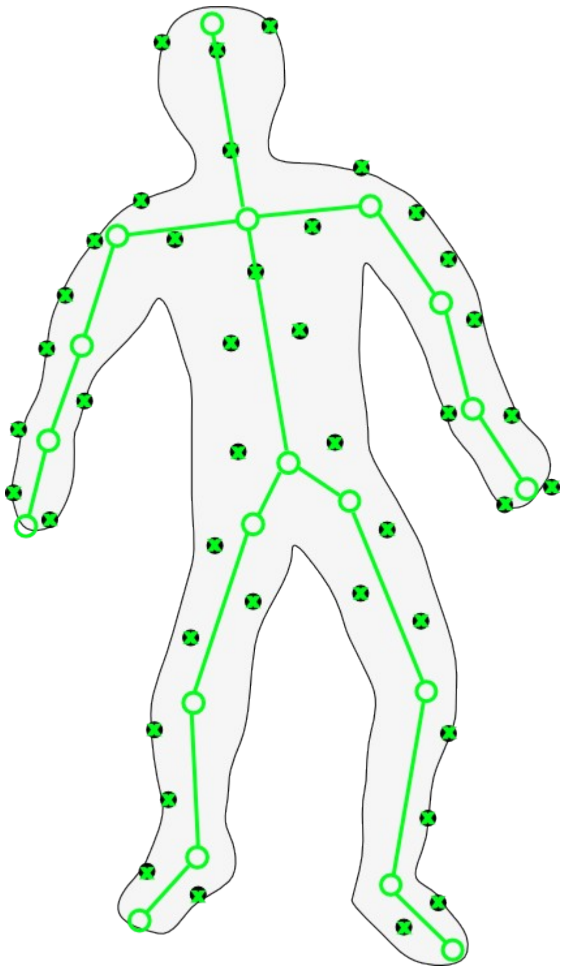
- Goal: given a time-series of feature positions, learn skeleton (structure) and pose
- Approach: formulate as a probabilistic model, unsupervised learning
- Subcomponents:
  - assigning features to sticks
  - connectivity of sticks
  - local geometry and motion of each stick
- Evaluation: on 2D and 3D datasets, including human mocap, multiple actors, video of giraffe

Obligatory CIFAR Slide

- We are learning a representation that is more amenable to higher level tasks
- Why not a deep belief net?
  - Very specific types of correlation that we're interested
  - Animation: adding a deformable mesh on top of skeleton
  - Generalizing to “similar” skeletons (stretch bones, etc.)
  - Ask Graham Taylor

Articulated Motion

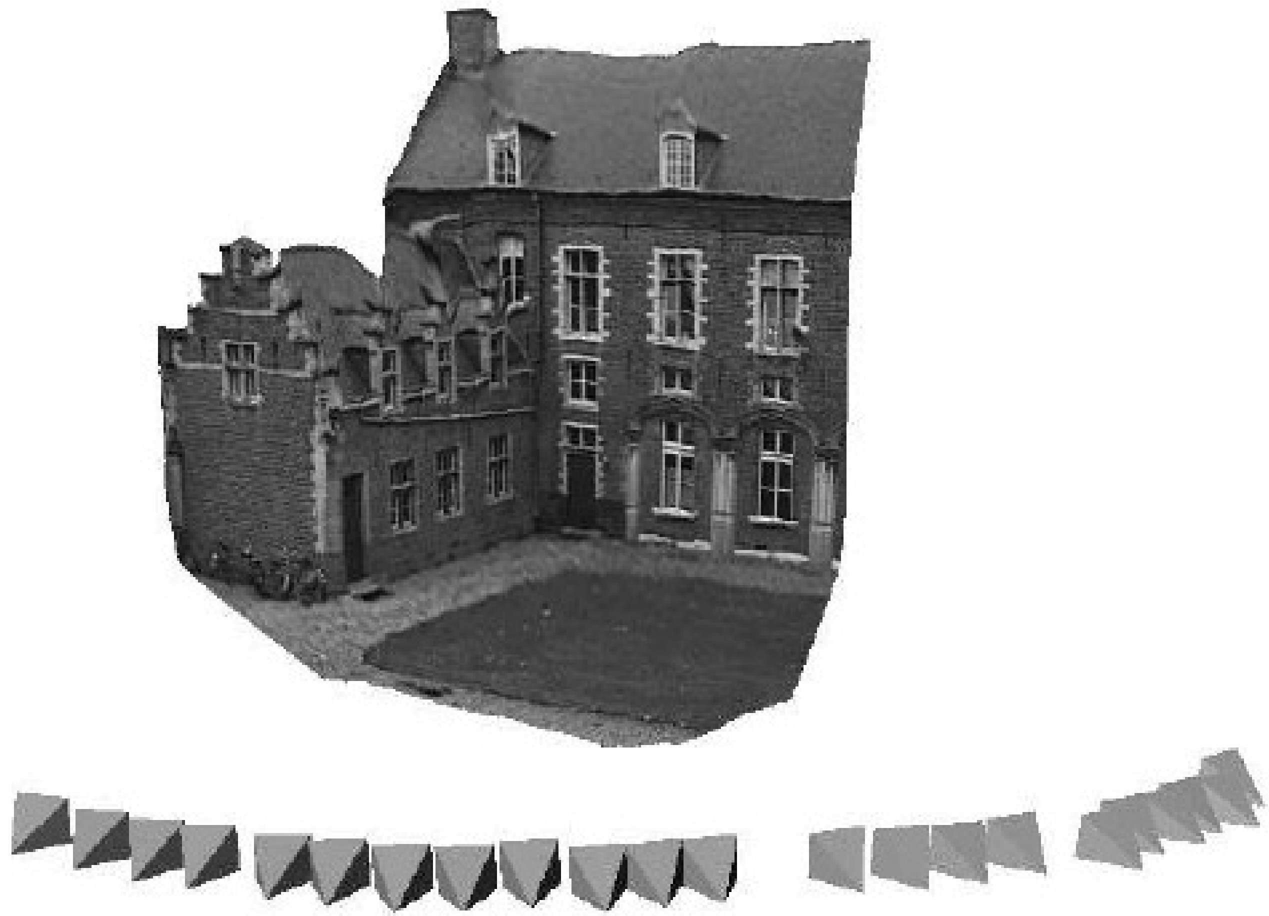
- Most interesting objects (humans, animals) aren't rigid
- Approximate as a connected set of rigid parts (i.e. stick figure)
- Multibody SFM won't work
  - motion dependence
  - doesn't recover connectivity, joint locations
- Our approach: probabilistic model of articulated stick figure(s)



Structure From Motion

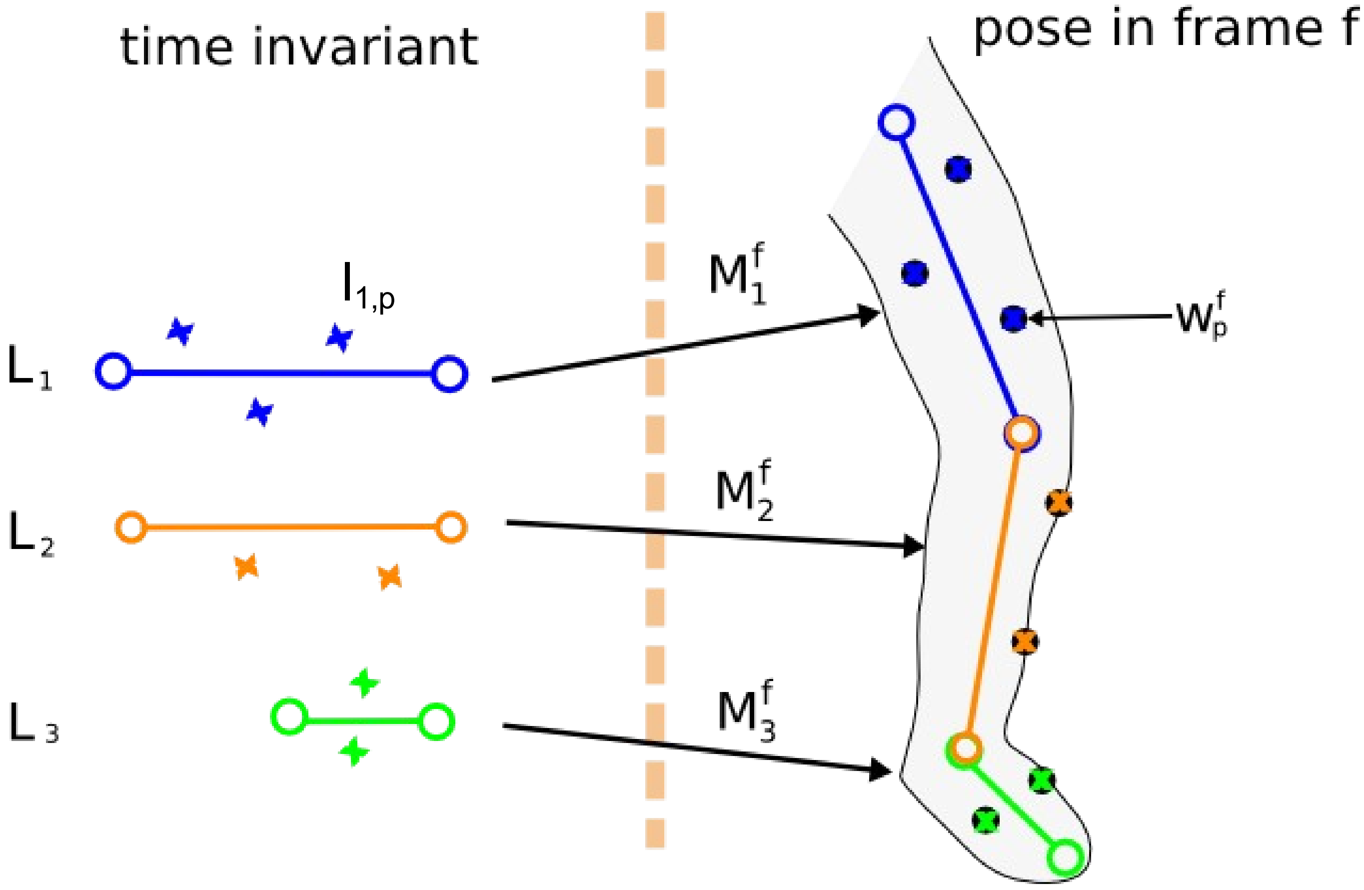
- Classic Problem: single rigid object viewed from multiple angles, from 2D feature point locations, recover:
  - relative position of features (3D structure)
  - pose of object in each frame (motion)
- Linear Solution: factorize W = M L using SVD
  - We assume orthogonal projection
- Multibody: segment feature points into objects, solve SFM independently for each
- We'll also deal with 3D from 3D (i.e. optical motion capture data)

Structure From Motion



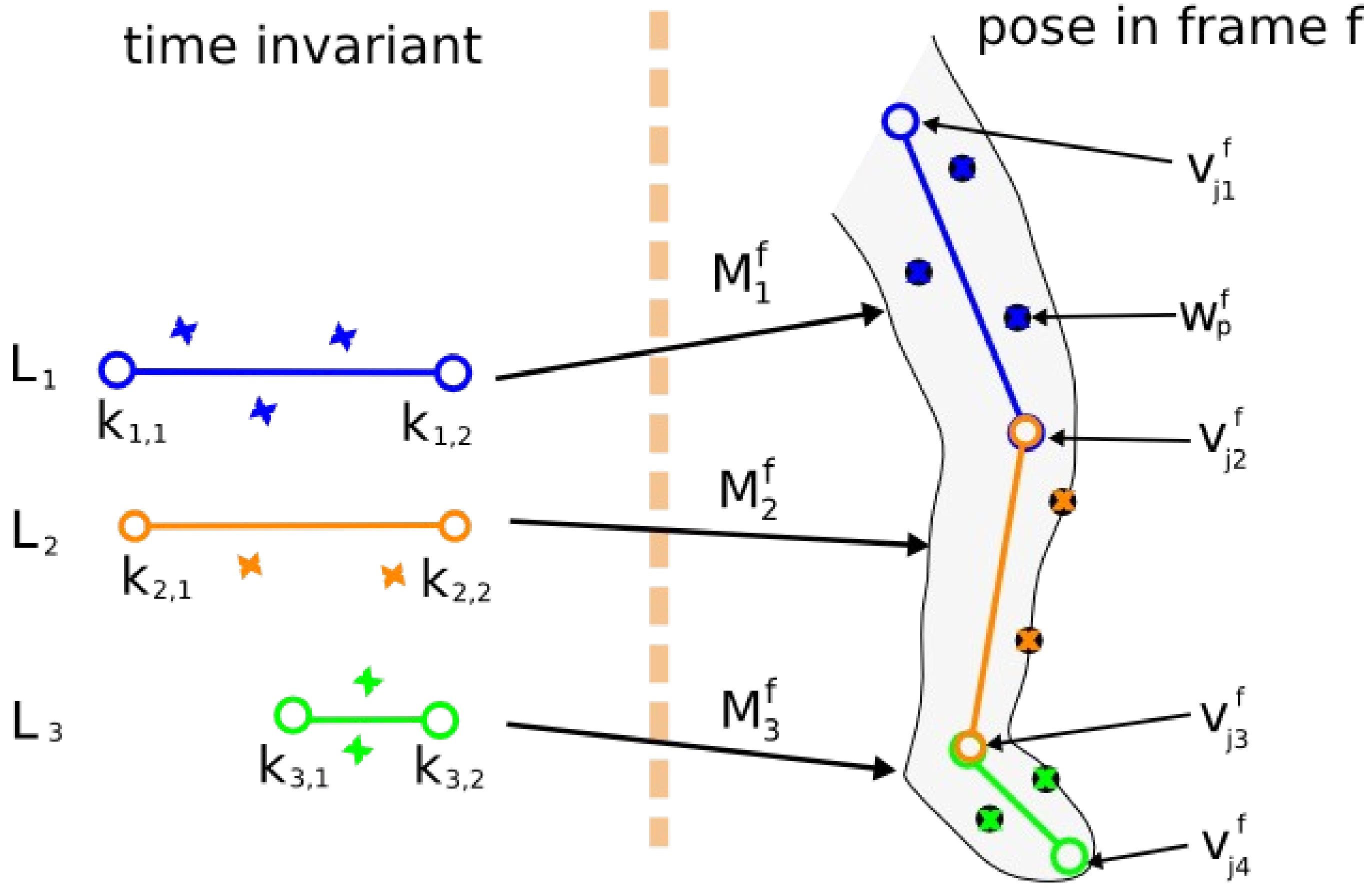
[image by Marc Pollefeys]

Local Geometry & Motion of Each Stick



- Probabilistic approach: $P(W|M,L) = \prod_{f,p,s} N(w_p^f \,|\, M_s^f l_{s,p},\ \sigma_w^2)^{r_{s,p}}$
- Related to factor analysis, fit using EM (talk to Yair)

Dependent Motion

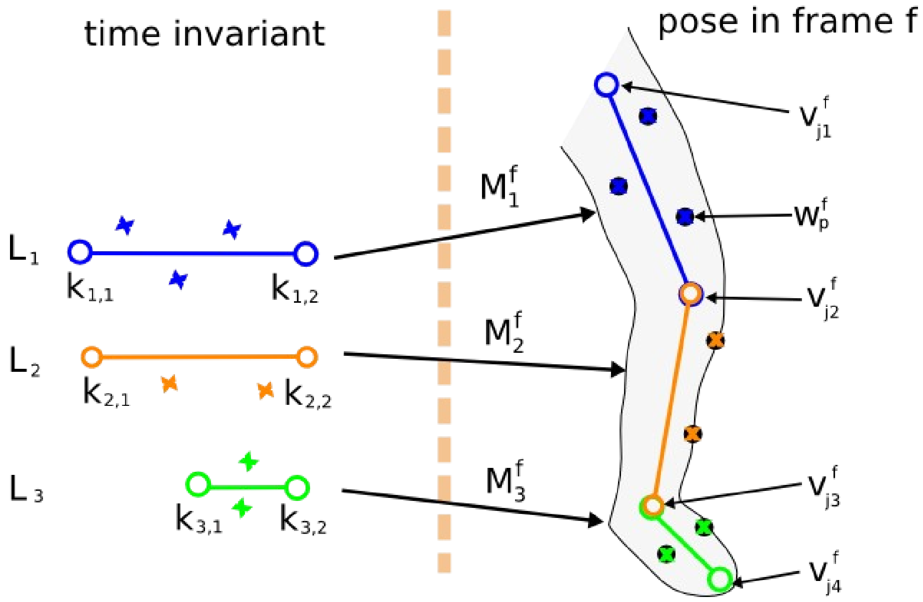


- Motions are constrained: $M_1^f k_{1,2} = M_2^f k_{2,2} = v_{j2}^f$
- Introduce auxiliary variables (endpoint & joint locations): factorizes into independent SFM problems

Dependent Motion: Details

$$P(M|S)=\iint P(M,V,K|S)\,\partial V\,\partial K$$

$$P(M,V,K|S)=P(V|M,K,S)\,P(M|S)\,P(K|S)$$

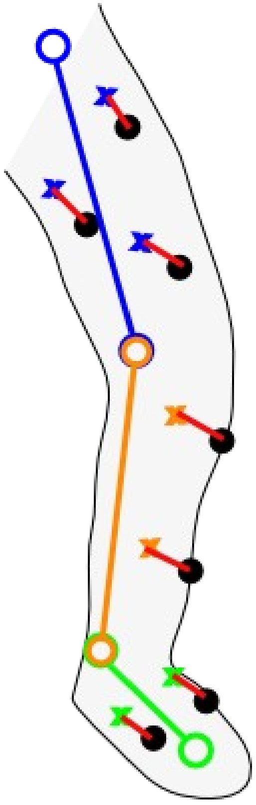


$$P(V|M,K,S)=\prod_{f,s,j,e} N(v_j^f \mid M_s^f k_{s,e},\, \sigma_v^2)^{g_{s,e,j}}$$

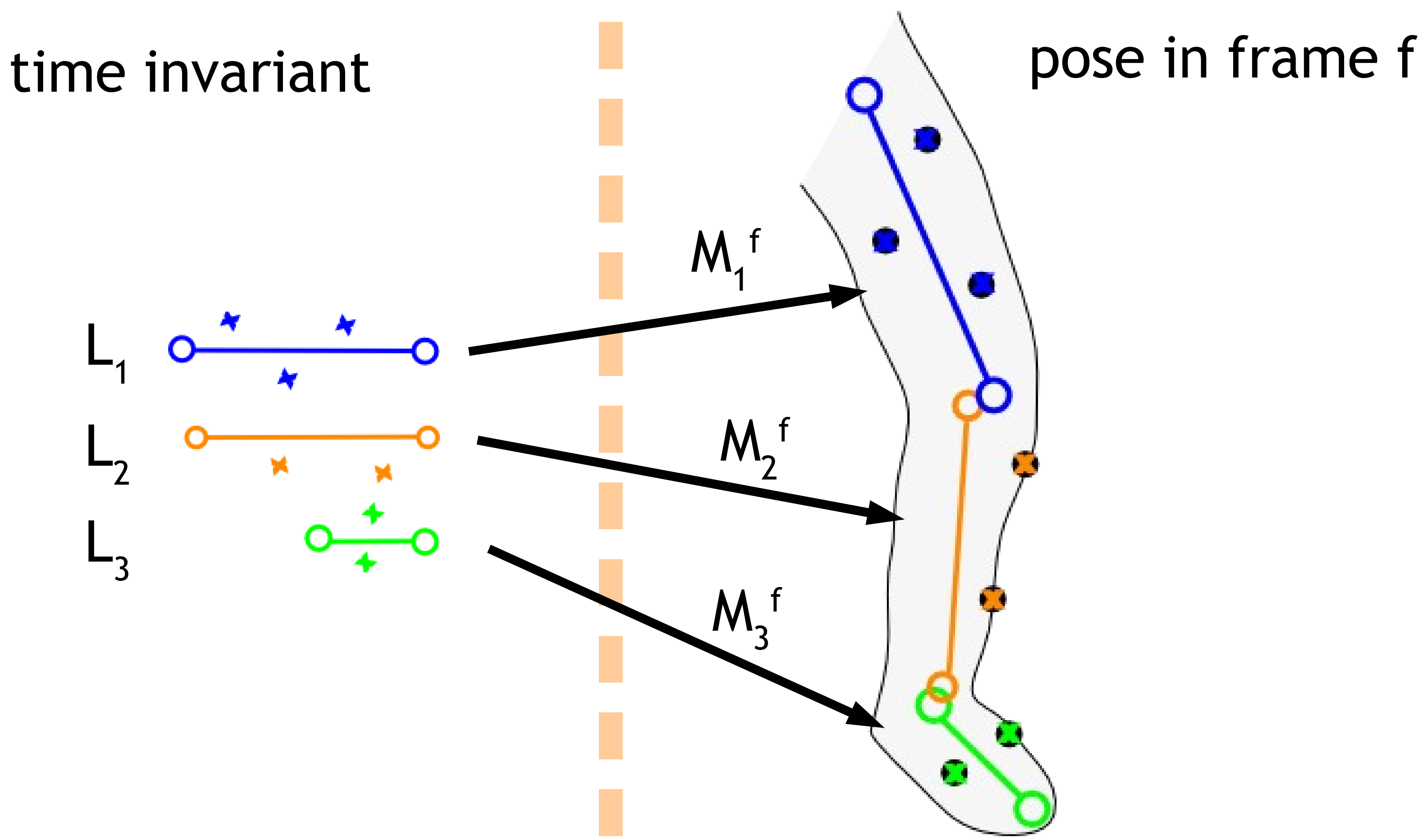
$$P(M|S)=\prod_{s,f} N(M_s^f \mid M_s^{f-1},\, \sigma_m^2)$$

$$P(K)=Broad\,Gaussian$$

Cost Function: Point Alignment



Cost Function: Joint Alignment



Stick Connectivity

- Computationally intractable to consider all skeletons
- Possible to solve for one unknown joint (via optimization of joint-probability)
- Greedy approach:
  - start with fully-disconnected skeleton
  - estimate change in cost for each possible joint (store these in a table)
  - incrementally connect stick endpoints until performance on validation set stops improving
- Efficiency: only a few costs must be reestimated after each stage

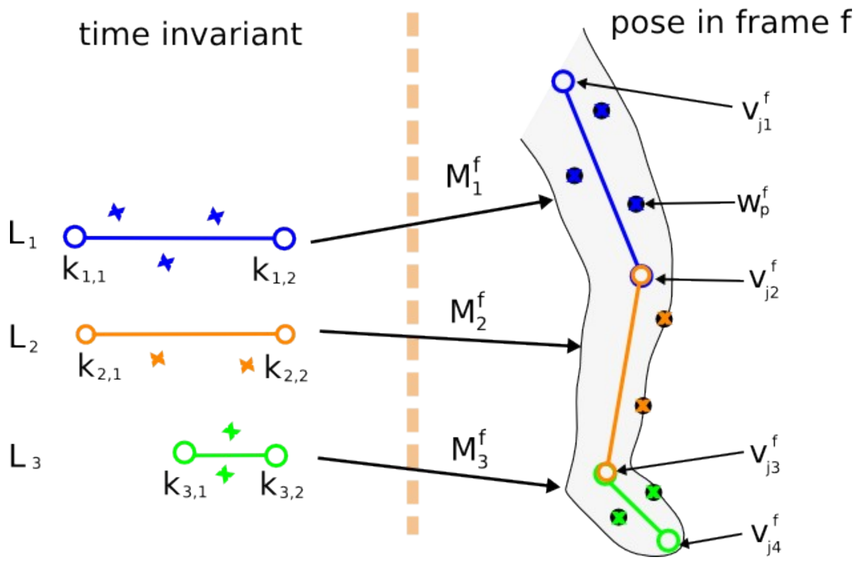
Identifying Sticks

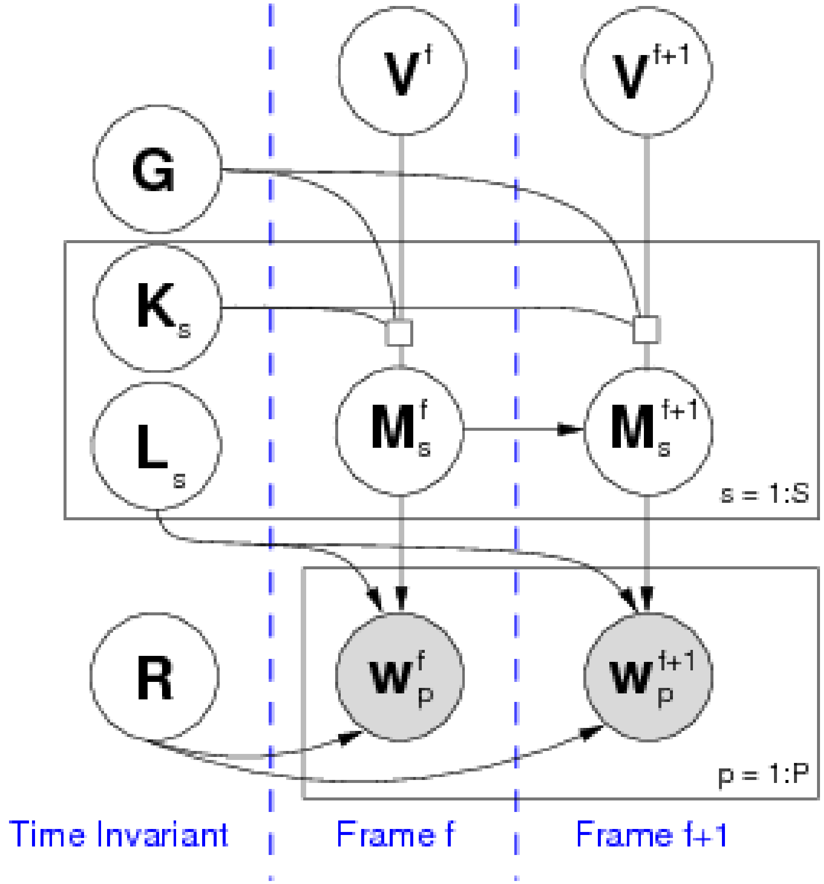
- How many sticks? Which points are connected to which sticks?
- Calculate a pairwise (dis)similarity measure:
  - 3D use standard deviation of distance [Kirk '05]
  - 2D use angle between local subspaces [Yan '06]
- Construct an empirical prior P(R), sample reasonable segmentations
- Use “Affinity Propagation” segmentation [Frey-Dueck '07]
- More recently, frame as CRP and alternate with local search for structure

Big Picture & Recap

1) Sample a segmentation of feature points trajectories
2) Assuming a disconnected skeleton, solve SFM independently for each stick
3) For each possible way to join sticks, compute cost (change in probability) save in a table
4) Iteratively join sticks (greedy), updating costs as necessary
5) Stop when validation error becomes large

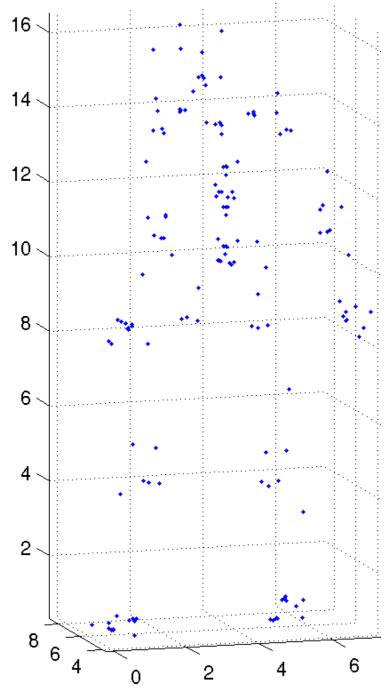
Graphical Model





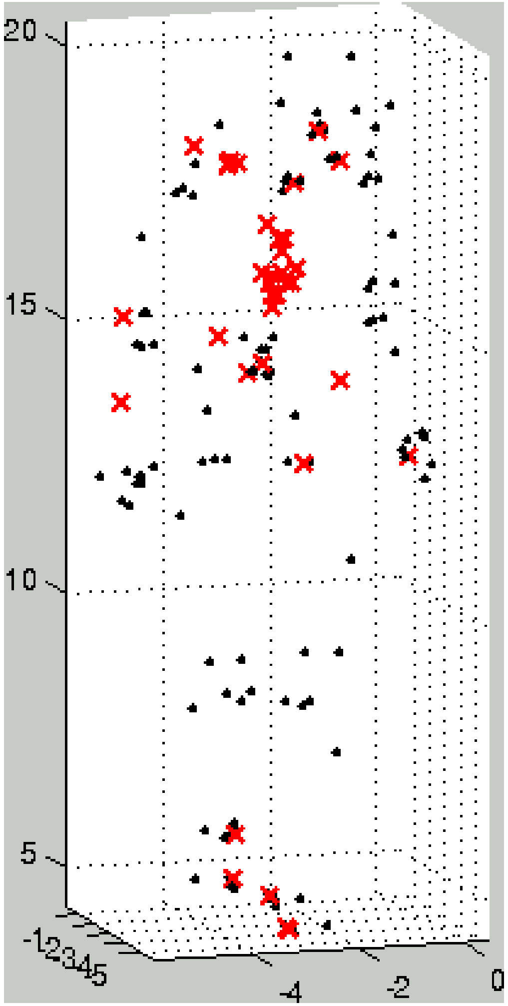
Experimental Evaluation

- Trained on 2D and 3D datasets
- Human motion capture data http://mocap.cs.cmu.edu/

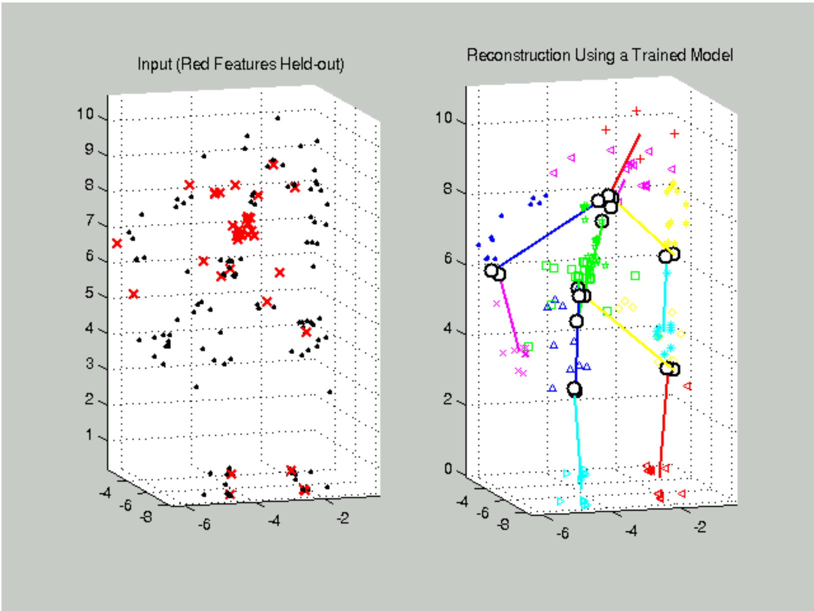


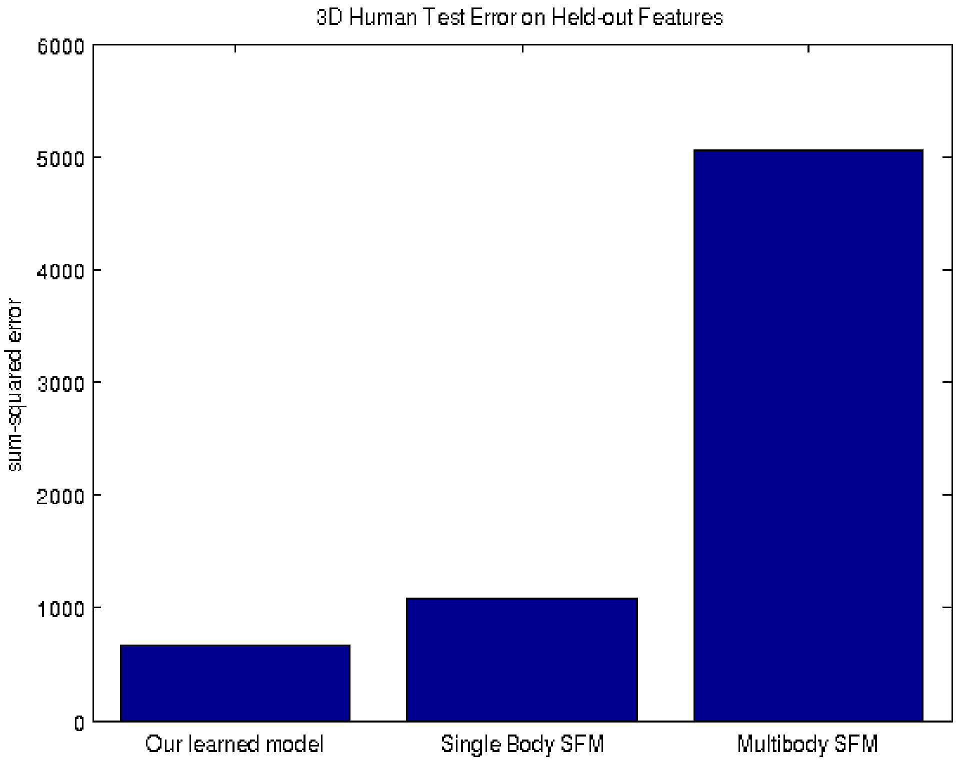
Experimental Methodology

- 60% of frames for learning, 20% for validation (model selection), 20% for measuring test performance
- validation & test sets, hold out 10% of feature points + one stick
- using learned model and visible features, estimate locations of held-out points
- compute squared error between estimated & true positions of heldout features



3D Human Reconstruction

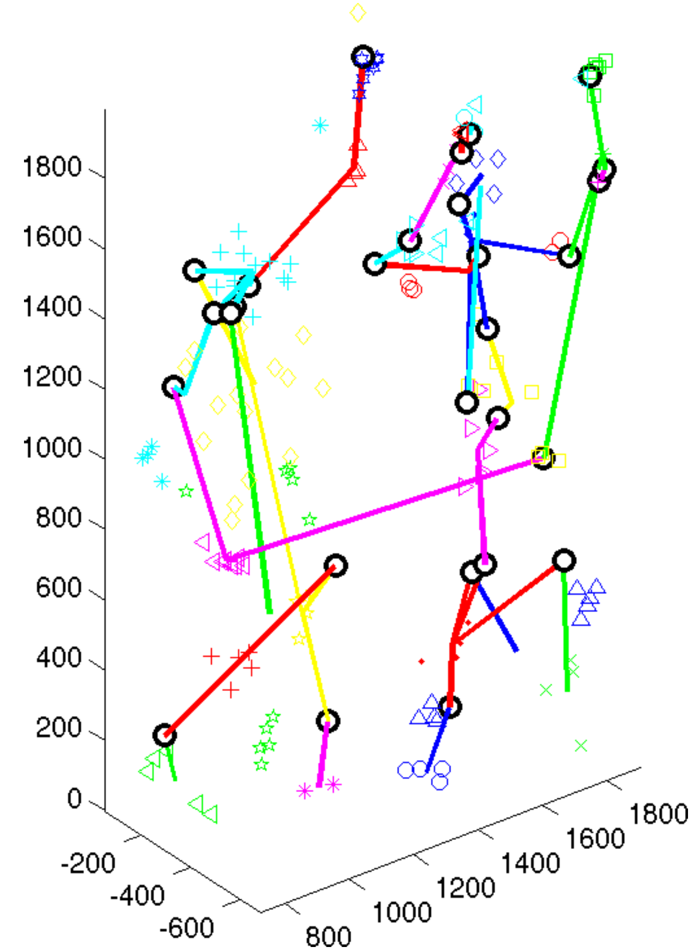




- Video
- Performance

Related Work

- Yan-Pollefeys (2005,6) mainly concerned with 2D segmentation; no global cost function
- Kirk-O'Brien-Forsyth (2005) works on 3D data only; uses spanning tree
- Anguelov (2004) works on 3D meshes; connectivity between sticks is known



KOF on Football data

Recent Directions

- 2 directions
  - Up: generalize structure learning model, more complex structures and motions.
  - Down: don't assume correspondences are known

The Correspondence Problem

- Take as input raw video... can we do the same stuff?
- Show giraffe video

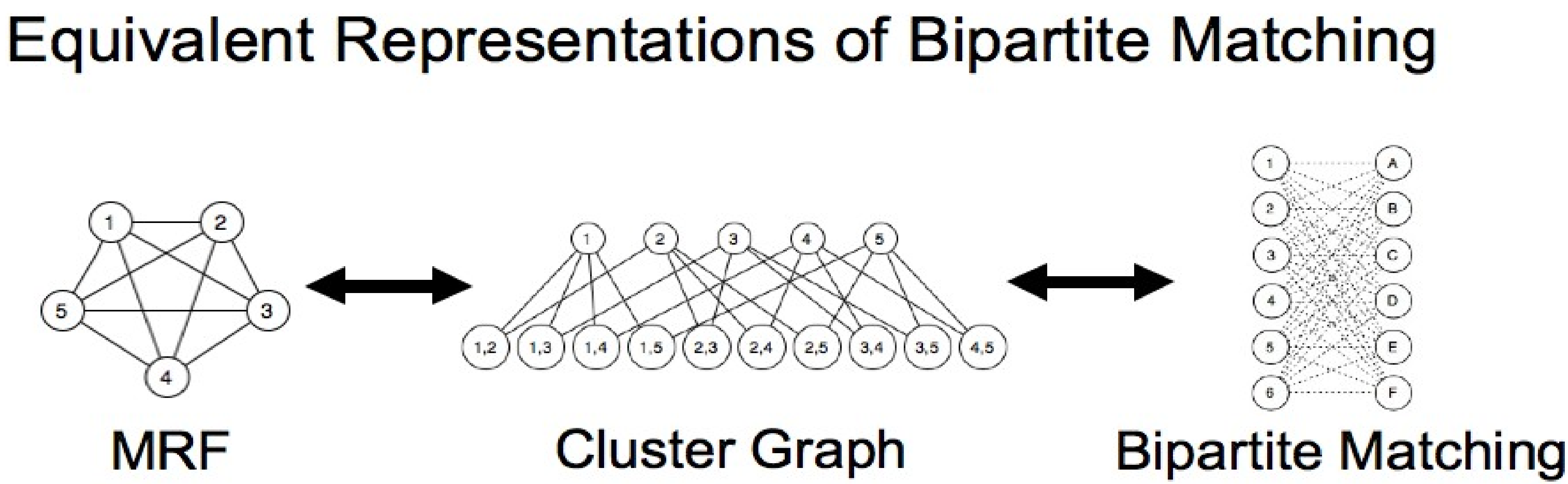
The Correspondence Problem

- Much harder than it seems.
  - Tried KLT tracker (optical flow)
    - Feature drift
    - Needed a lot of hand-corrections
  - Tried SIFT matching
    - Expensive to run on every frame
    - Didn't match anything on legs
    - Still needs distinct textures

The Correspondence Problem

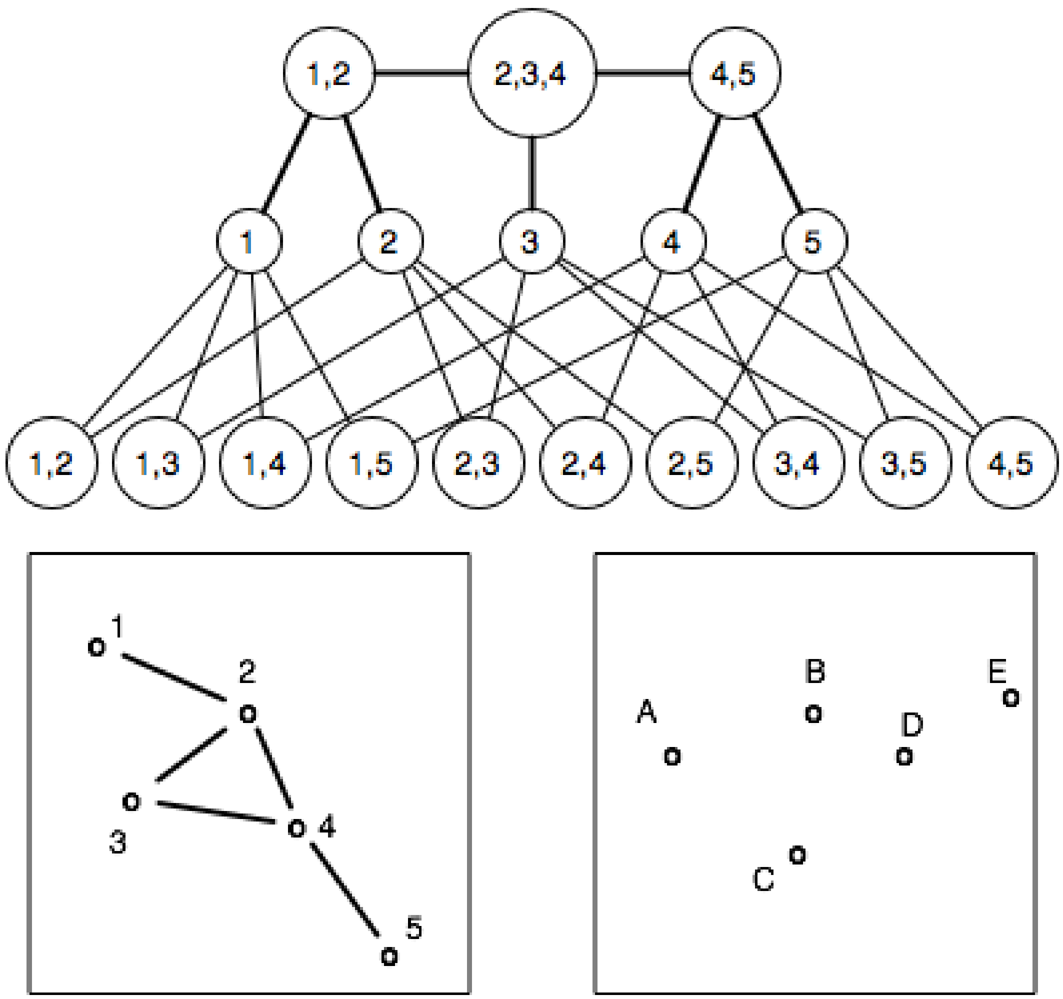
- Many different approaches, all(?) leverage some subset of:
  - Appearance (SIFT features, image neighborhood intensities)
  - Temporal smoothness / small movement prior
  - 2D Geometric Constraints
  - 3D Geometric Constraints / Rank-based Constraints
- Matching can be:
  - One-to-one (weighted bipartite matching problem)
  - Nearest neighbor
  - Ratio of nearest to second-closest neighbor (Lowe)

The Correspondence Problem



- Correspondence problem has a lot of structure
  - This diagram just helps make it explicit

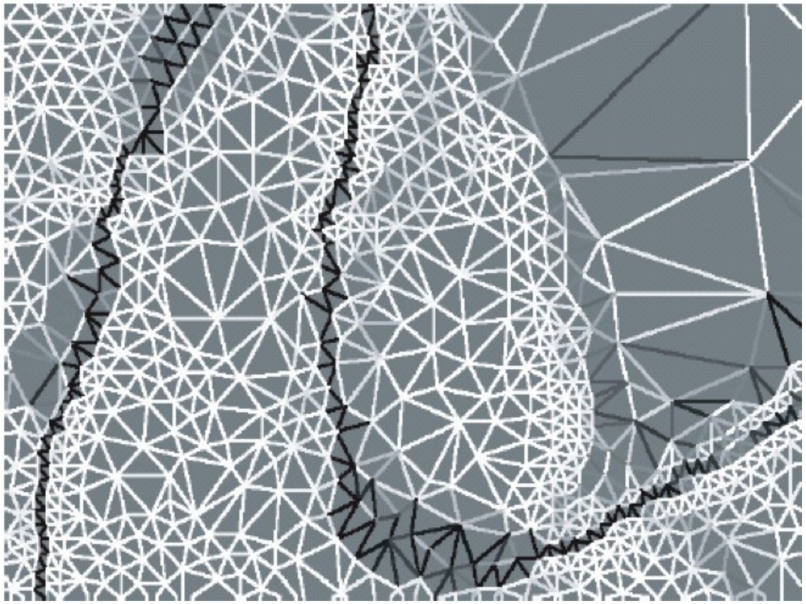
2D Geometric Correspondence Constraints



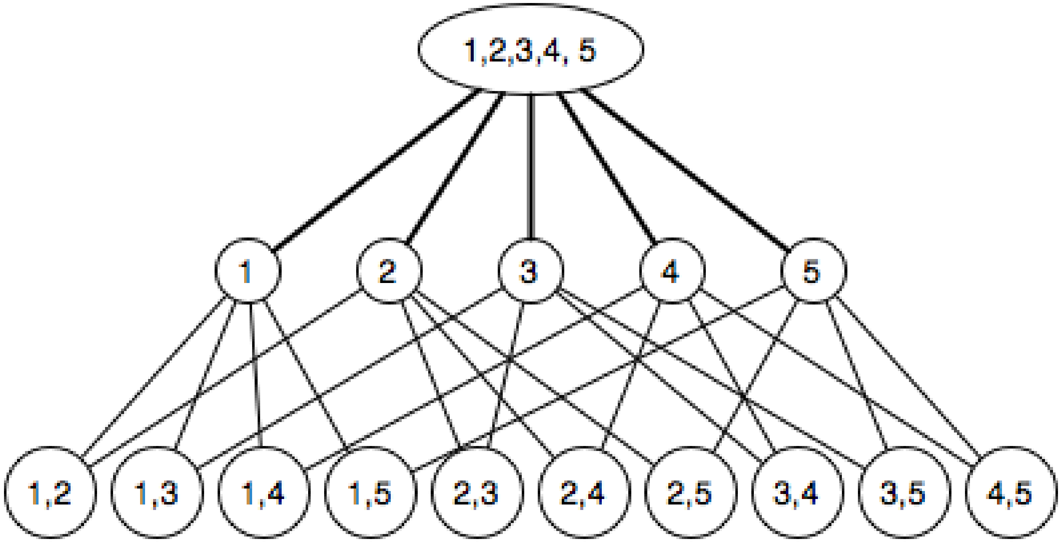
* Ask me afterwards for as much detail as you want

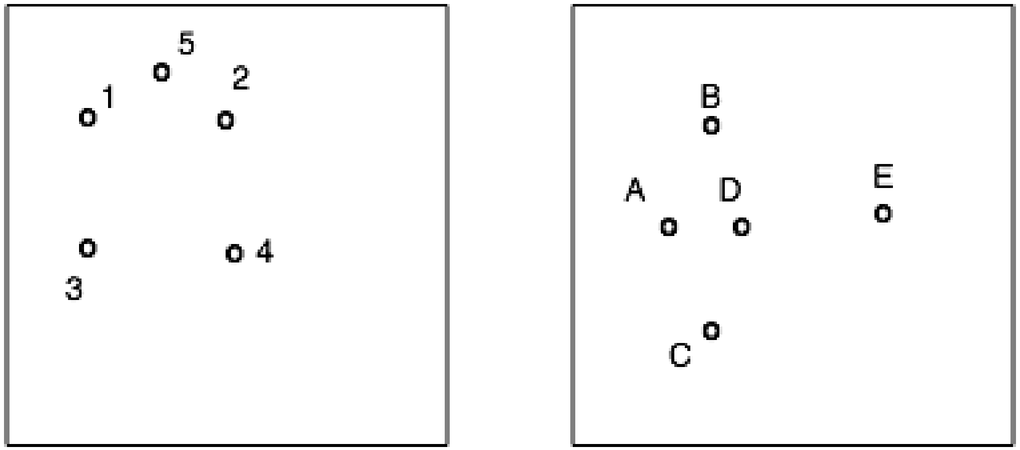
2D Geometry in Video

- This can be made to work surprisingly well
  - P. Sand and S. Teller. *Particle video: Long-range motion estimation using point trajectories.* CVPR 2006.



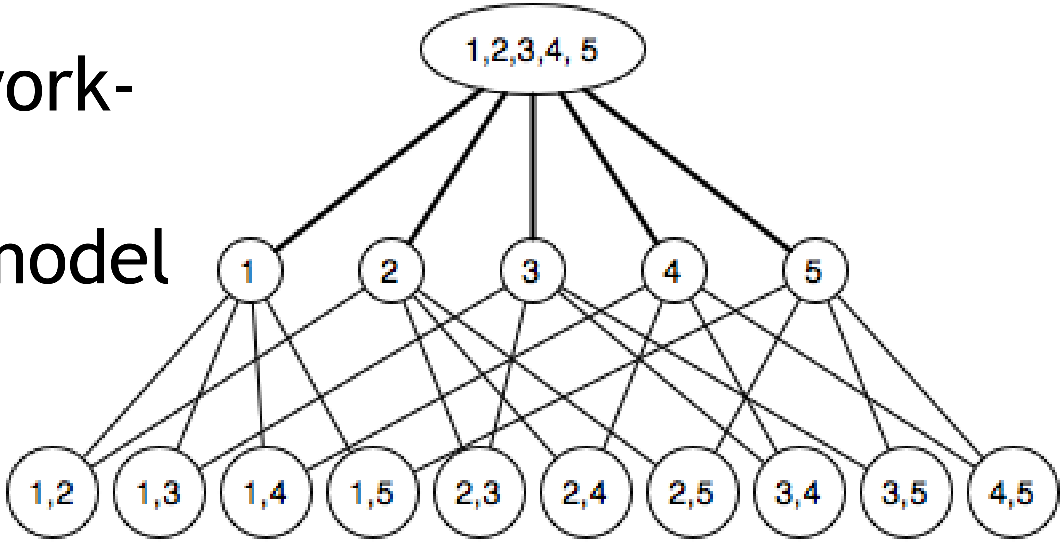
3D Geometric Correspondence Single Rigid Body Constraints

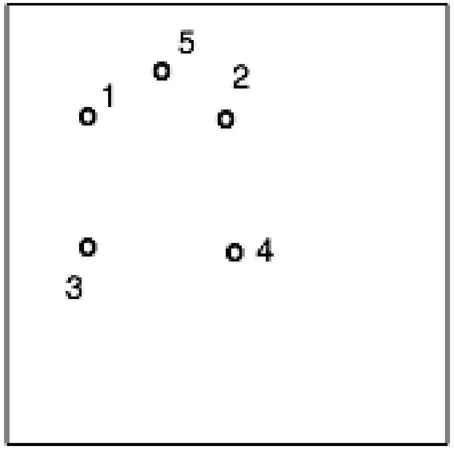


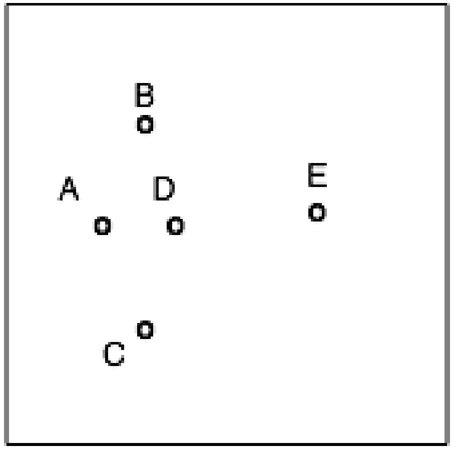


3D Geometric Correspondence
Single Rigid Body Constraints

Two common work-arounds:
- EM + outlier model
- RANSAC

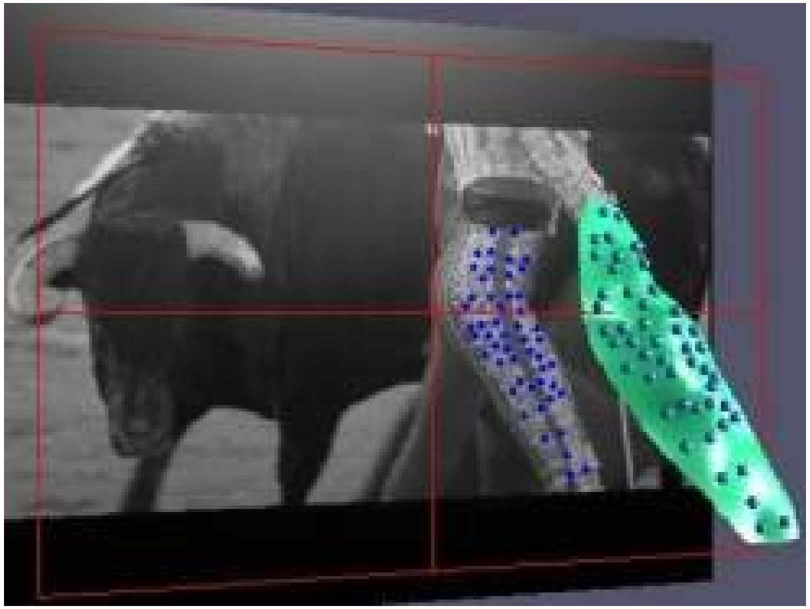






3D Geometry Video

- This can be made to work surprisingly well
  - Lorenzo Torresani and Aaron Hertzmann. *Automatic Non-Rigid 3D Modeling from Video*, ECCV 2004.



Correspondences Overview

- Don't assume correspondences are known
  - This opens up a whole new set of issues
- Still want our 3D model of complex underlying structure
  - At least multibody, maybe articulated later
  - But there is lots of information available from 2D and temporal information
- Temporal information?
  - Yes and no.
  - Camera cuts?

End

- Comments / questions?