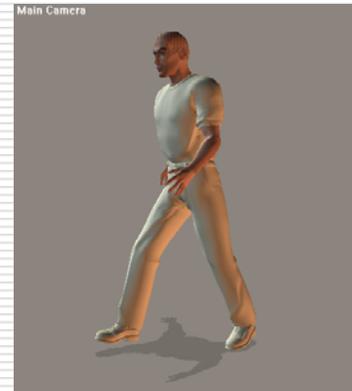




Predicting 3D People from 2D Pictures



Leonid Sigal

Michael J. Black

Department of Computer Science

Brown University

<http://www.cs.brown.edu/people/lis/>

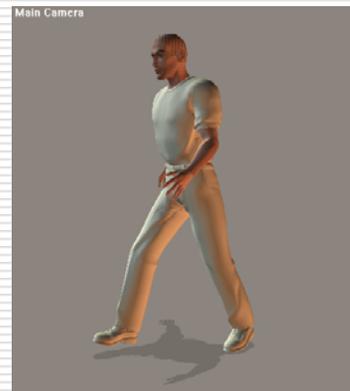


Introduction

Articulated pose estimation from single-view monocular image(s)



(2D) Picture



(3D) Person

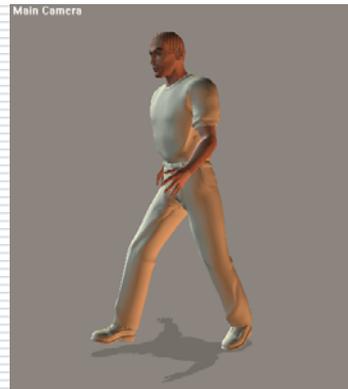


Introduction

Articulated pose estimation from single-view monocular image(s)



(2D) Picture



(3D) Person

- Entertainment:** Animation, Games
- Clinical:** Rehabilitation medicine
- Security:** Surveillance
- Understanding:** Gesture/Activity recognition





Why is it hard?

- Appearance/size/shape of people can vary dramatically



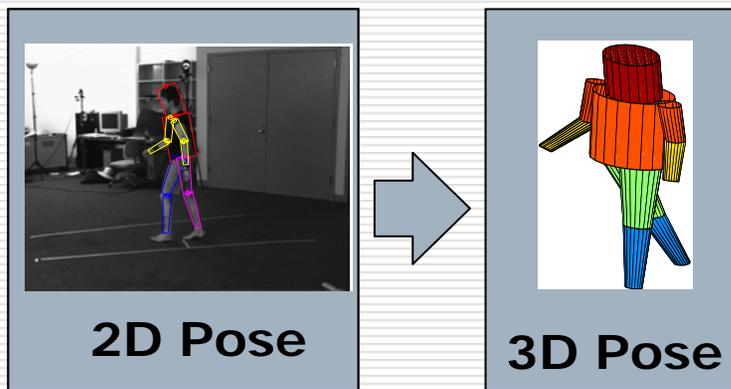
- The bones and joints are observable indirectly (obstructed by clothing)

- Occlusions
- High dimensionality of the state space
- Lose of depth information in 2D image projections



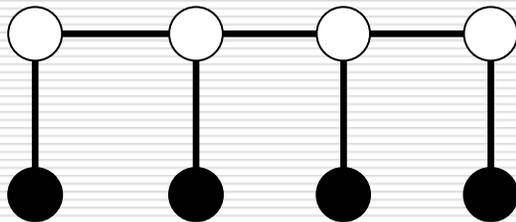
Approach

- Break up a very hard problem into smaller manageable pieces

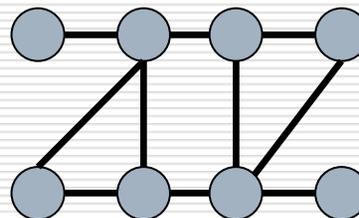


□ Tools

■ Graphical models



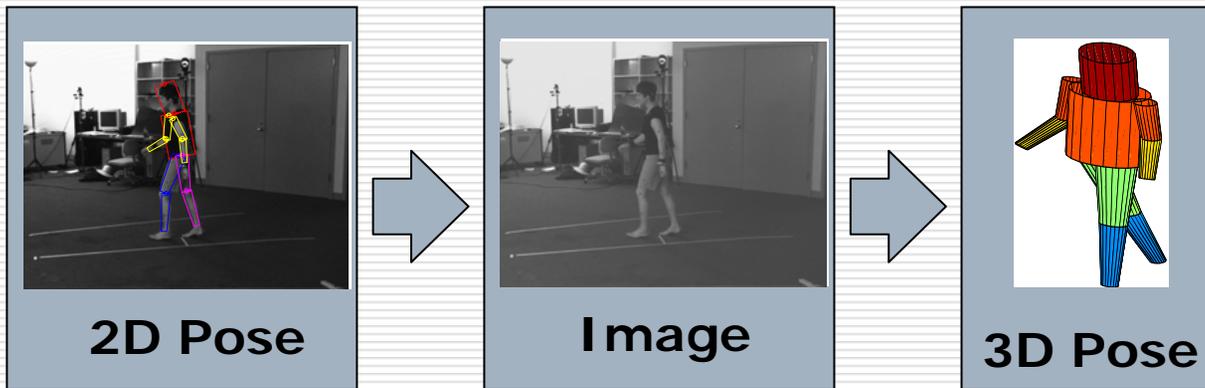
■ Belief Propagation





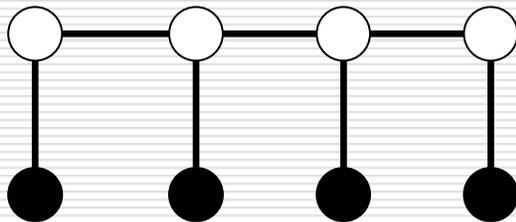
Approach

- Break up a very hard problem into smaller manageable pieces

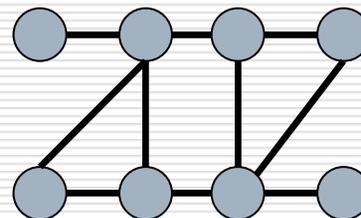


□ Tools

■ Graphical models



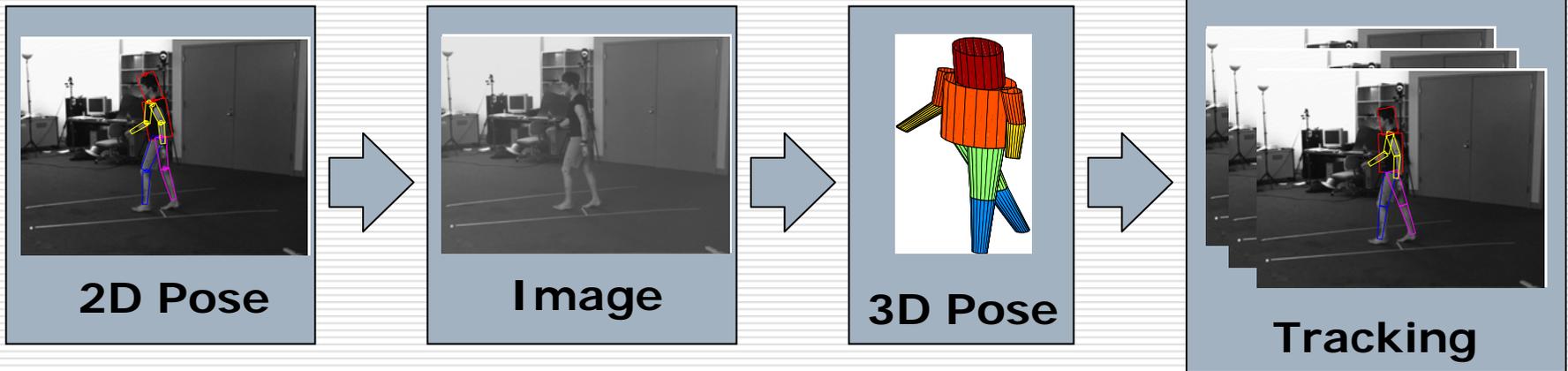
■ Belief Propagation





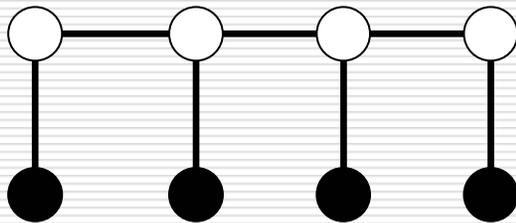
Approach

- Break up a very hard problem into smaller manageable pieces

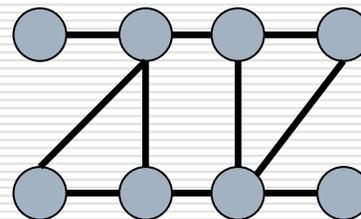


□ Tools

■ Graphical models



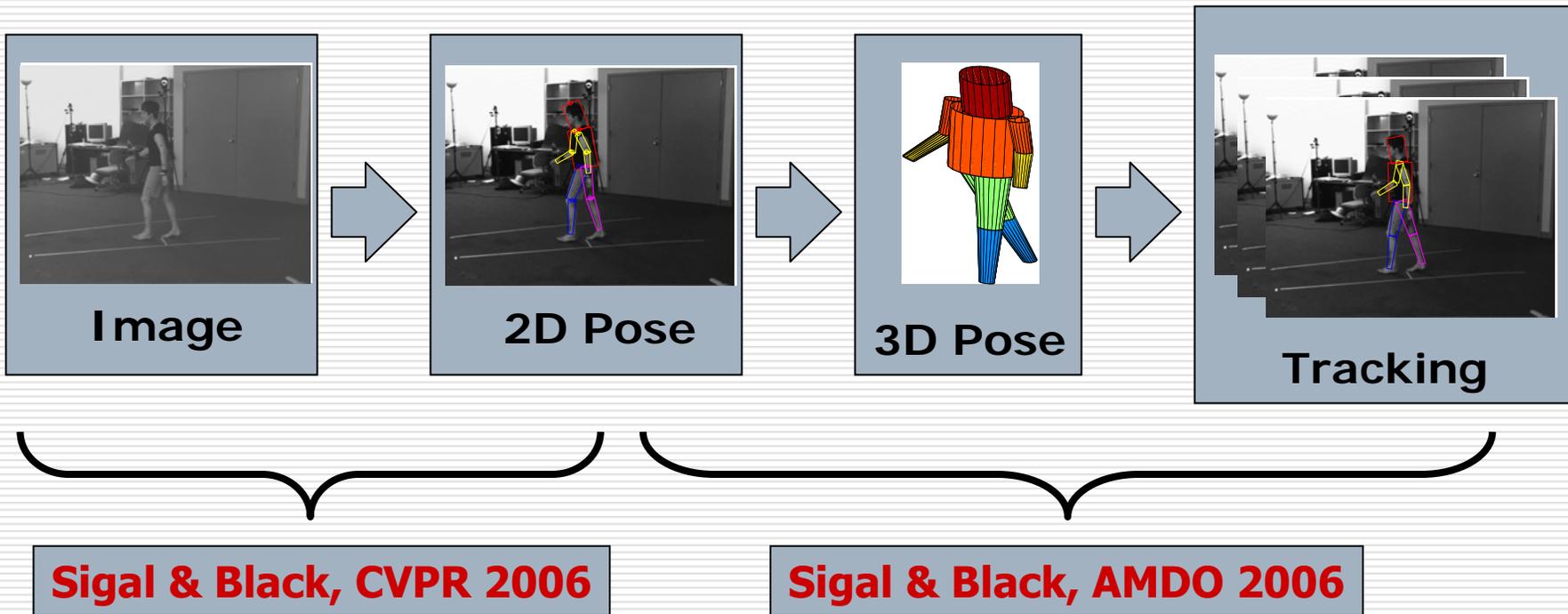
■ Belief Propagation





Hierarchical Inference Framework

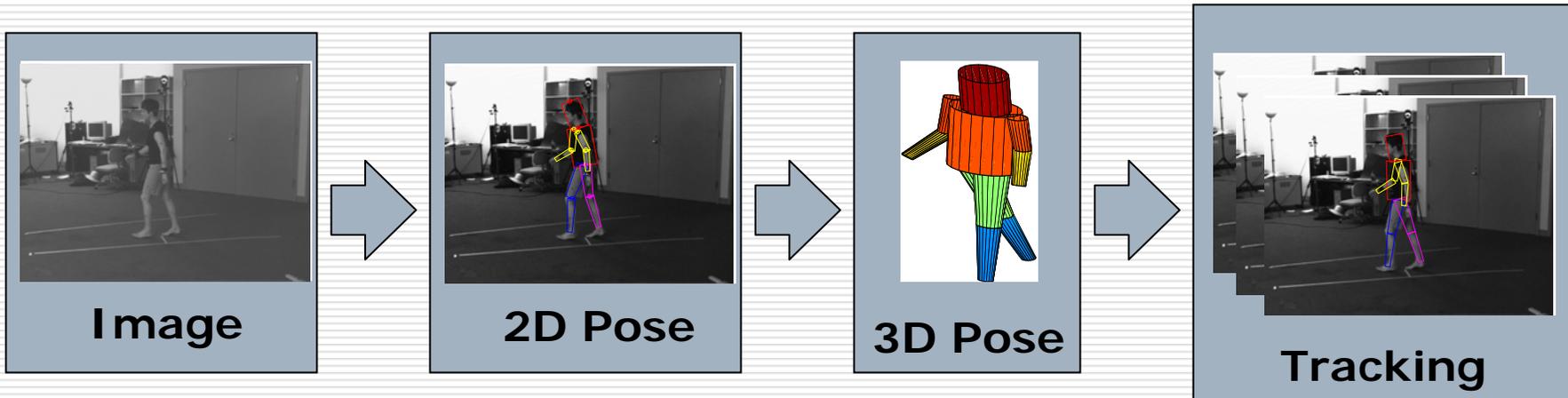
- Break up a very hard problem into smaller manageable pieces



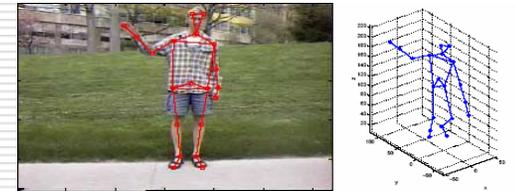


Hierarchical Inference Framework

- Break up a very hard problem into smaller manageable pieces



- We are able to infer the 3D pose from a single image
- But, are still able to make use of temporal consistency when it is available



Howe, Leventon, Freeman, '00



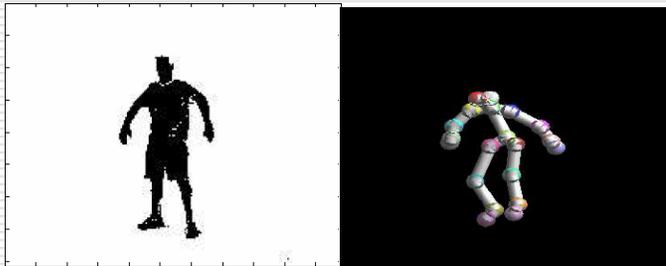
Discriminative Approaches



Sminchisescu, Kanaujia, Li, Metaxas, '05



Agarwal & Triggs, '04



Rosales & Sclaroff, '00

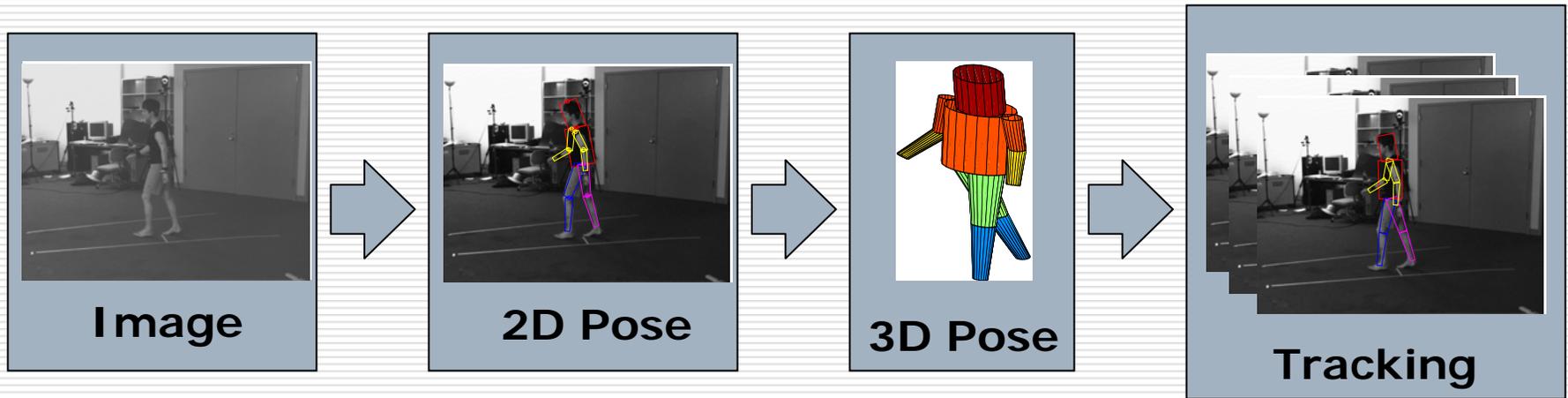


Shakhnarovich, Viola, Darrell, '03

- ❑ Tend to be very fast
- ❑ Work well on the data they are trained on
- ❑ **Generalize poorly to data they have never seen**



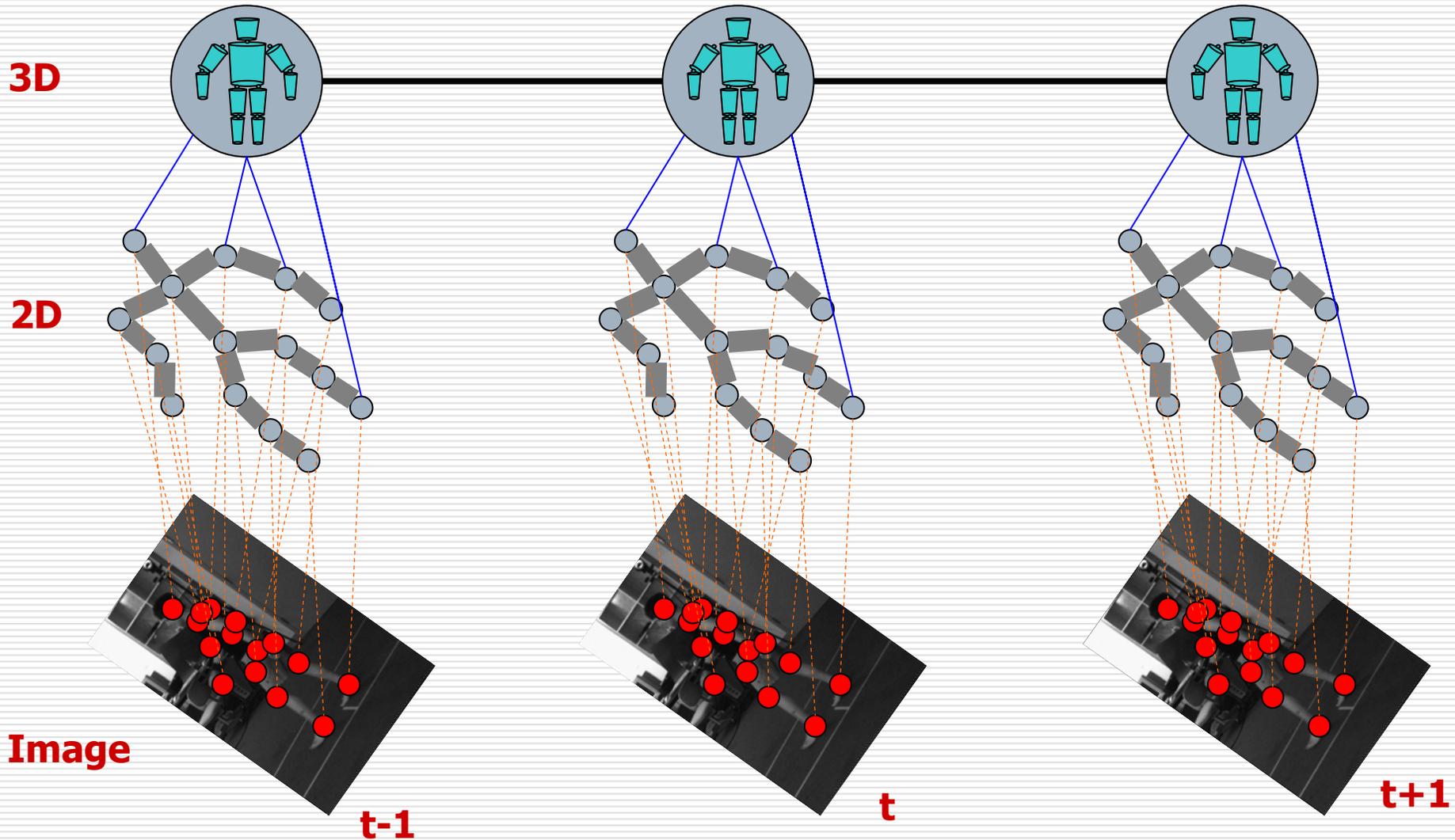
Advantage of Hierarchical Inference



- **Better generalization in situations where good features are unavailable (lack of good silhouettes)**
 - via the use intermediate Generative 2D pose estimation
- **Modularity**
 - Can easily substitute different 2D pose estimation modules
- **Fully probabilistic approach**

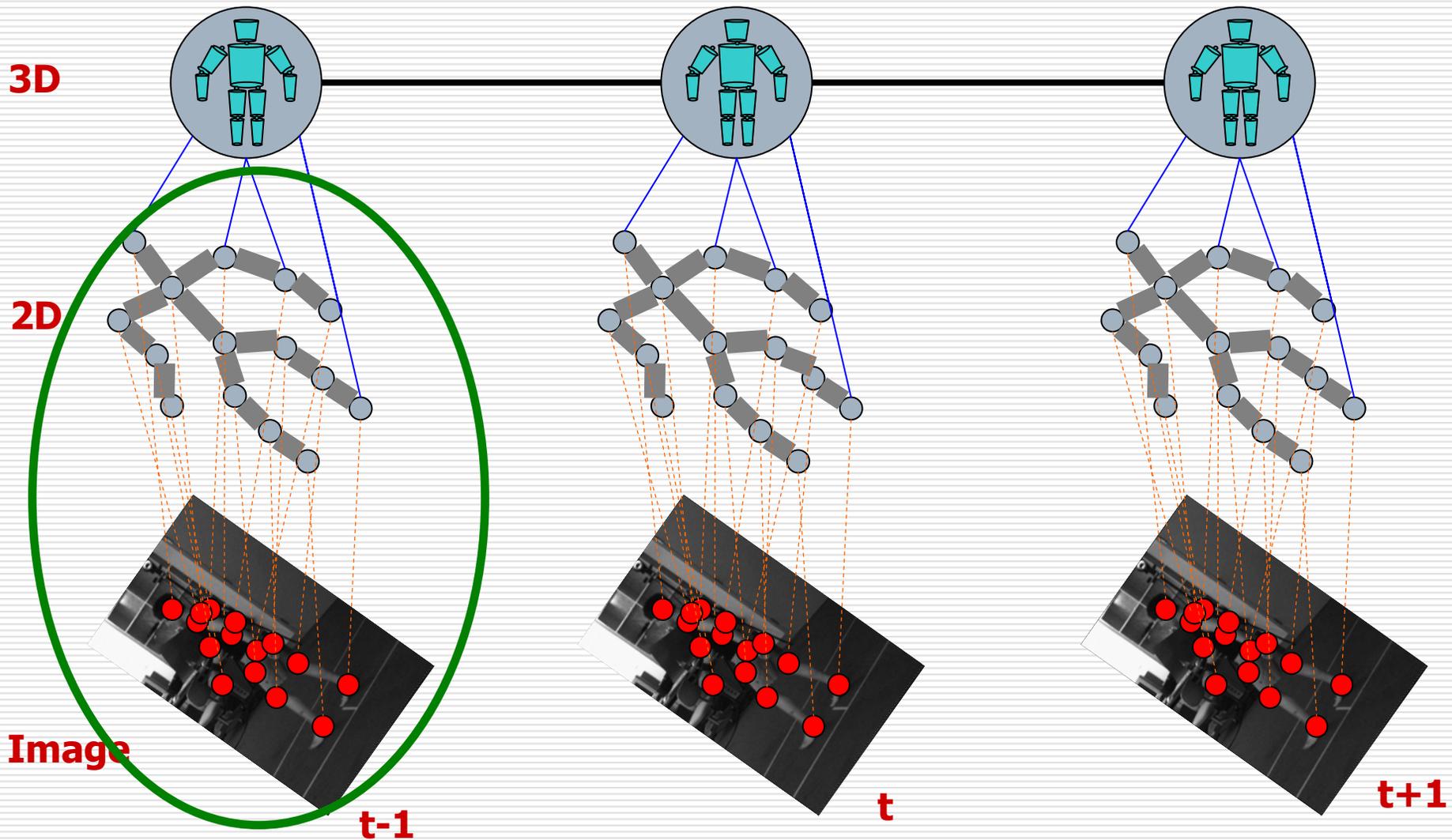


Hierarchical Graphical Model Structure



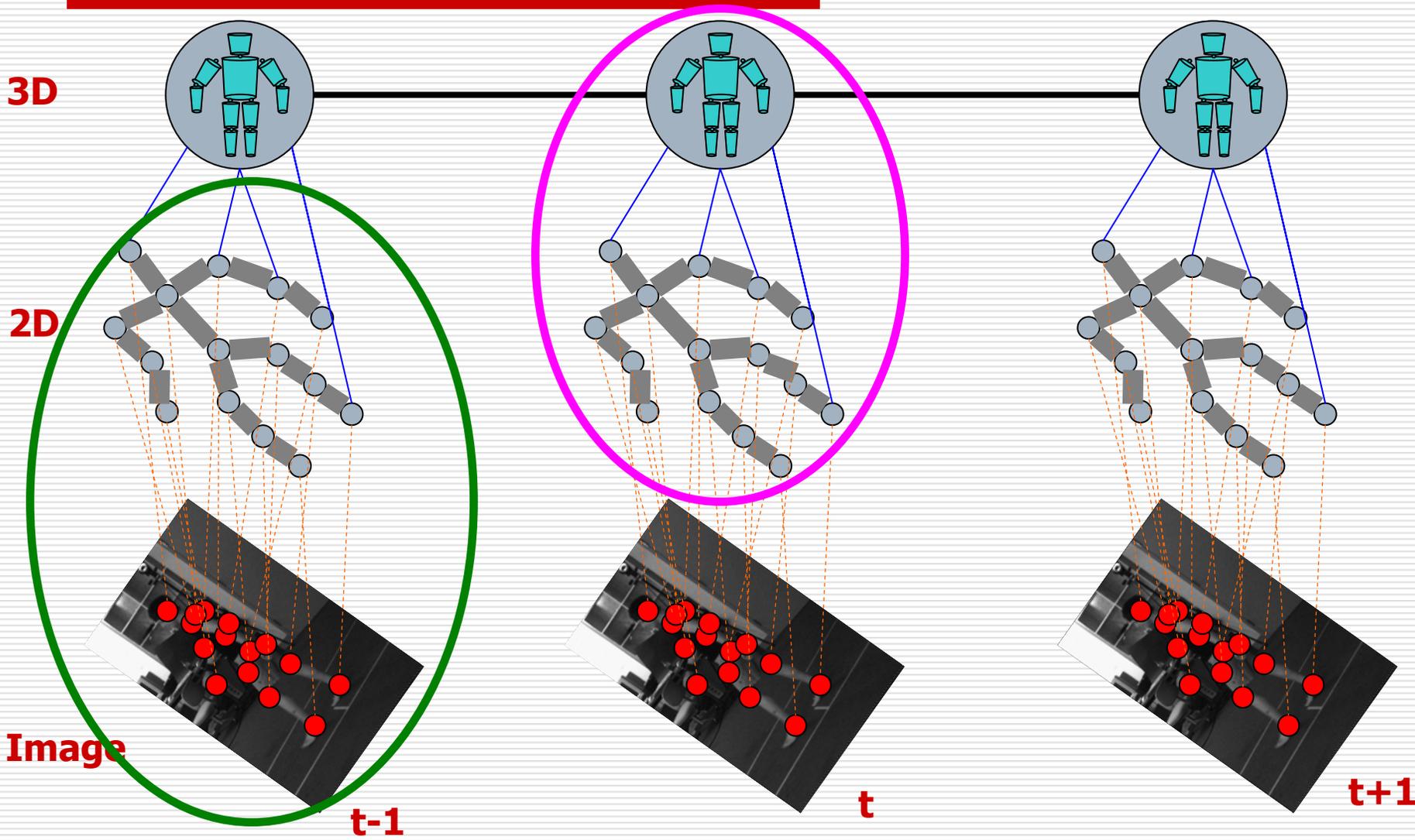


Hierarchical Graphical Model Structure



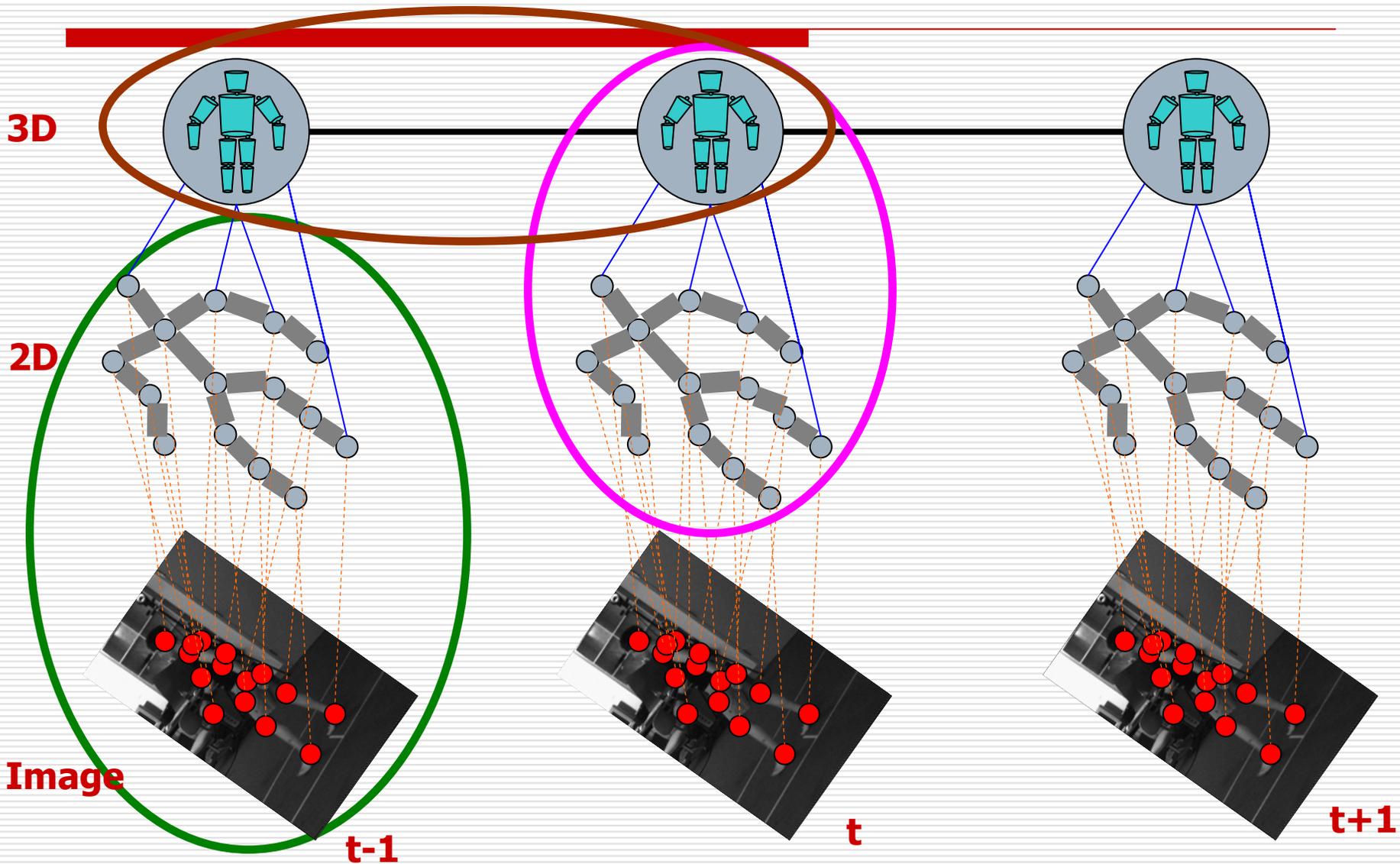


Hierarchical Graphical Model Structure



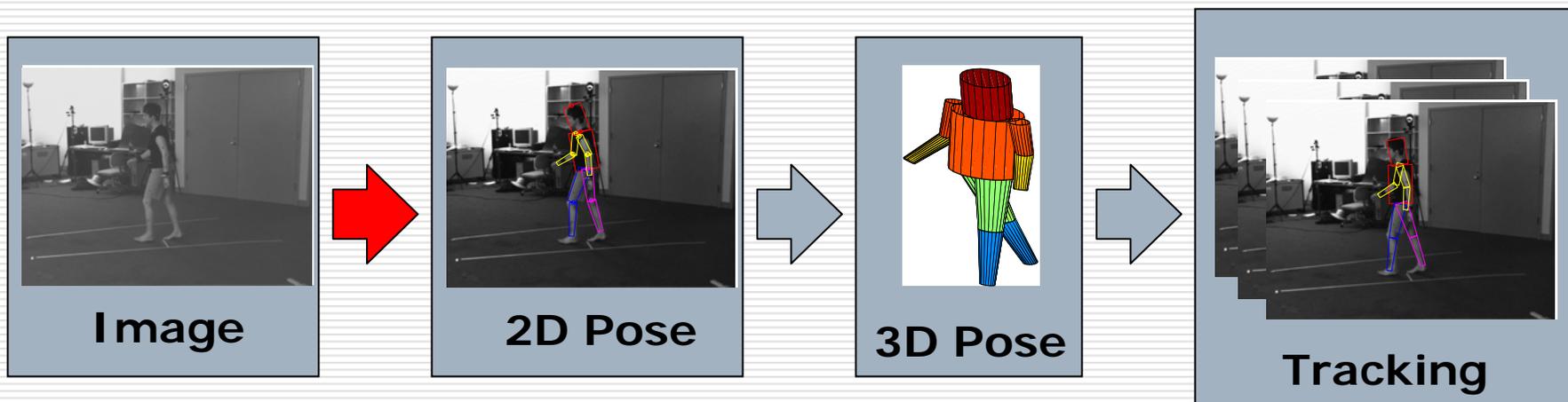


Hierarchical Graphical Model Structure





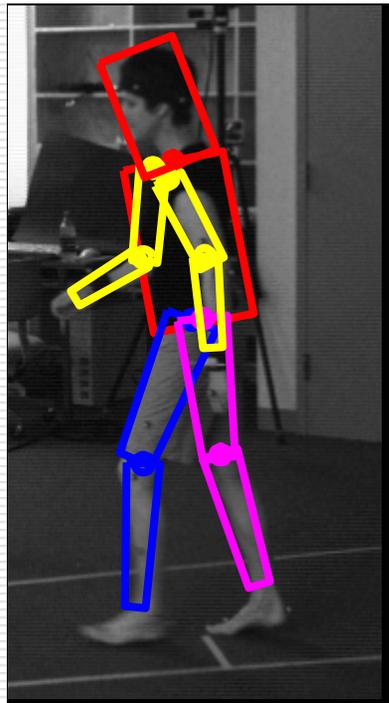
Inferring 2D pose



⎵
**Occlusion-sensitive
“Loose-limbed” body
model**

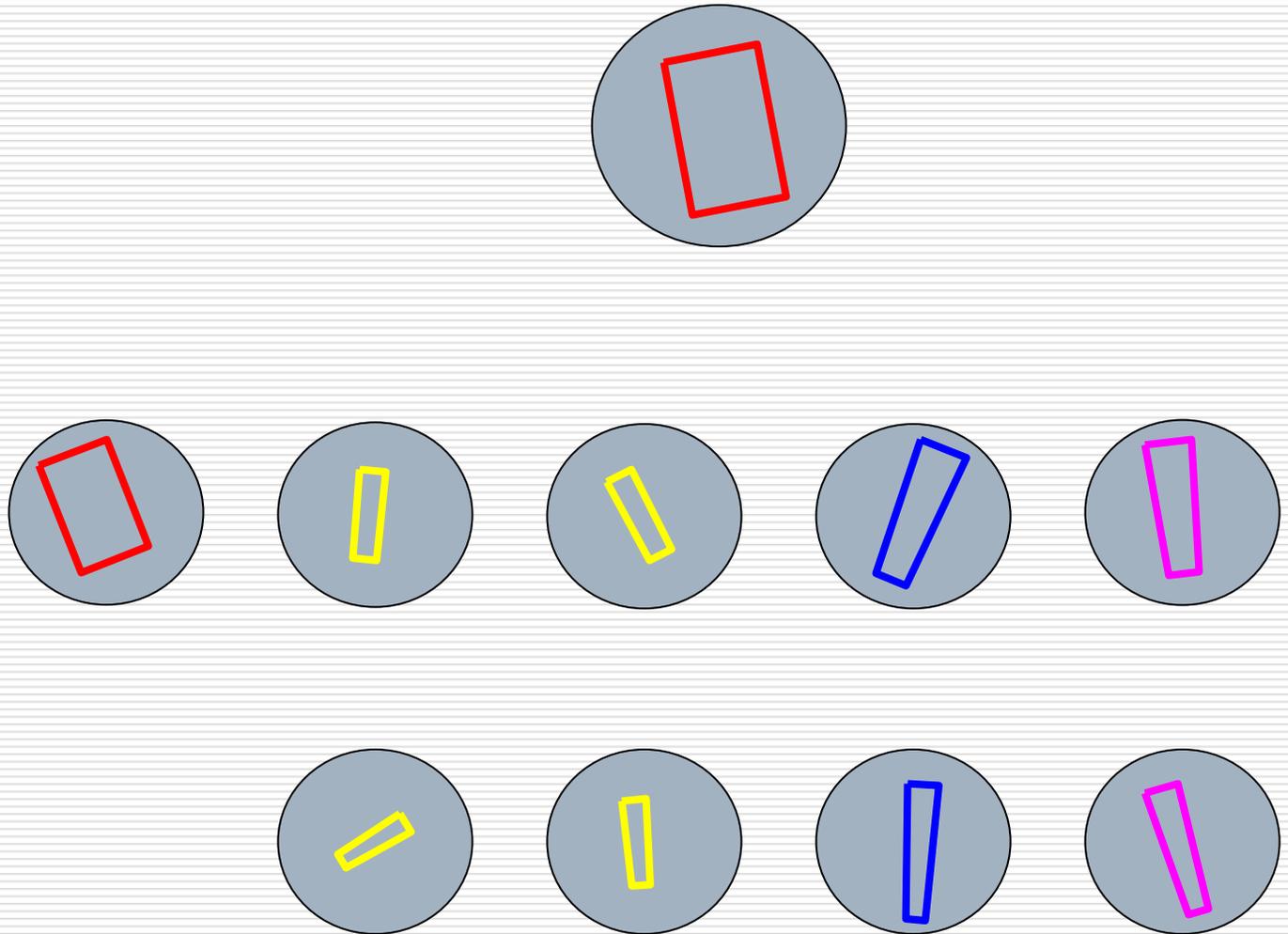
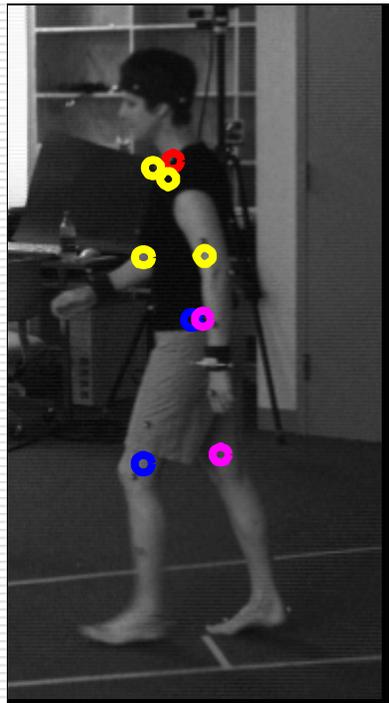


2D "Loose-limbed" Body Model



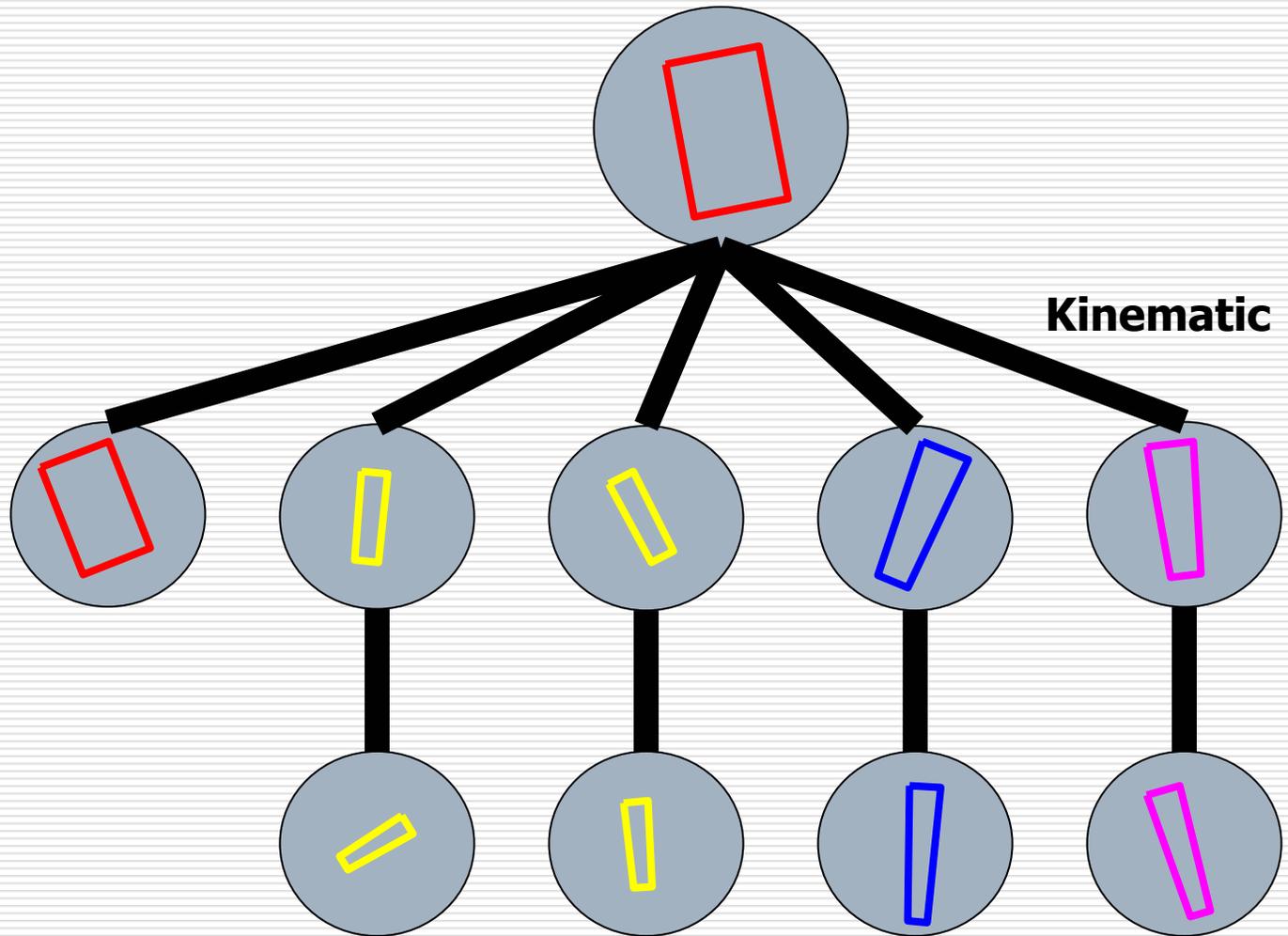
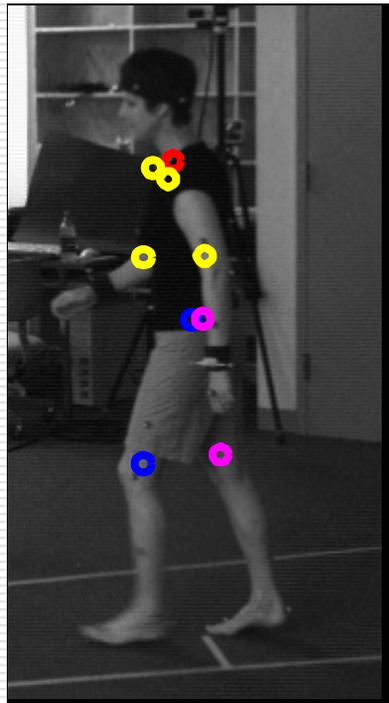


2D "Loose-limbed" Body Model



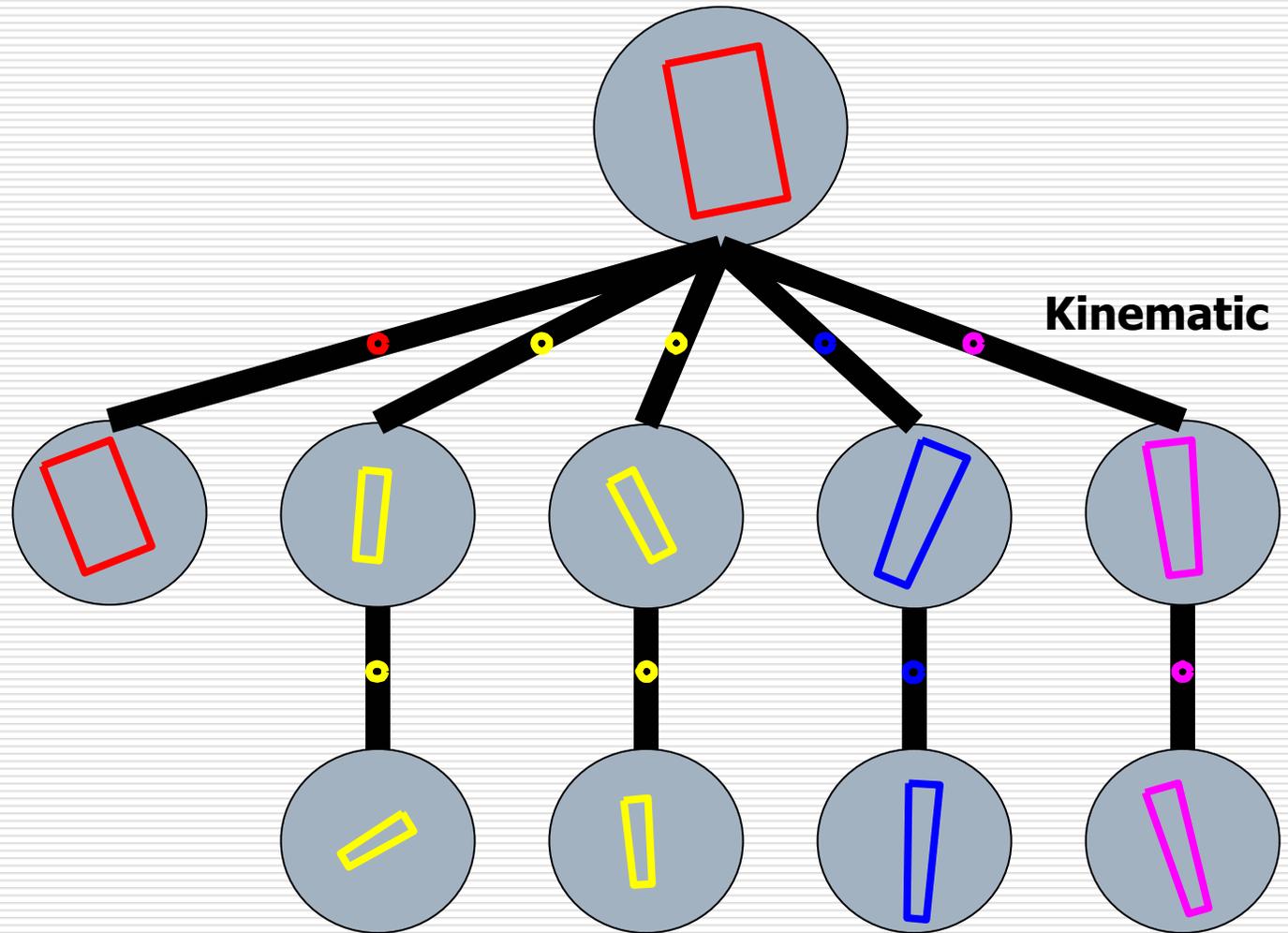


2D "Loose-limbed" Body Model



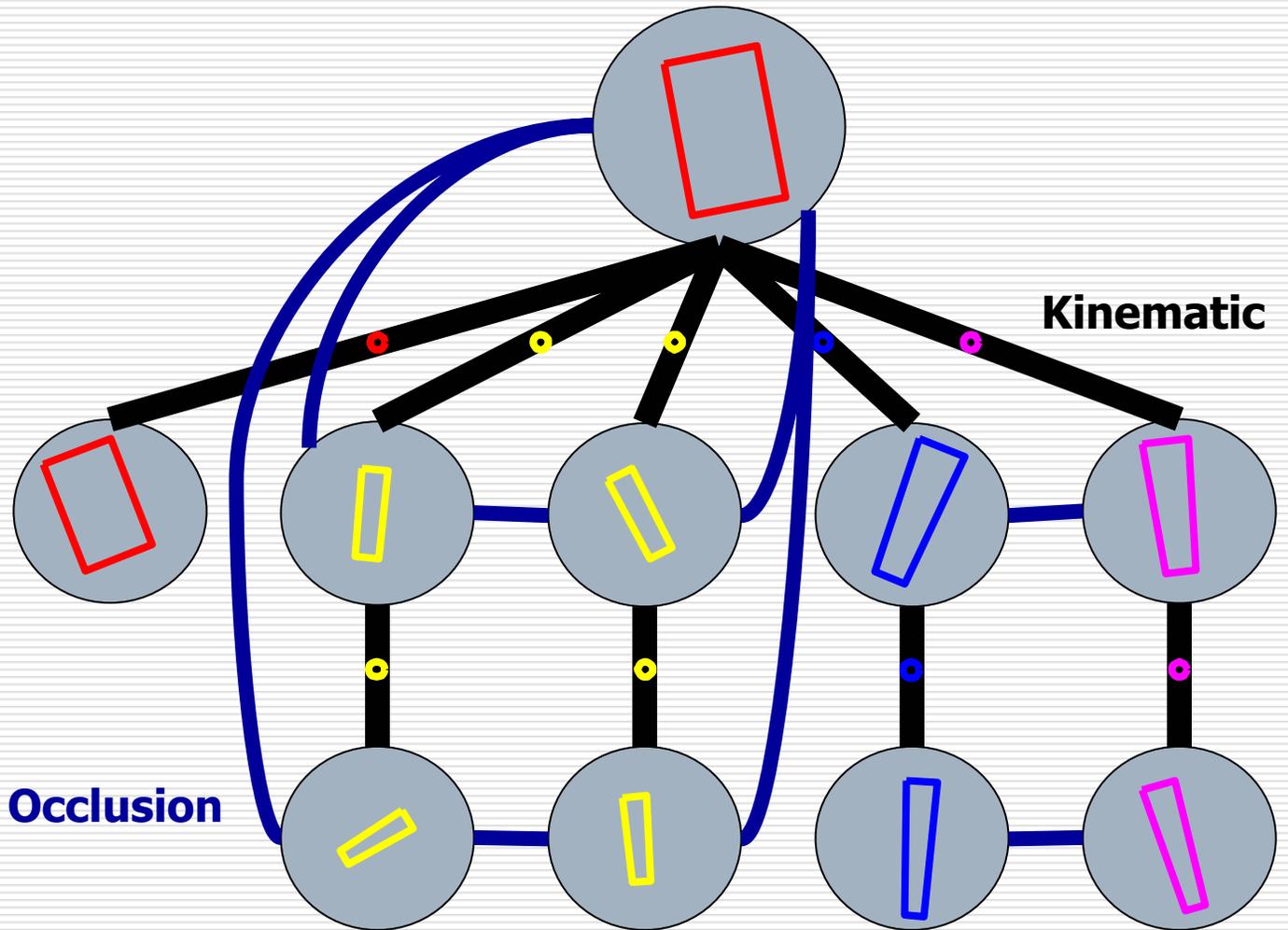


2D "Loose-limbed" Body Model



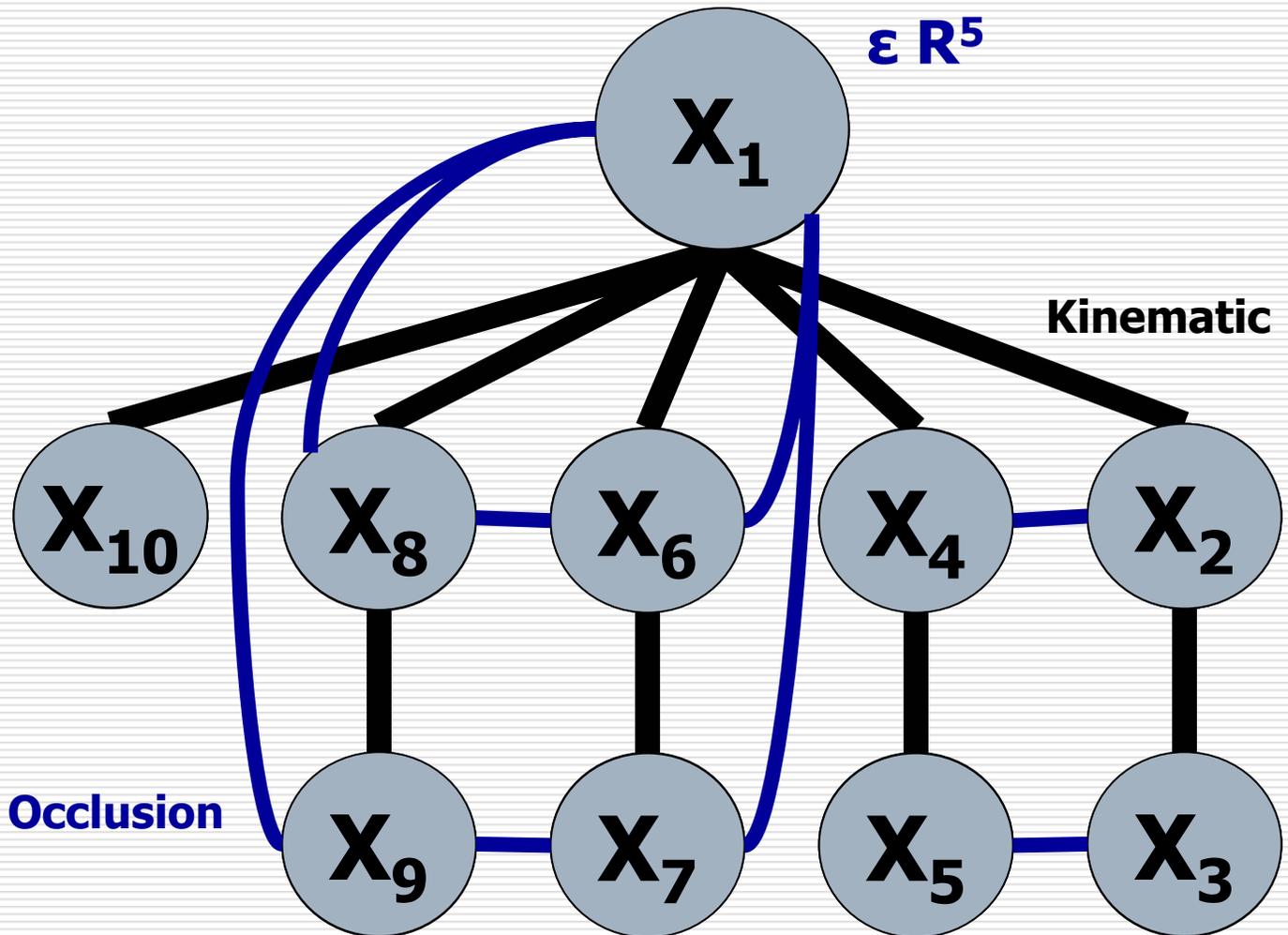
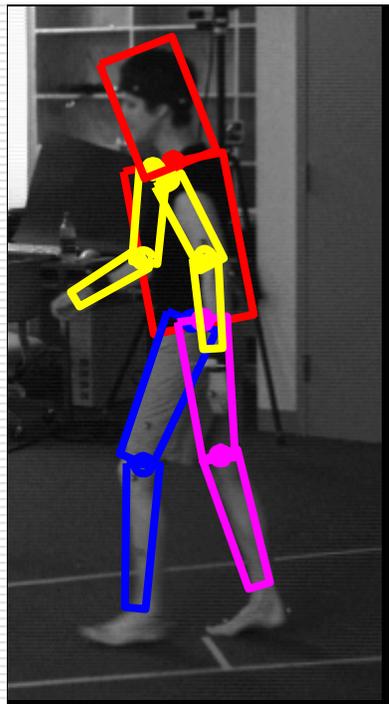


2D "Loose-limbed" Body Model





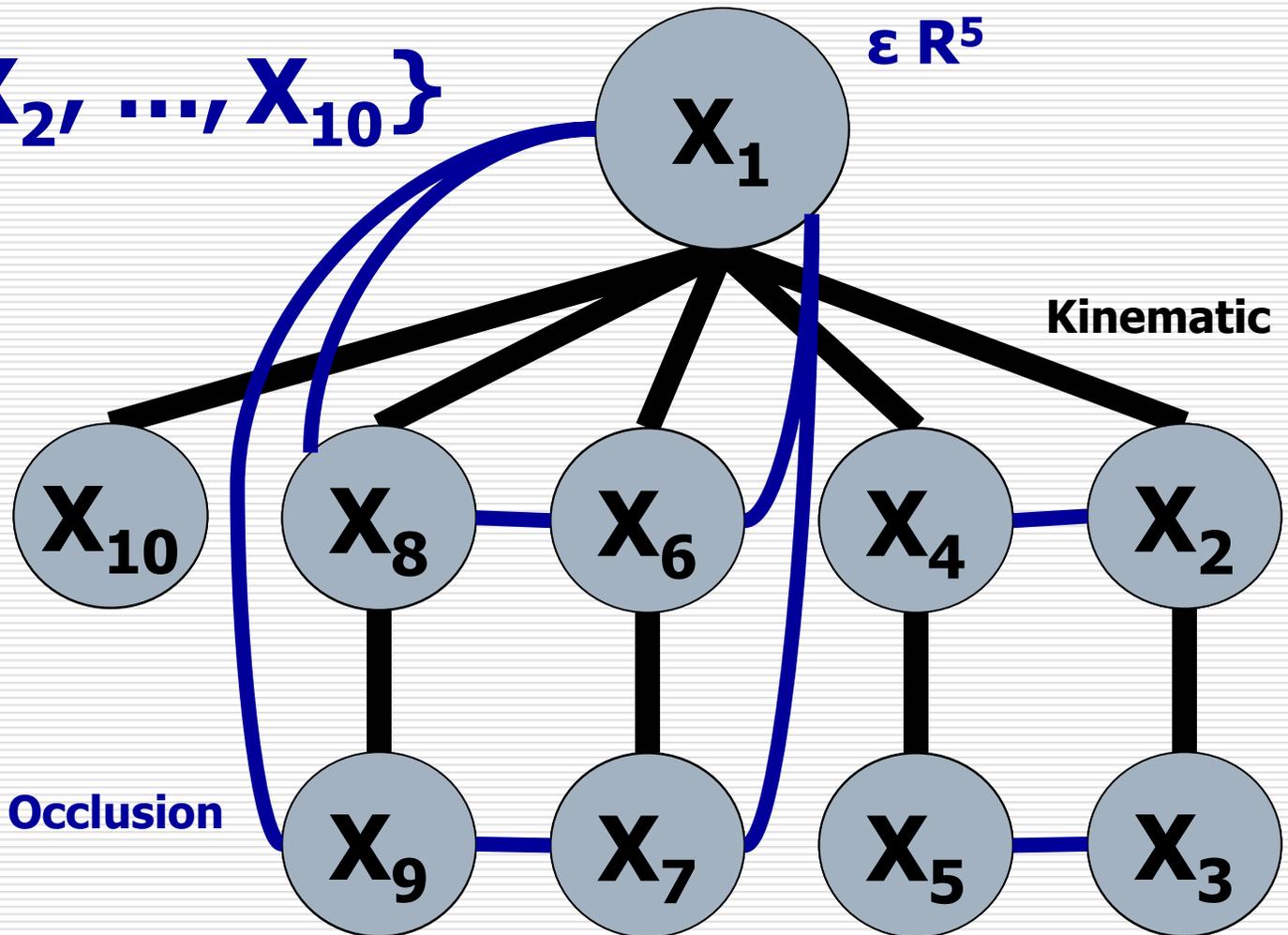
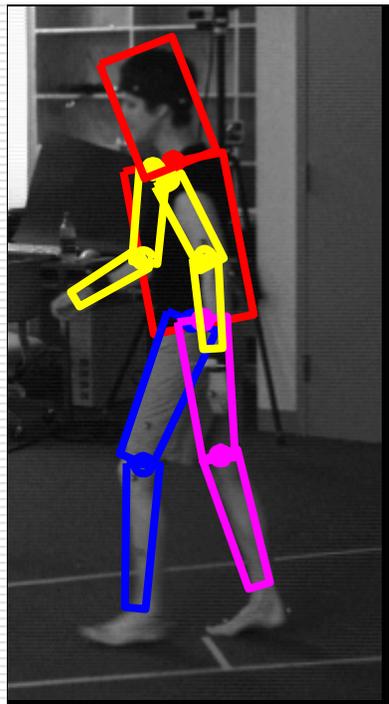
2D "Loose-limbed" Body Model





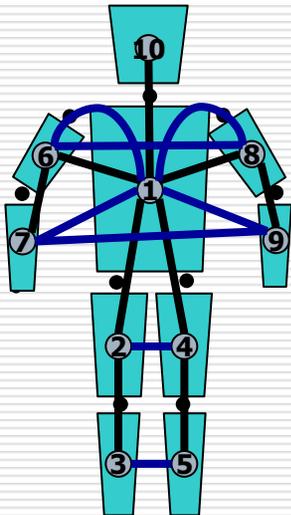
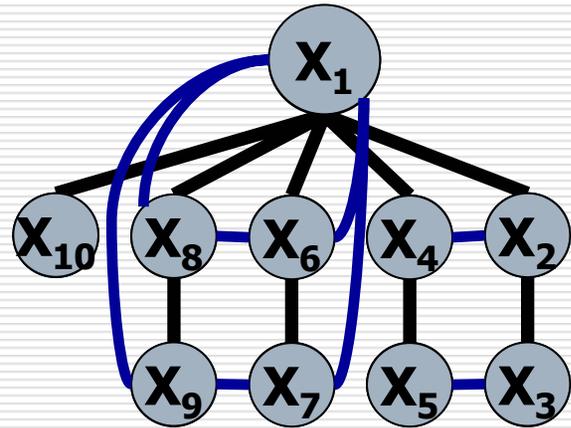
2D "Loose-limbed" Body Model

$$\mathbf{X} = \{ \mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_{10} \}$$





2D "Loose-limbed" Body Model

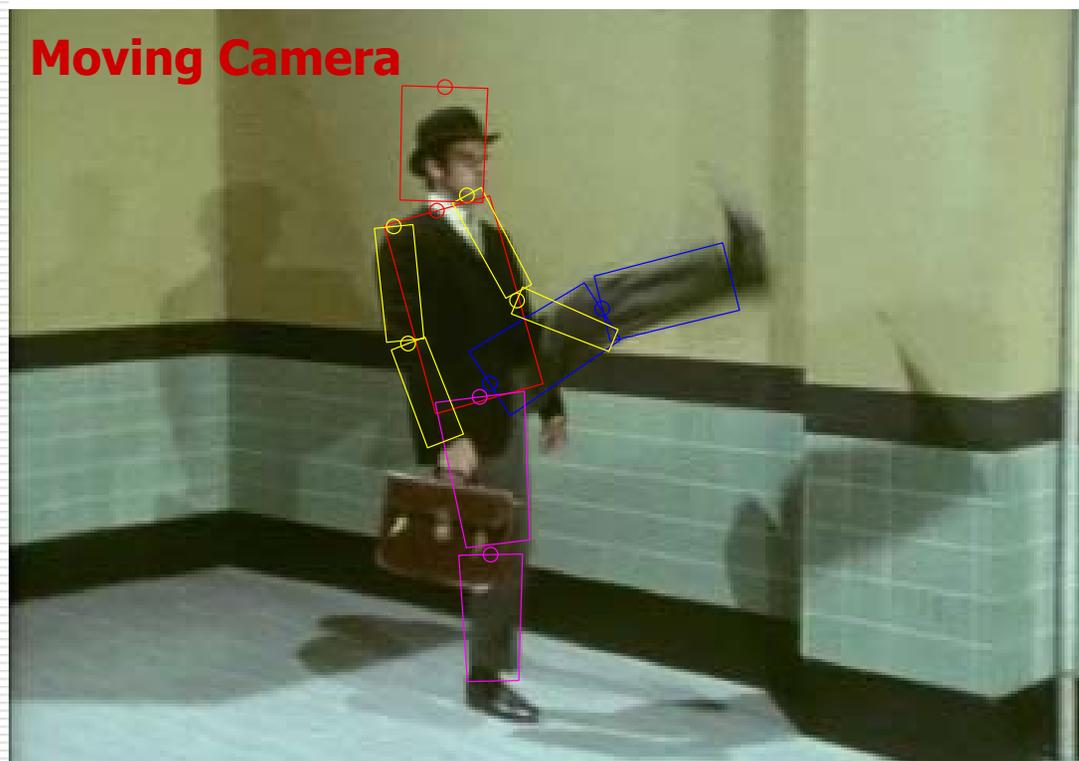


- **Exact inference in tree-structured graphical models can be computed using BP**
- **But, not when**
 - State-space is continuous
 - Likelihoods (or potentials) are not Gaussian
 - Graph contains loops
- **This forces the use of approximate inference algorithms**
 - **PAMPAS:** M. Isard, '03
 - **Non-Parametric BP:** E. Sudderth, A. Ihler, W. Freeman, A. Willsky, '03



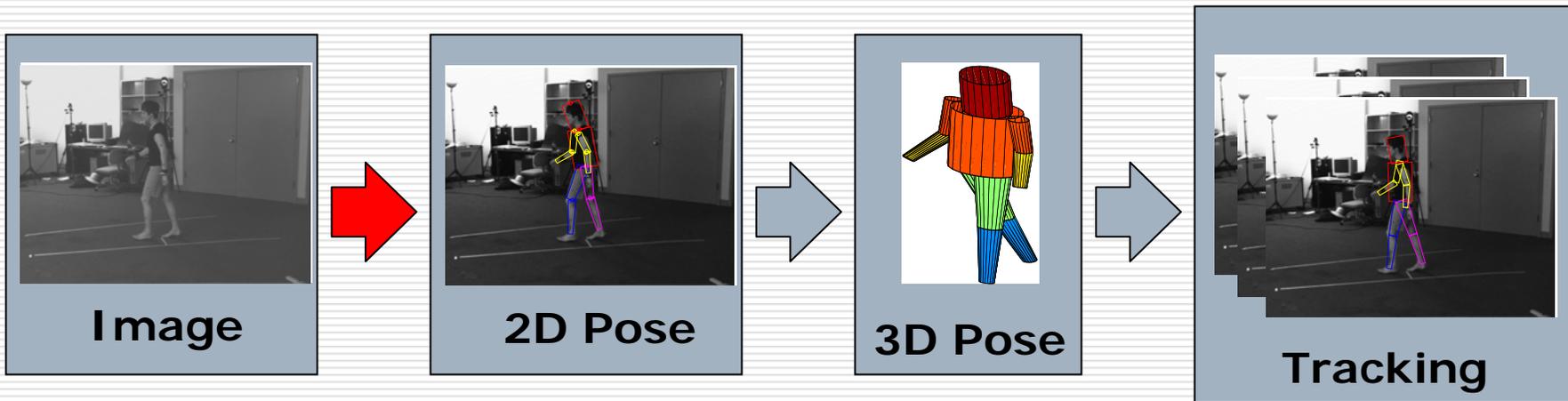
Inferring 2D pose

- ❑ **Occlusion-sensitive “Loose-limbed” body model allows us to infer the 2D pose reliably**
- ❑ **Even when motions are complex**





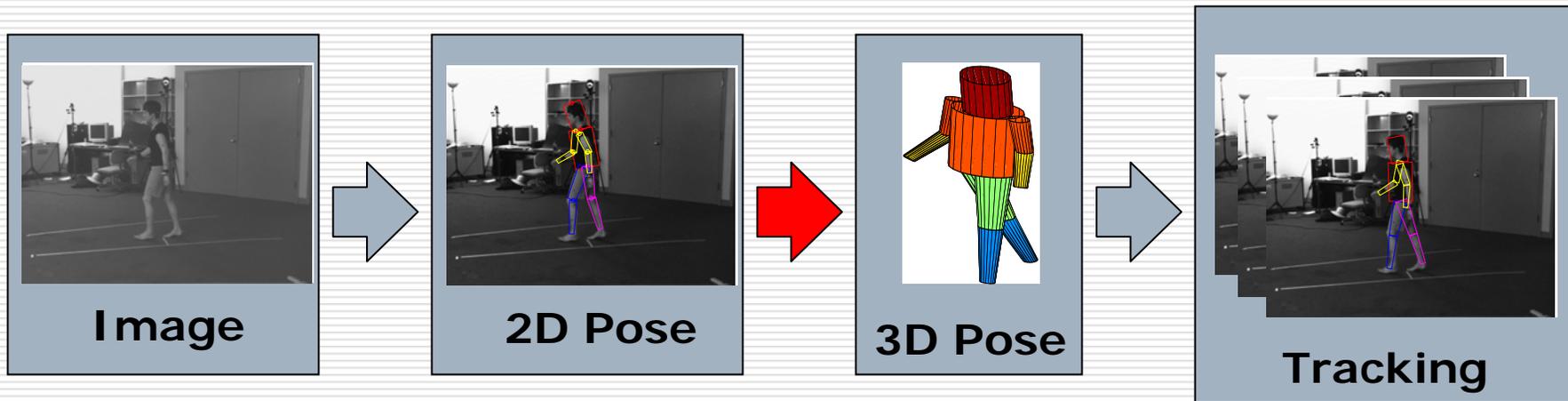
Summary so far ...



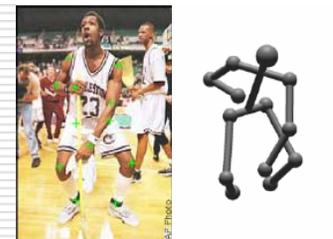
⎵
**Occlusion-sensitive
“Loose-limbed” body
model**



Inferring 3D pose from 2D pose



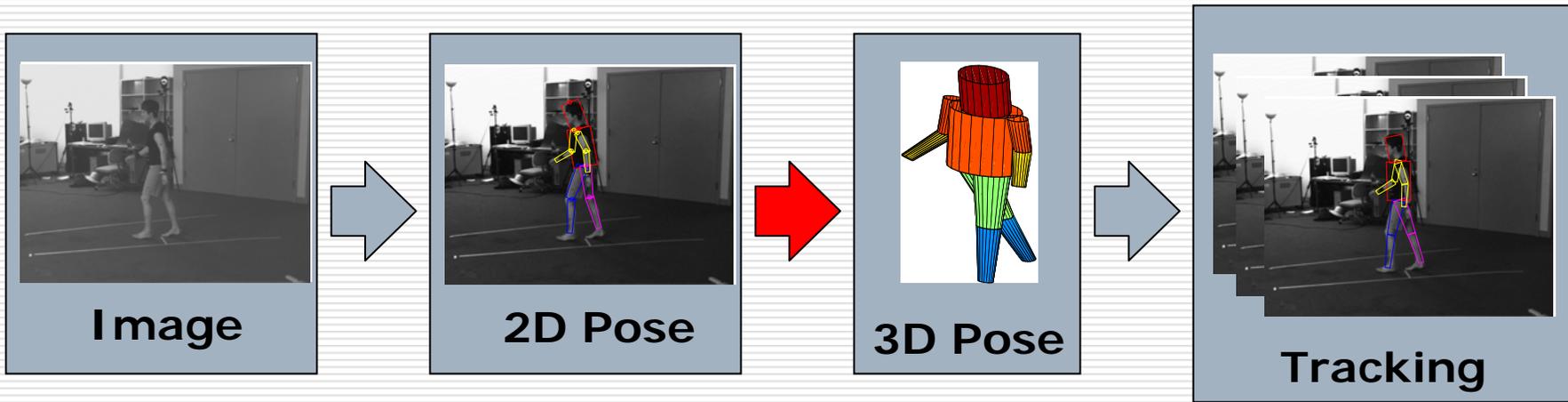
- **We obtain estimates for the joints automatically**
- **We learn direct probabilistic mapping**



Camillo J. Taylor, '00



Inferring 3D pose from 2D pose



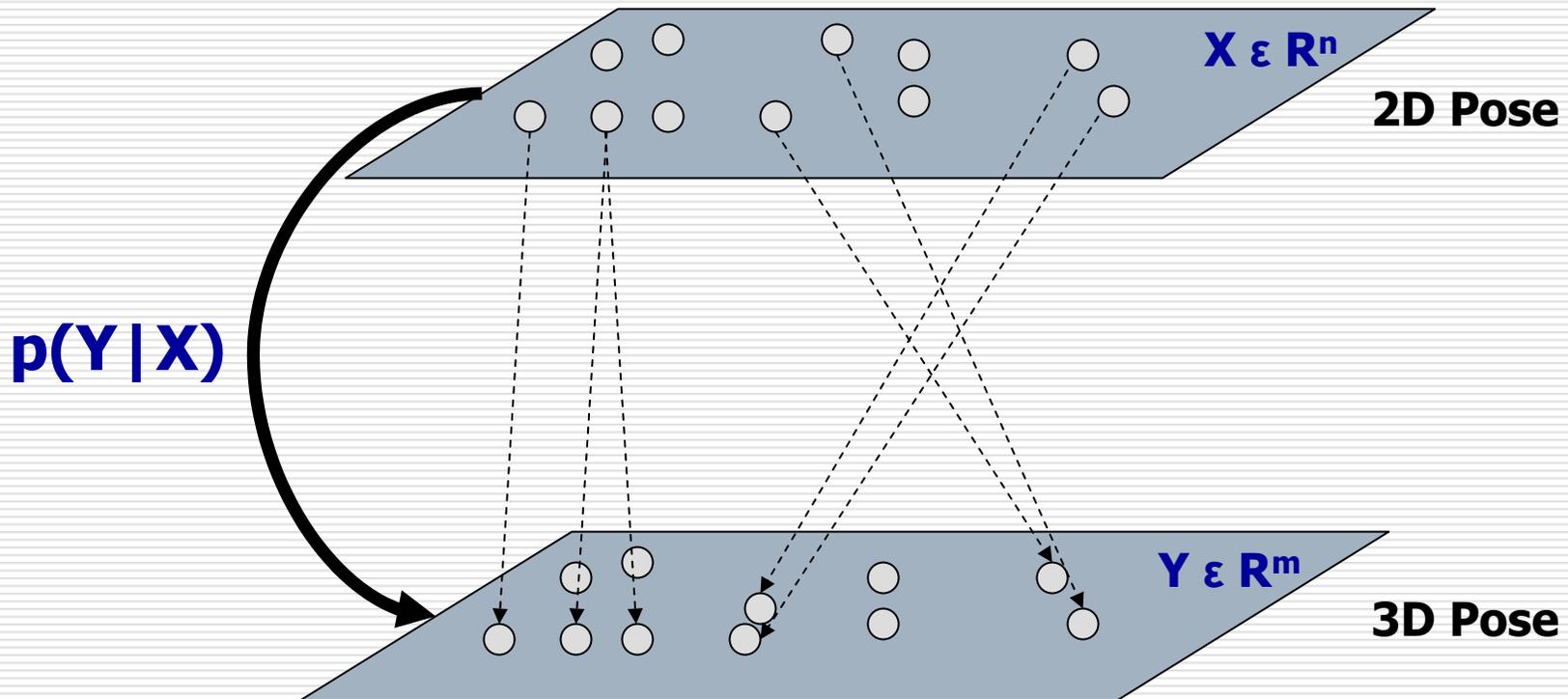
Mixture of Experts (MoE)

Sminchisescu et al, '05



Inferring 3D pose from 2D pose

We want to estimate a distribution/mapping $p(\text{3D Pose} | \text{2D Pose})$

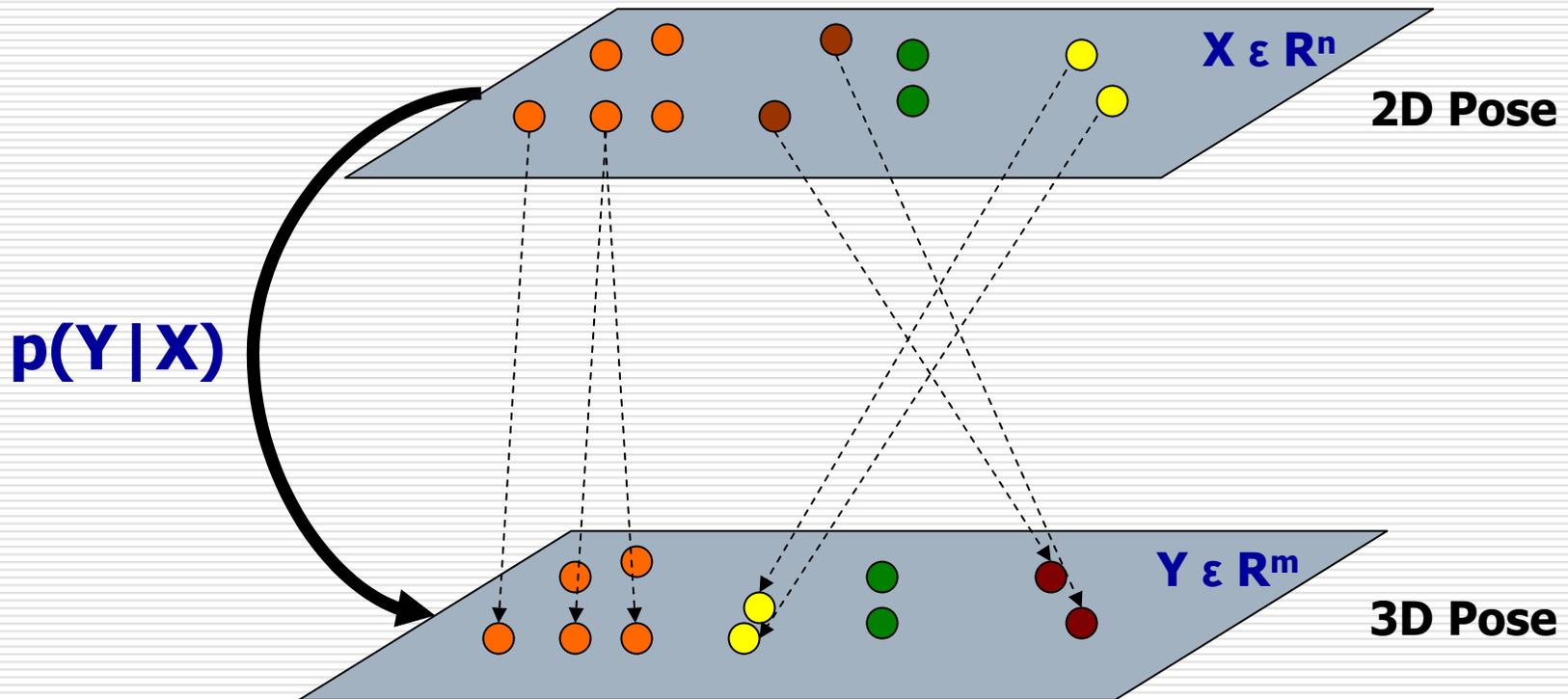


Problem: $p(Y|X)$ is non-linear mapping, and not one-to-one



Mixture of Experts (MoE)

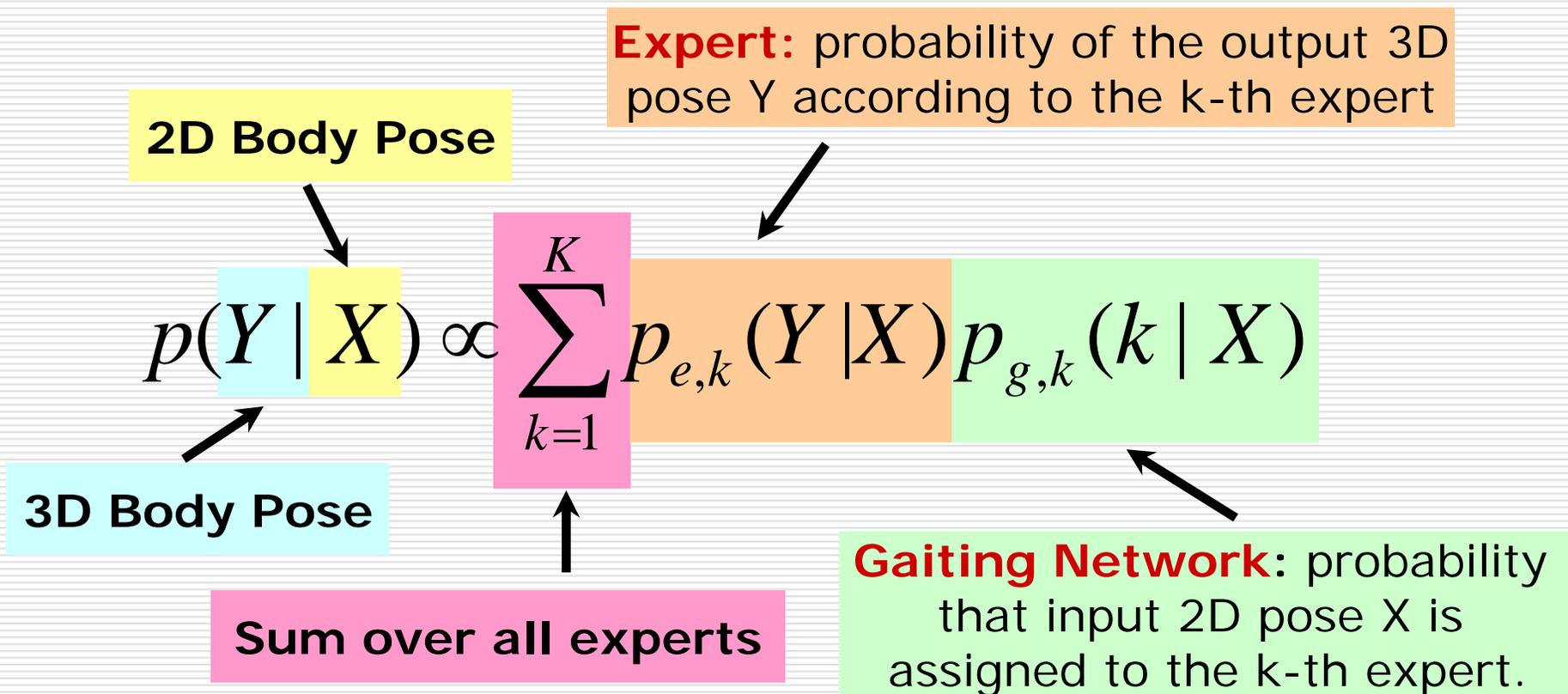
We want to estimate a distribution/mapping $p(\text{3D Pose} | \text{2D Pose})$



Solution: $p(Y | X)$ may be approximated by a locally linear mappings (experts)



MoE Formally



- **Training of MoE is done using EM procedure (similar to learning Mixture of Gaussians)**



Illustration of 3D pose inference

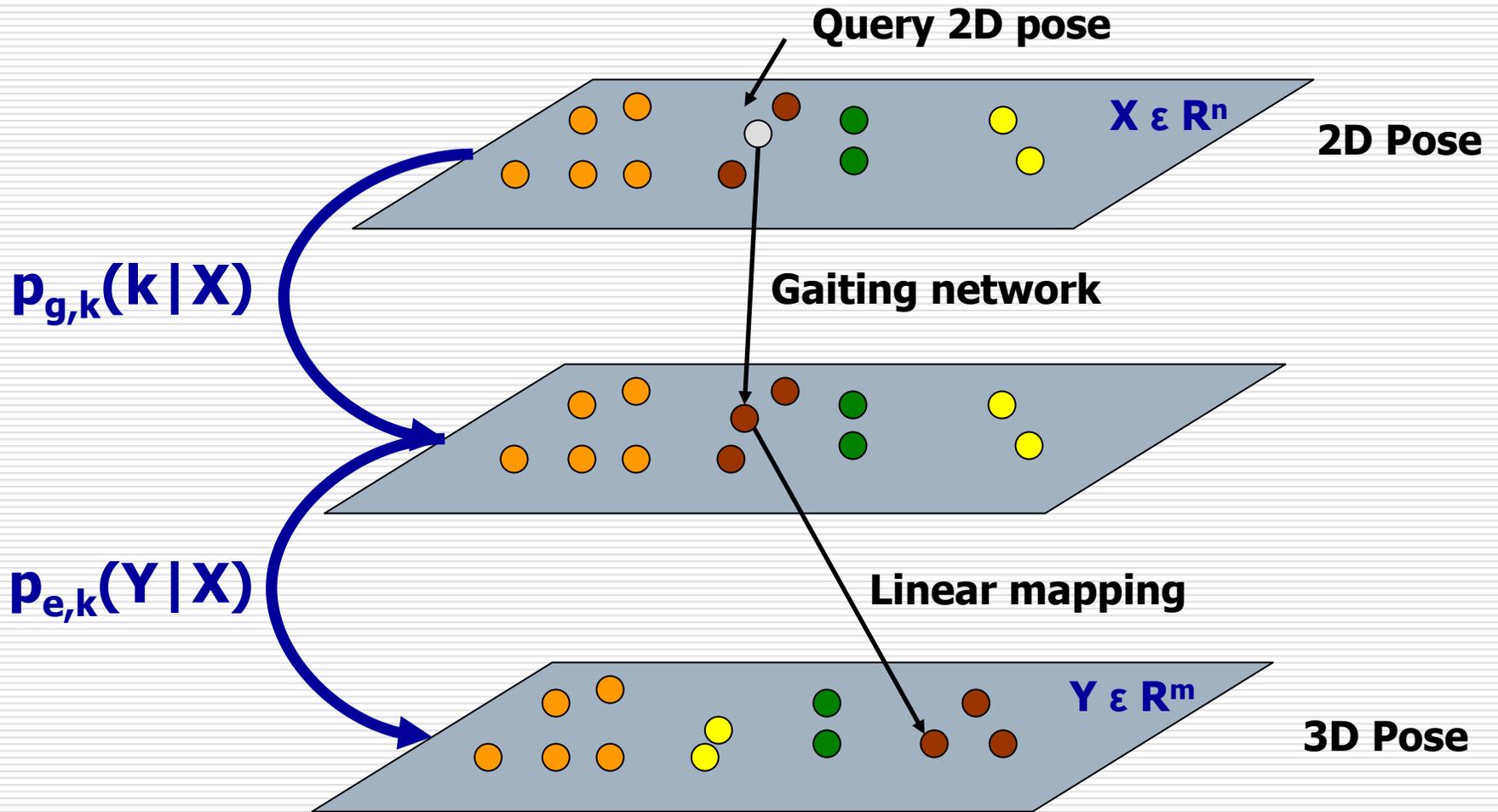
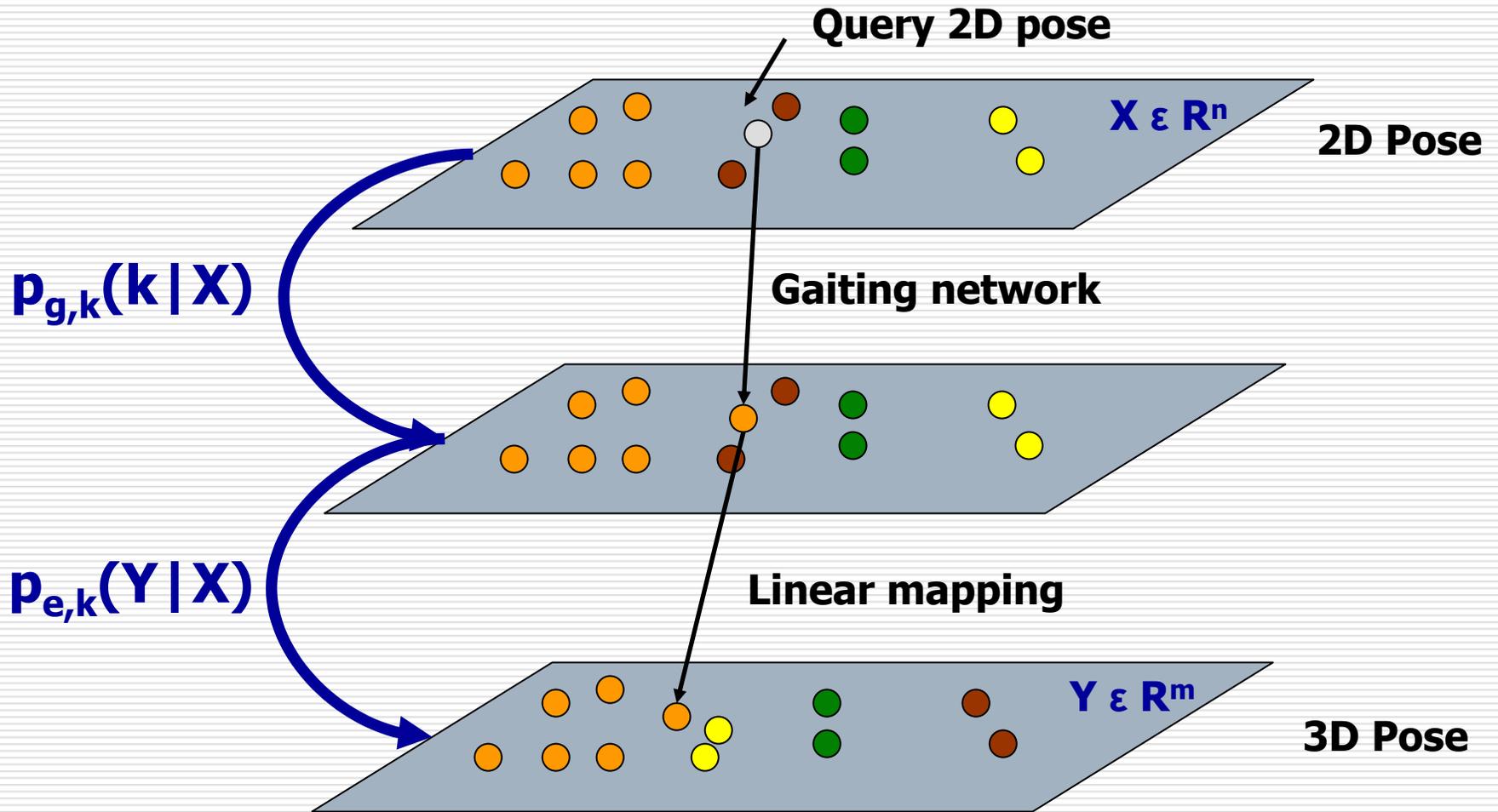




Illustration of 3D pose inference





Performance

□ Two action-specific MoE models are trained

Walking

- 4587 2D/3D MOCAP pose pairs for training
- 1398 video frames used for testing

Performance

- **View only:** 14 mm
- **Pose only:** 23 mm
- **Overall:** 30 mm

Dancing

- 4151 2D/3D MOCAP pose pairs for training
- 2074 video frames used for testing

Performance

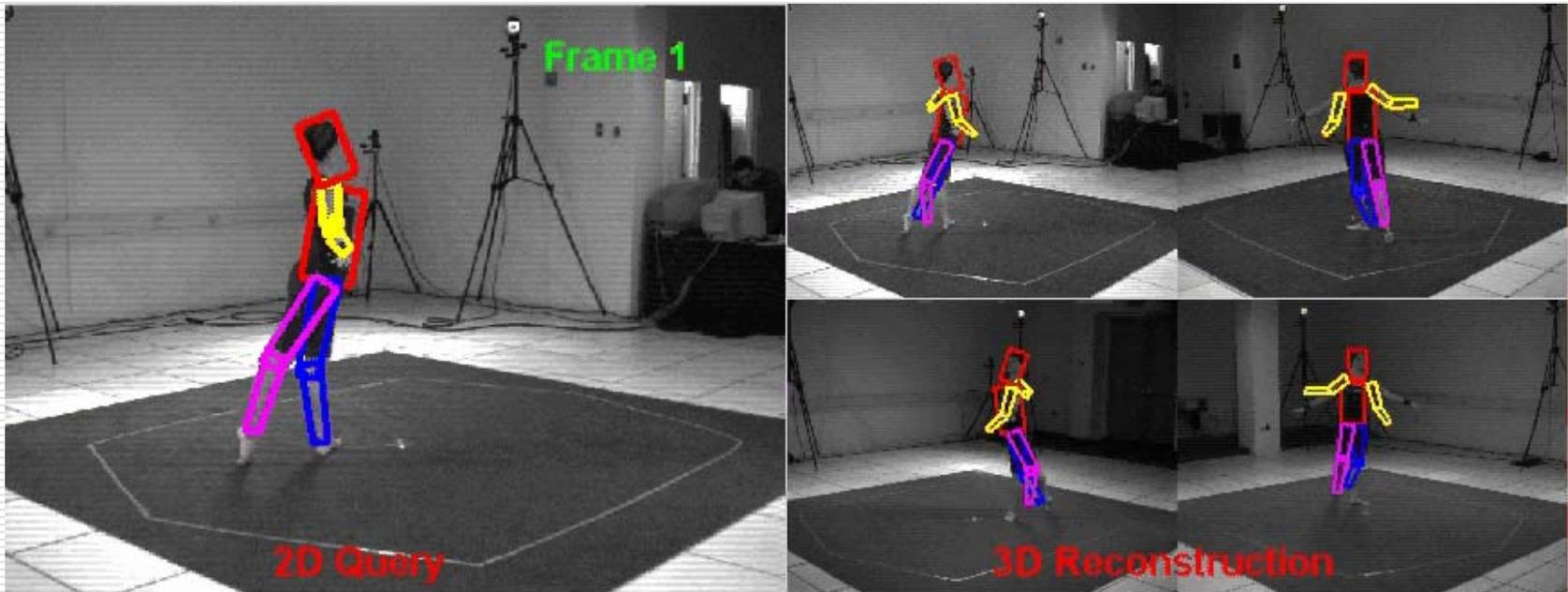
- **View only:** 22 mm
- **Pose only:** 59 mm
- **Overall:** 64 mm



Structured motion / Performance

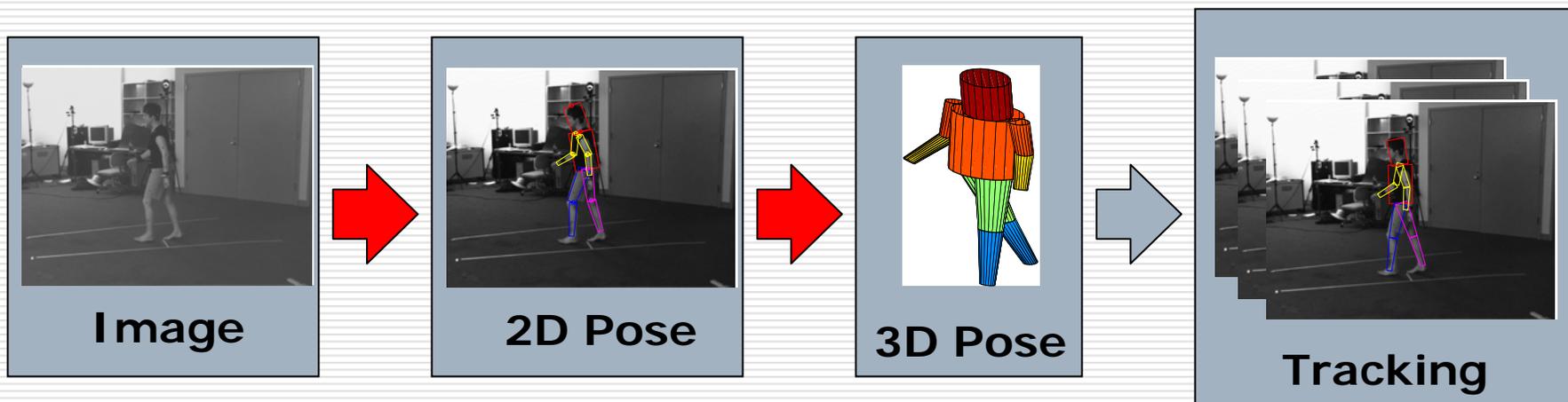


How well does MoE model work?





Summary so far ...

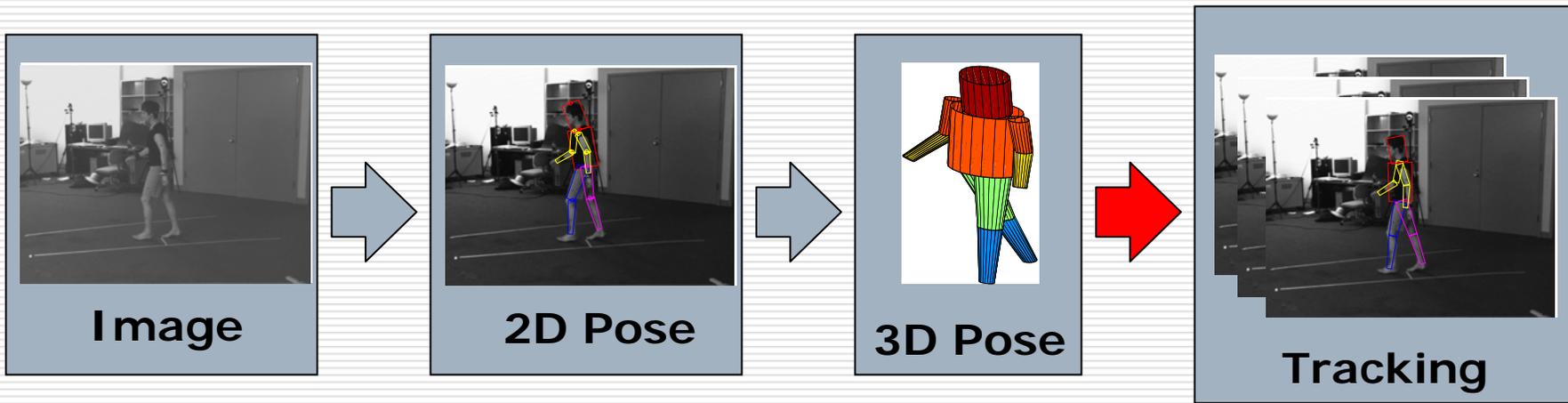


Occlusion-sensitive
"Loose-limbed" body
model

Mixture of
Experts (MoE)



Summary so far ...

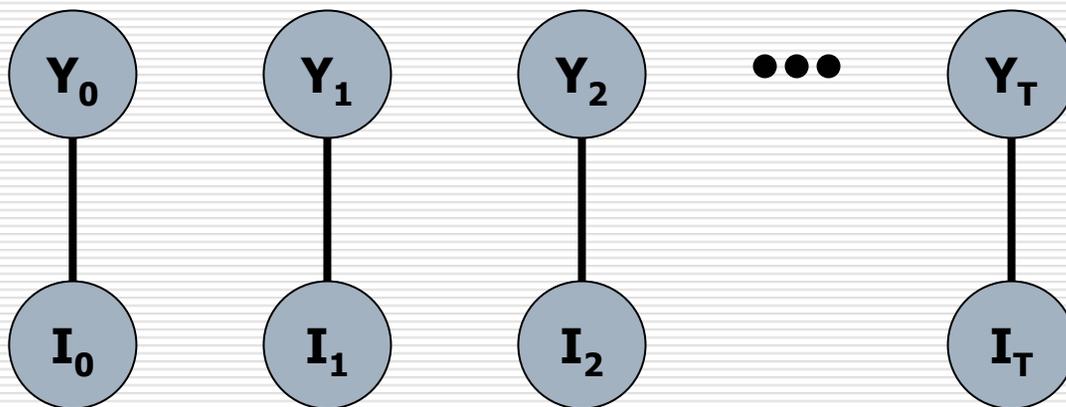


Hidden Markov Model (HMM)



Tracking in 3D

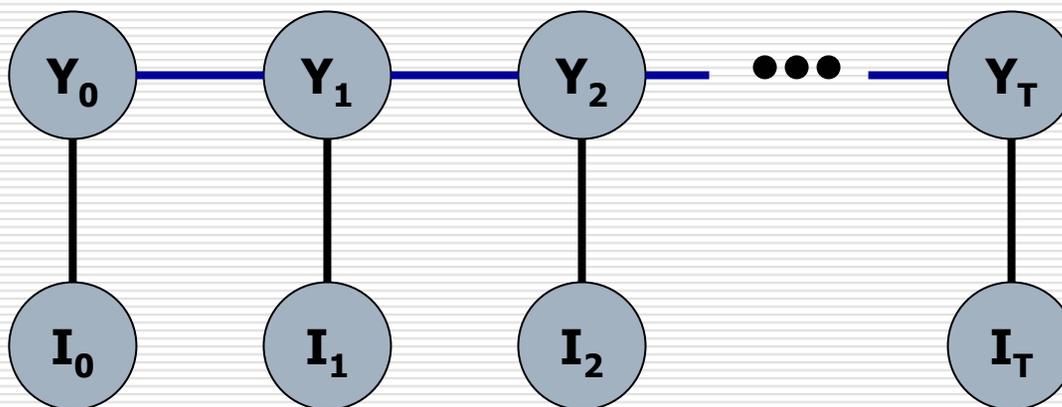
- We have a distribution over 3D poses at every time instance





Tracking in 3D

- We have a distribution over 3D poses at every time instance

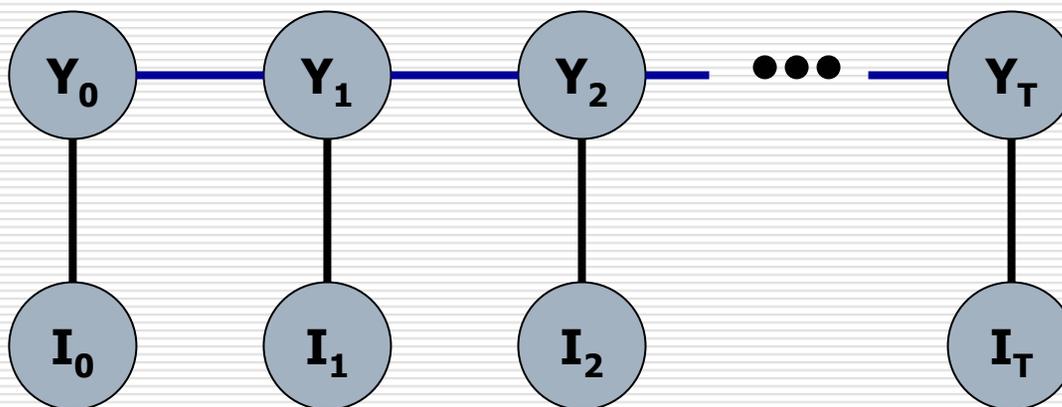


- Assuming that 3D pose at time t is conditionally independent of the state at time $t-2$ given state at $t-1$



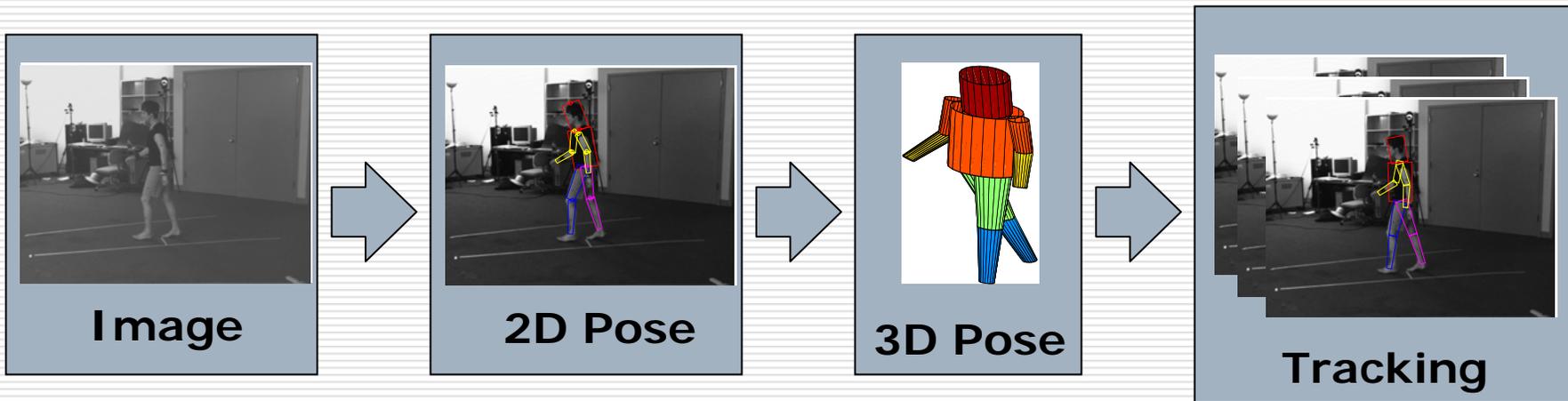
Tracking in 3D (inference)

- **Inference in this graphical model can be done using the tools we already have**
 - PAMPAS/Non-parametric belief propagation





Summary so far ...



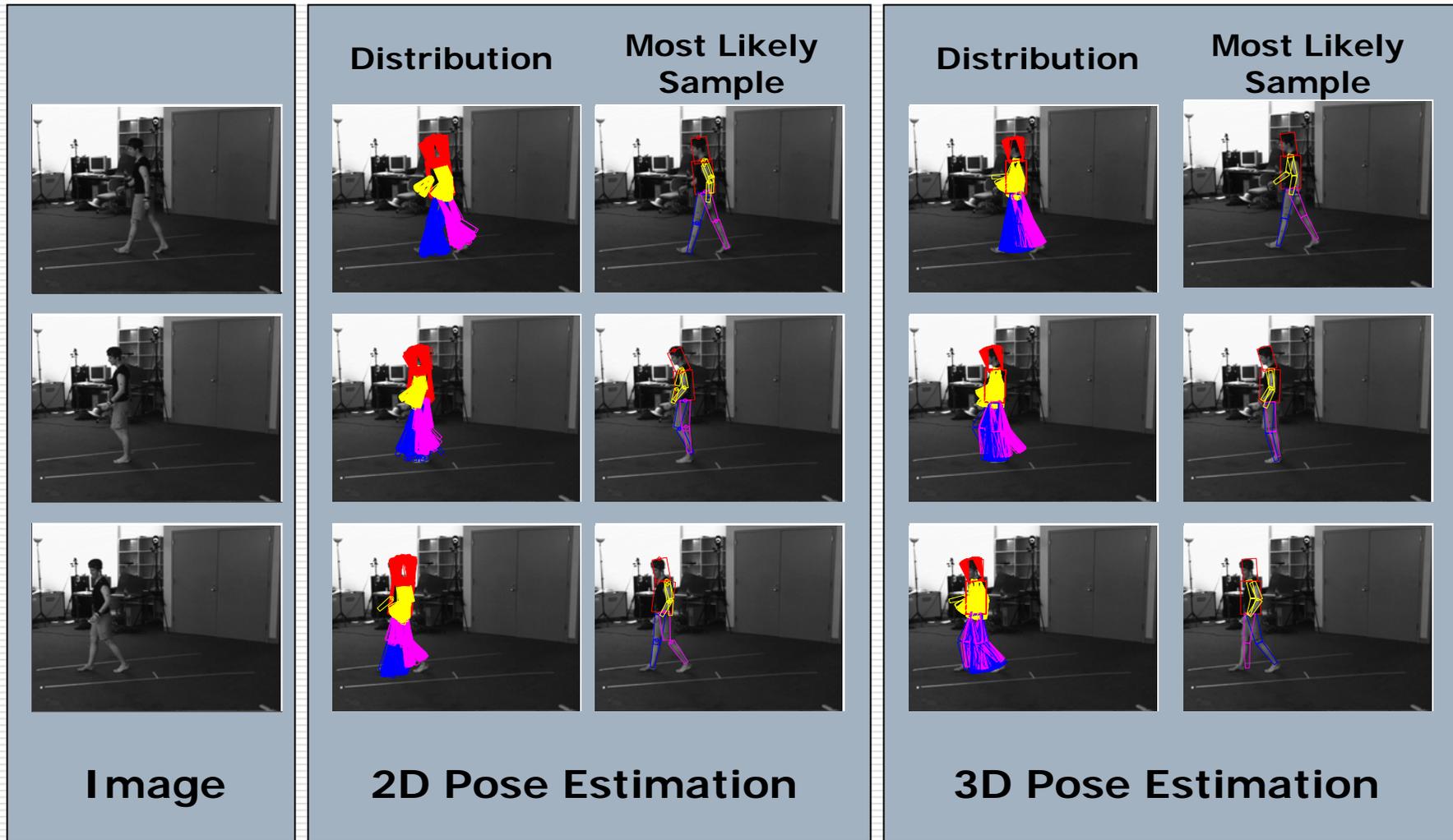
Occlusion-sensitive
“Loose-limbed” body
model

Mixture of
Experts (MoE)

Hidden Markov
Model (HMM)

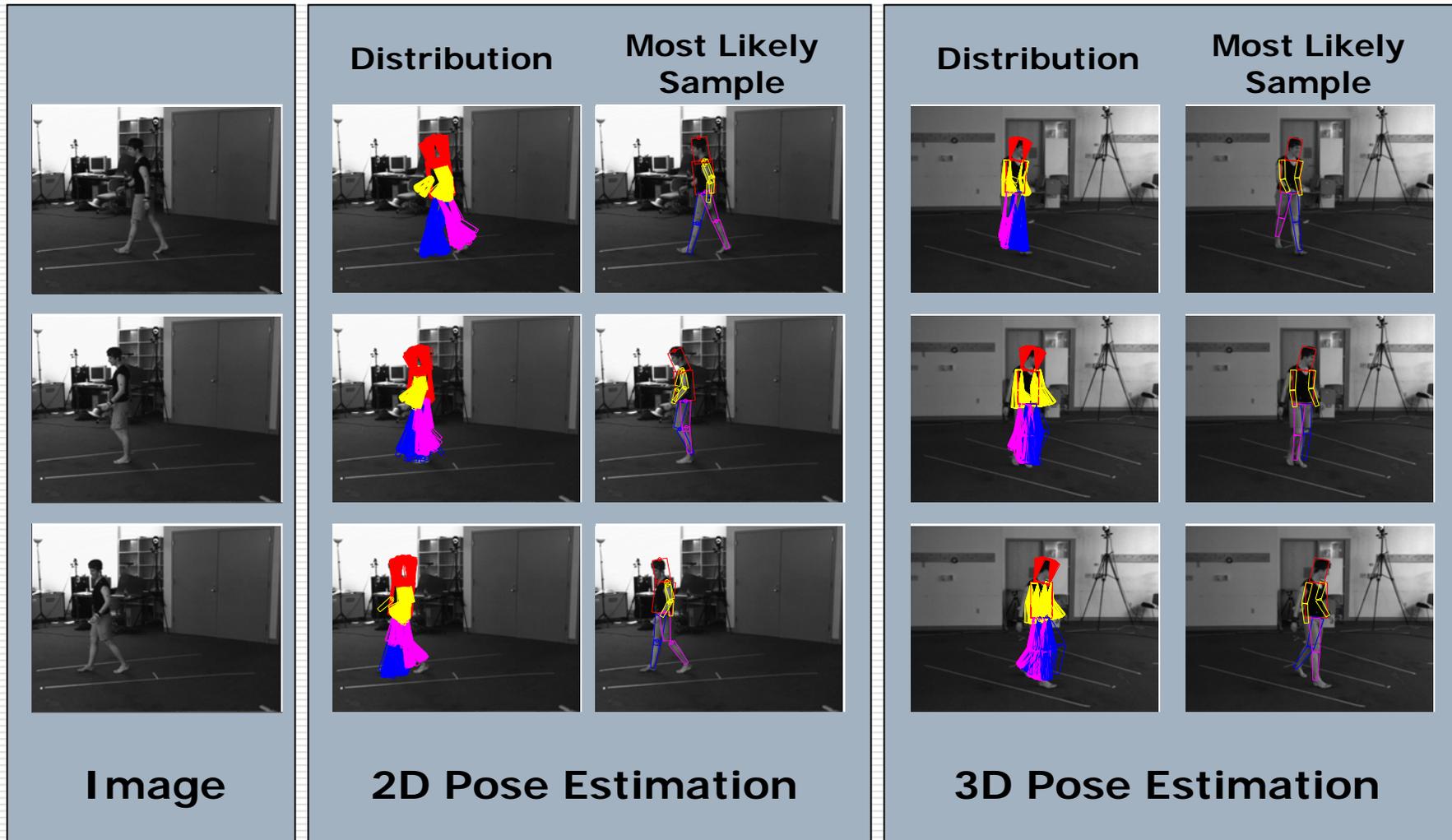


Hierarchical 3D Pose Estimation from Single View Monocular Images



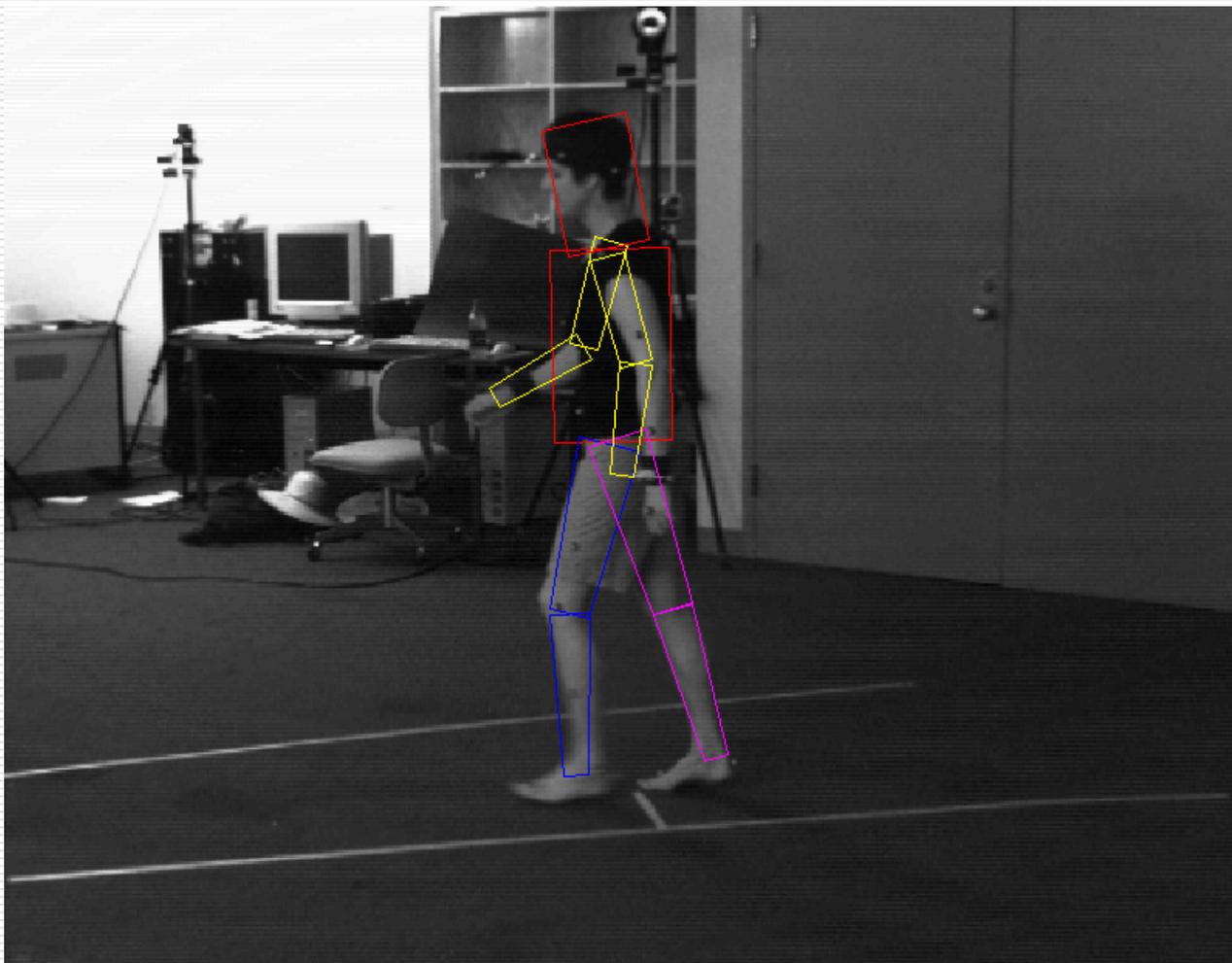


Hierarchical 3D Pose Estimation from Single View Monocular Images





Tracking in 3D



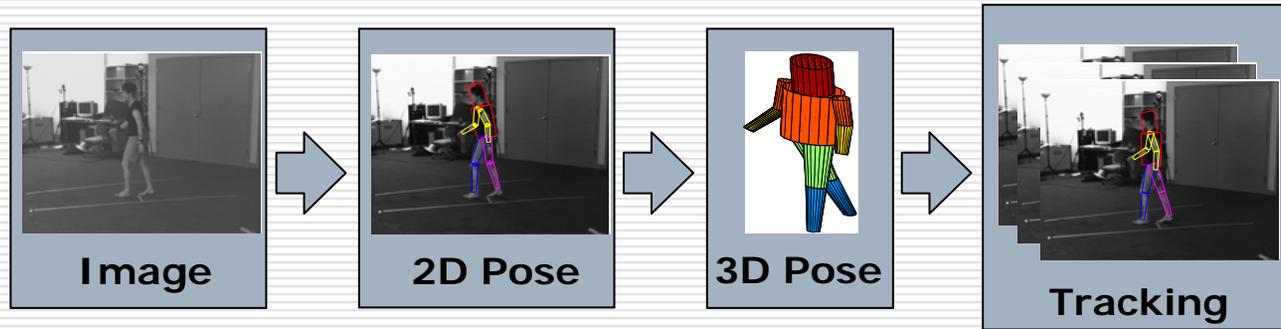


Tracking in 3D





Summary



- **We introduced a novel hierarchical inference framework**
 - Where we mediate the complexity of single-image monocular 3D pose estimation by intermediate 2D pose estimation stage
- **Inference in this framework can be tractably done using a variant of Non-parametric Belief Propagation**
- **Results obtained are very encouraging**



THANK YOU