Visual Analysis of High-Dimensional Motion: A Collaborative Approach

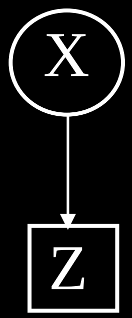
Aggelos K. Katsaggelos
Work by Ying Wu

Dept. EECS
Northwestern University
Evanston, IL 60208

http://www.ece.northwestern.edu/~yingwu
yingwu@ece.northwestern.edu

Preamble

- Computer vision includes many inverse problems, i.e., the inference of "hidden factors" from images
- Motion analysis is one of these inverse problems
  - e.g., estimating rigid/non-rigid, simple/complex motions so as to track moving targets and recognize motion patterns.
- These tasks can be apparently accommodated by the Bayesian framework
  - X: hidden factors, e.g., motion parameters
  - Z: image observations
  - $p(X|Z) \propto p(Z|X)p(X)$

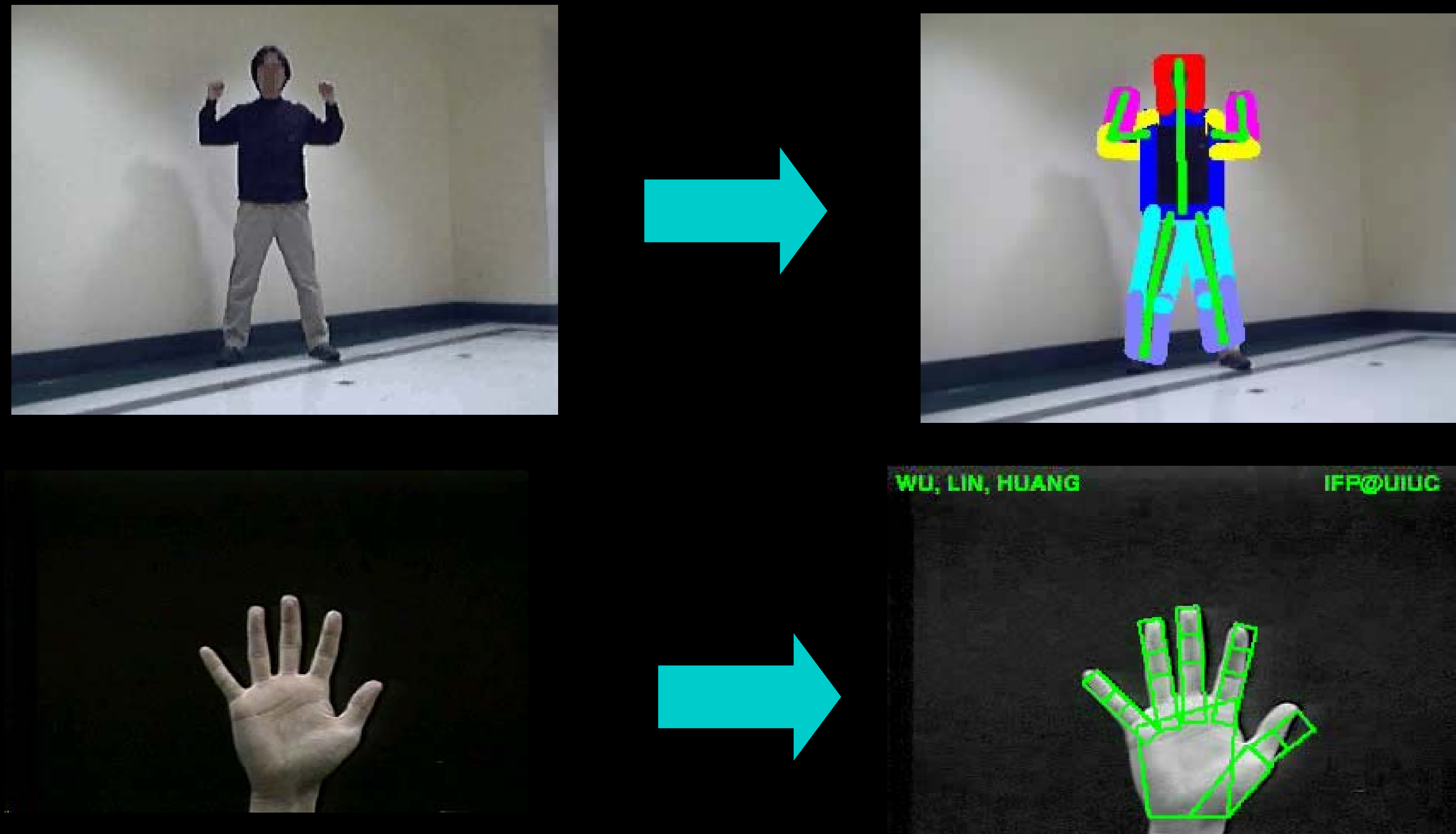


- Things are fine when talking about low-dim motion, such as rigid motion and affine motion.
- But …

High-dimensional Motion

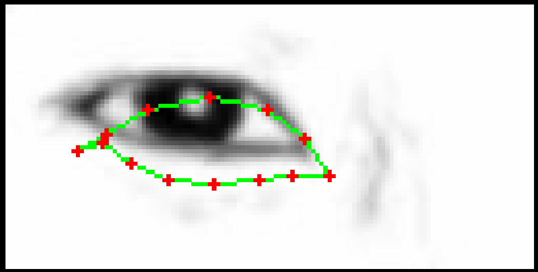
- What about those complex motions with a larger number of degrees of freedom?
- High-dimensional motion (HDM)
  - Articulation of linked kinematical structures
  - Deformation of elastic contours or surfaces
  - Multi-motion of multiple occluding targets
  - …
- Applications
  - Intelligent video surveillance
  - Human computer interaction
  - Video understanding and multimedia databases
  - Medical imaging

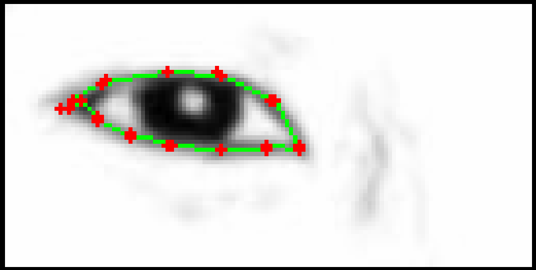
The Problem (articulation)

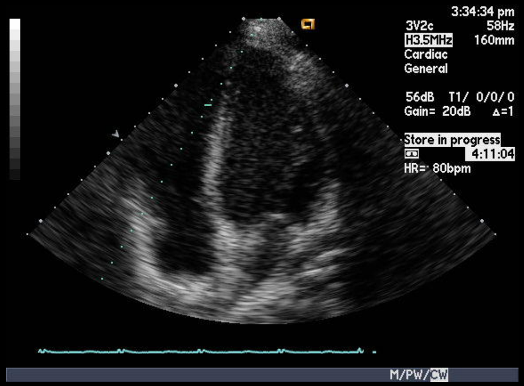


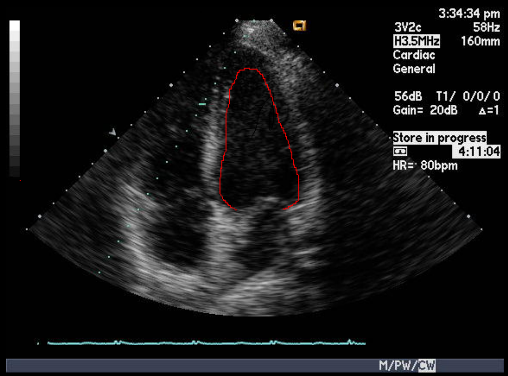
tracking a complex articulated structure in monocular video

The Problem (deformation)







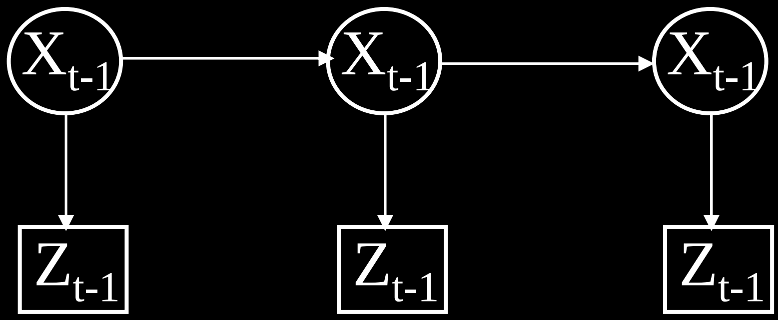


The Problem (multi-motion)

The Problem

- X_t : motion at time t
- Z_t : image measurement (evidence) at time t
- $\underline{Z}_t = \{Z_1, \ldots, Z_t\}$: measurement history
- A major task of motion analysis is to calculate the posterior $p(X_t | \underline{Z}_t)$ of X_t given measurement history
- Easy to see the recursive form

$$p(X_t | \underline{Z}_t) \propto p(Z_t | X_t) \int p(X_t | X_{t-1}) p(X_{t-1} | \underline{Z}_{t-1}) dX_{t-1}$$



State-of-the-Art

- Basic approaches
  - differential-based approaches (bottom-up)
    - ✓ construct an objective function and minimize it
  - Prediction-correction approaches (top-down)
    - ✓ parametric methods
      - Kalman filtering
    - ✓ non-parametric (or sampling-based) methods
      - Particle filtering
- Particle Filtering
  - A p.d.f. is represented by a set of particles
  - The solution is found by the evolution of the particles
  - It is flexible for non-Gaussian densities
- Obviously, the dimension of X and the prior p(X) determine the solution space.
- Both work for low-dim motion, e.g., rigid motion, since dim(X) is low and p(X) is simple.

State-of-the-Art

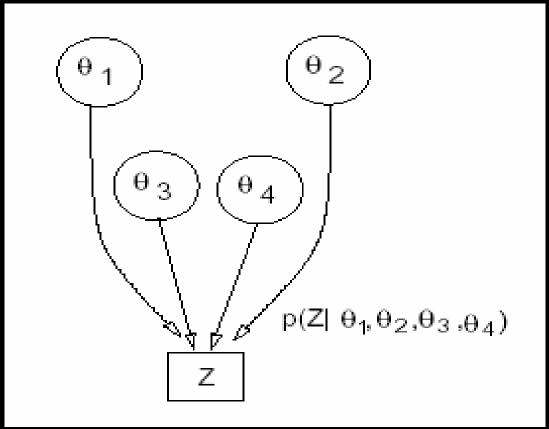
- But HDM has a completely different story.
- these approaches are confronted by the "*curse of dimensionality*"!
  - i.e., tremendous performance degradation of effectiveness and efficiency when the dimensionality increases
  - differential approach
    - ✓ difficult to calculate high-dim derivatives
    - ✓ too many local optima
  - sampling-based approach
    - ✓ exponential requirements for samples
    - ✓ computationally prohibitive

The other way around?

- Reducing the dimension
  - To seek for the lowest dimensional subspace of X
    - ✓ linear subspaces
    - ✓ nonlinear manifolds
  - To model p(Y), where Y is the low-dim projection of X
    - ✓ configuration space
- But …

Inevitability of a high-dim space ?!

- The intrinsic complexity of HDM itself
  - Low-dimensional manifolds (linear/nonlinear) may exist
    - ✓ by reducing motion correlation or motion constraints
    - ✓ for specific motion (like walking, hand grasping)
  - But the intrinsic complexity is irreducible
  - It may be quite high for those less-constrained HDM
    - ✓ E.g., arbitrary body articulation
- The conditional dependency of HDM given images



$$p(\Theta|\mathbf{Z}) \neq \prod_{i=1}^{4} p(\theta_i|\mathbf{Z}).$$

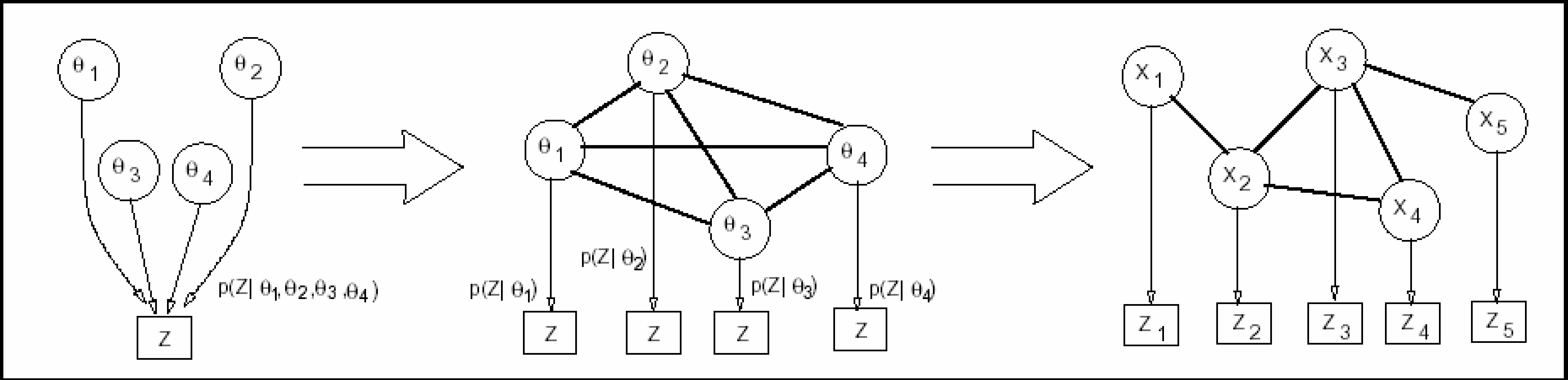
The ROOT of the curse

- The centralized methodology
  - Motion is modeled in a centralized fashion
    - ✓ Centralized models are compact (and low-dimensional)
    - ✓ But the intrinsic complexity bounds the dimensionality
    - ✓ Motion parameters are tightly correlated
    - ✓ i.e., we have to deal with p(X) as a whole
  - Then image observation also has to be centralized
    - ✓ Image observation Z is produced by X
    - ✓ Thus, we have to deal with p(Z|X) as a whole
- It is very tight, since we have to work with a pretty high dimensional but irreducible space.
- Is it a dead end?

New Approach

- "dimension reduction" → "dimension redundancy"
- Why not going to an even higher dim. space?
  - A distributed motion representation
    - ✓ A relaxed representation
    - ✓ highly redundant but loosely correlated
    - ✓ Exploiting motion correlations rather than eliminating them
    - ✓ enables distributed image observations
  - A collaborative motion analyzer
    - ✓ !! Completely different from the conventional approach, which uses one single but high-dim and super-powerful motion analyzer.
    - ✓ We try to use a group of mutually-dependent (collaborative) low-dim motion analyzer to do the job.

The Idea: an illustration



The conventional centralized approach:

(1) Motion is modeled by Θ

(2) Difficult to model motion prior $p(\Theta)$

(3) $p(\Theta|Z) \propto p(Z|\Theta)p(\Theta) \neq \Pi\, p(\theta_k|Z)$

(4) Computationally infeasible, and seems to be no way to turn

The new collaborative approach:

(1) Motion model is redundant

(2) X is a networked subpart X_i

(3) Motion prior is distributed in the network

(4) Image observations are also distributed

(5) A set of low-dim motion analyzer $p(X_i|Z_i)$ collaborates and solves the problem

Theory and Benefits

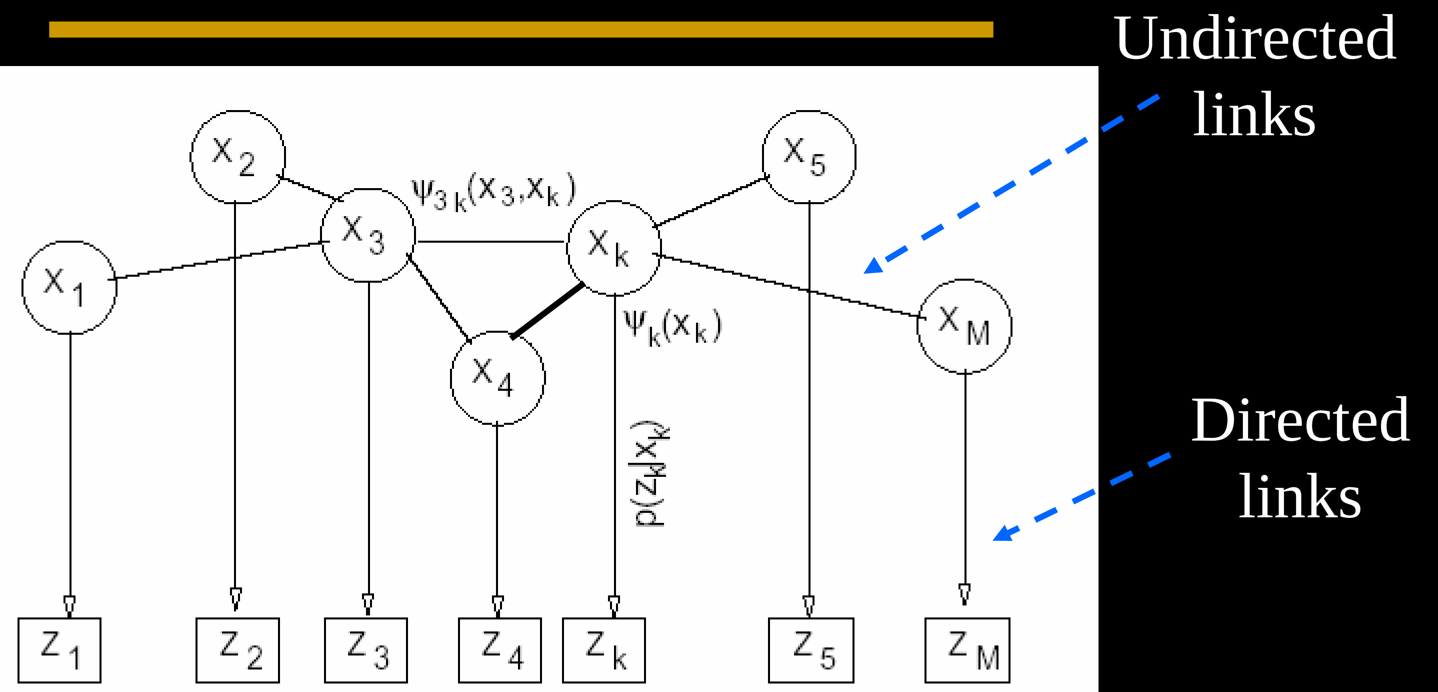
- Theoretical foundation (see later slides)
  - Markov networks,
  - Variational analysis and mean field theory
  - Collaboration and competition mechanisms
- Benefits of the new approach
  - dramatic reduction of computation
    - ✓ from exponential → close-to-linear
  - robustness to occlusion and clutters
    - ✓ handling conditional dependency

Theory

- Distributed representation & Markov network
- Tool: probabilistic variational analysis
- The beauty of the mean field theory
- A new computational paradigm

Distributed Motion Model

Markov Network



$$p(\mathbf{X}) = \frac{1}{Z_c} \prod_{(i,j)\in\mathcal{C}^2} \psi_{ij}(\mathbf{x_i}, \mathbf{x_j}) \prod_{i\in\mathcal{C}^1} \psi_i(\mathbf{x_i})$$

$$p(\mathbf{Z}|\mathbf{X}) = \prod_{i=1}^{n} p_i(\mathbf{z_i}|\mathbf{x_i}).$$

Variational Analysis

- We want to infer $p(\mathbf{x}_i|\mathbf{Z})$
- It is difficult, because of the networked structure.
- We perform probabilistic variational analysis
- The idea is to find an optimal approximation $q^*(\mathbf{X})$, su$p(\mathbf{X}|\mathbf{Z})$t the Kullback-Leibler (KL) divergence of these two distribution is minimized:

$$
\begin{aligned}
q^*(\mathbf{X}) &= \arg\min_q KL(q(\mathbf{X}||p(\mathbf{X}|\mathbf{Z})) \\
&= \arg\min_q \int_x q(\mathbf{X}) \log \frac{q(\mathbf{X})}{p(\mathbf{X}|\mathbf{Z})}
\end{aligned}
$$

Mean Field Theory

- When we choose a full factorization variation:

$$q(\mathbf{X}) = \prod_{i}^{M} q_i(\mathbf{x}_i)$$

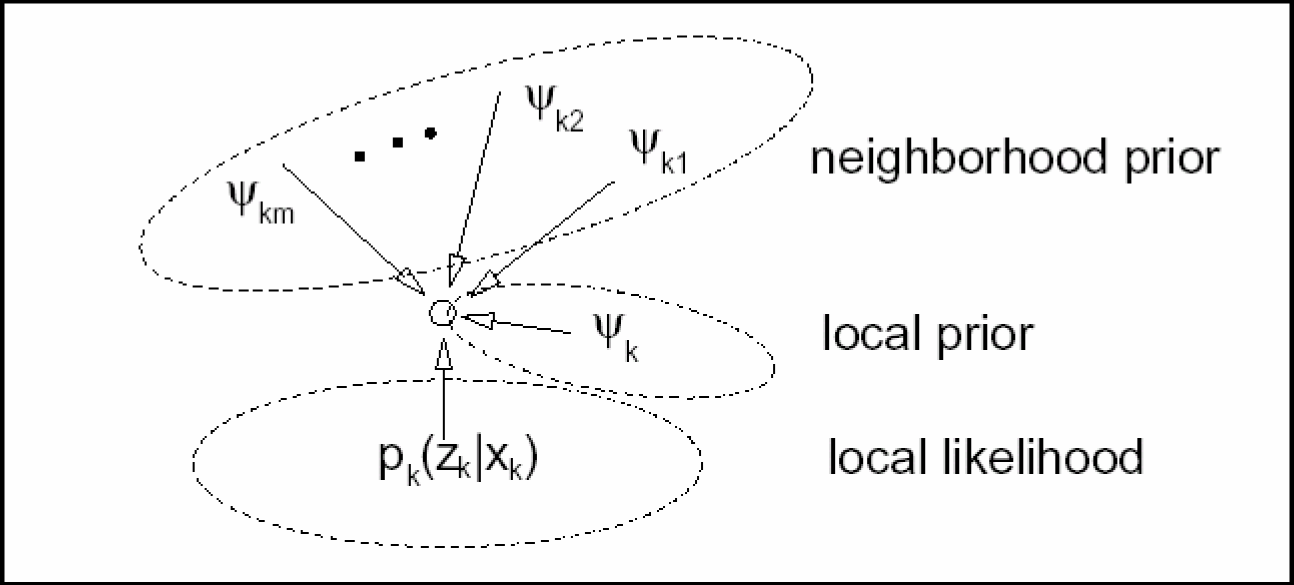
- We end up with a very interesting result: a set of fixed point equations:

$$q_i(\mathbf{x}_i) \longleftarrow \frac{1}{Z_i'} p_i(\mathbf{z}_i|\mathbf{x}_i)\psi_i(\mathbf{x}_i)M_i(\mathbf{x}_i), \qquad \text{where}$$

$$M_i(\mathbf{x}_i) = \exp\{\sum_{k\in\mathcal{N}(i)} \int_{x_k} q_k(\mathbf{x}_k)\log\psi_{ik}(\mathbf{x}_i, \mathbf{x}_k)\},$$

- This is very similar to the Mean Field theory in statistical physics.

Computational Paradigm



- Three factors affect the posterior of a node:
  - Local prior
  - Neighborhood prior
  - Local likelihood

Algorithms

- Collaborative particle networks
- Example: Mean field Monte Carlo (MFMC)
- Complexity: from exponential to linear
- Unsolved problems

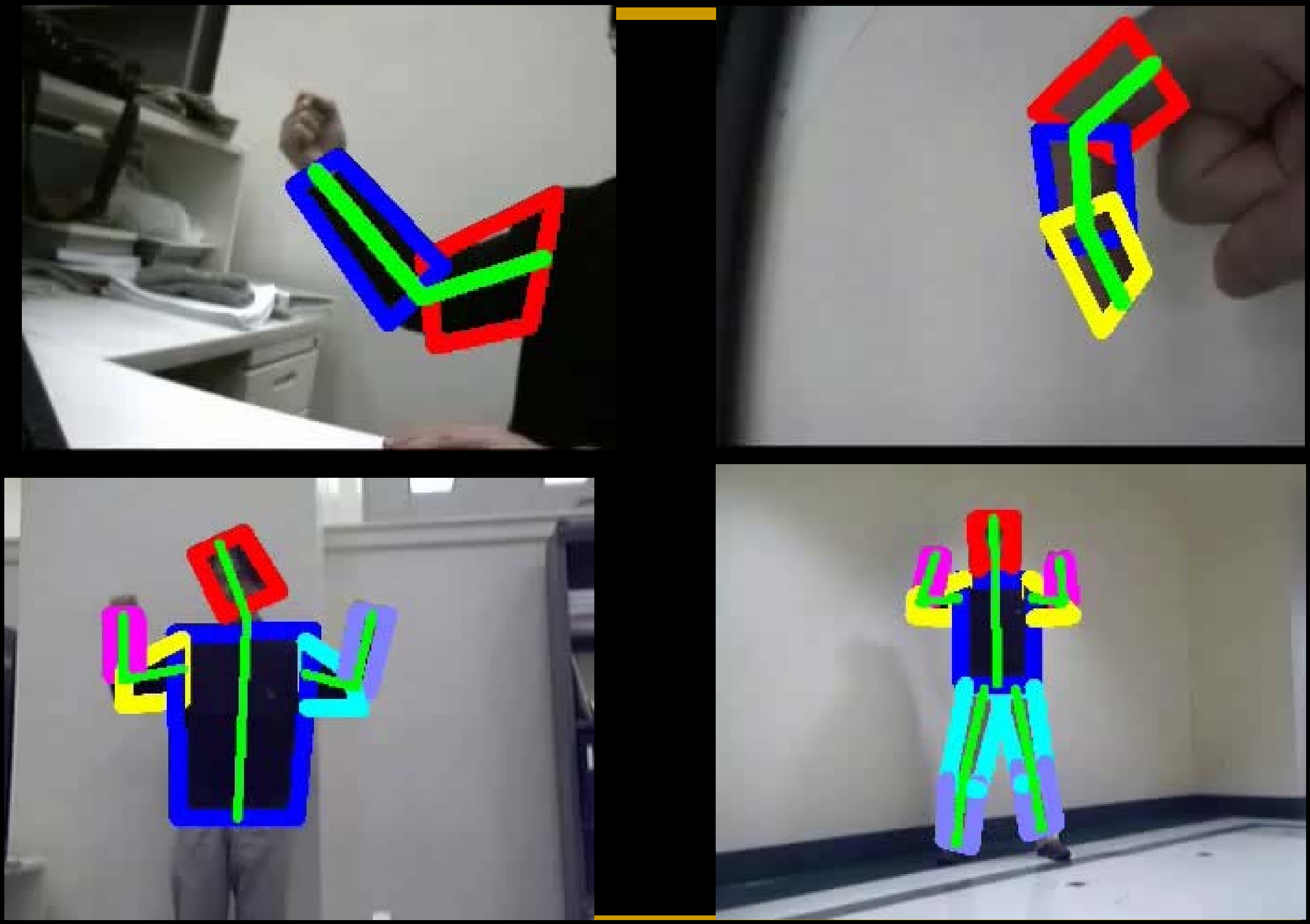
Case Studies

- Articulated motion capturing
- Multi-target motion tracking
- Deformable motion alignment

Cooperation and articulation

- Cooperation
  - Provides inclusive information to others
    - ✓ *"I am here, you probably should be somewhere around"* ☺
  - Two sources
    - ✓ Physical constraints, e.g.,
      - Connectivity
      - Smoothness
      - Distance
    - ✓ Purposive constraints, e.g.,
      - Specific motion correlations
- Articulation is a good example for cooperation

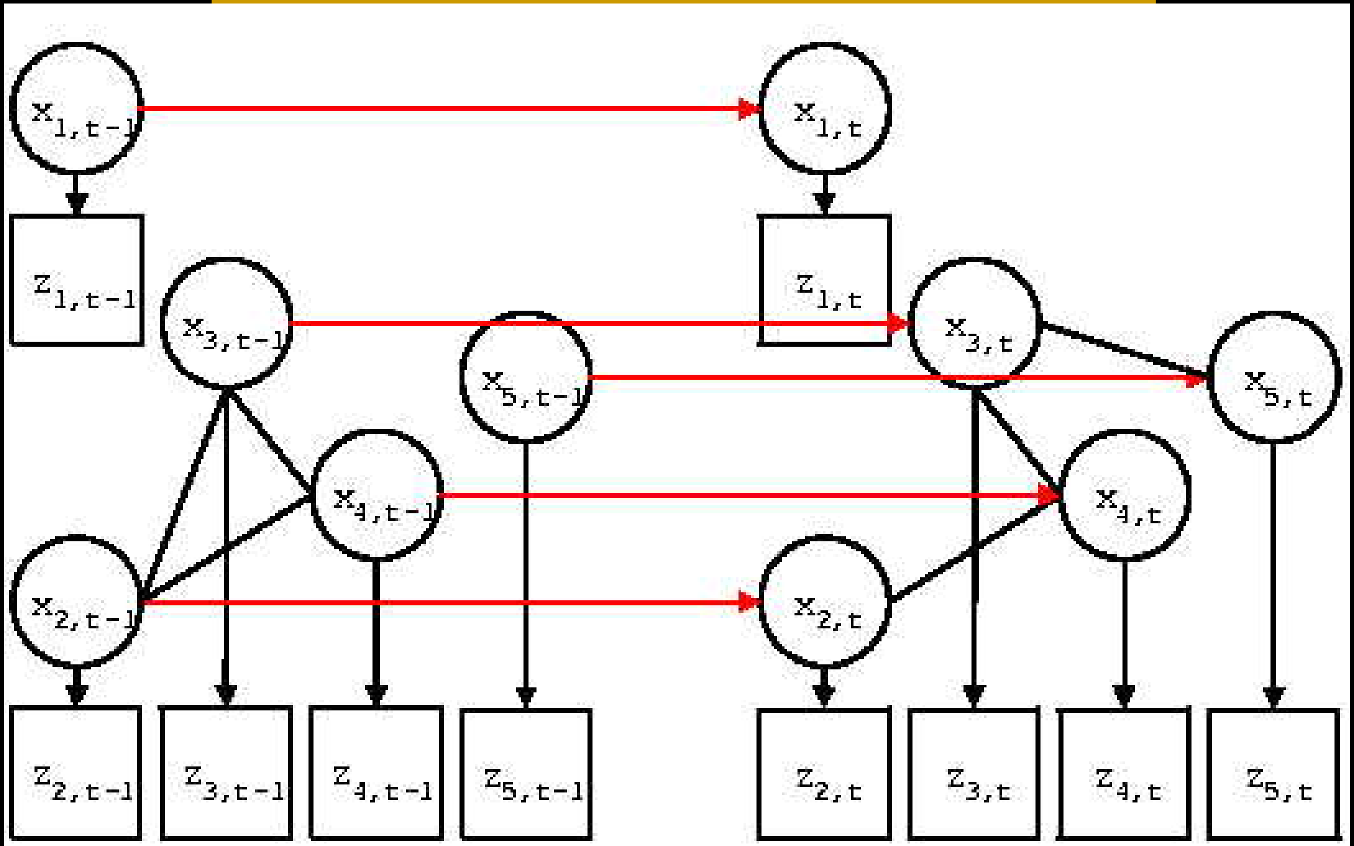
Initial Results



Competition and Multi-motion

- Competition
  - Provides exclusive information to others
    - ✓ *"I am here, you probably should not be somewhere here also"* ☹
  - Why competition?
    - ✓ competing for common image resources
    - ✓ to handle conditional dependency
- Multi-target tracking is a suitable case
  - The motion of multiple targets are obviously independent when they are far apart
  - However, when they get closer and occlude each other, since it is difficult to distinguish from images which is which, they become conditionally dependent once image observations are made.
- Ad hoc Markov network and ad hoc MFMC

Ad hoc Markov Network

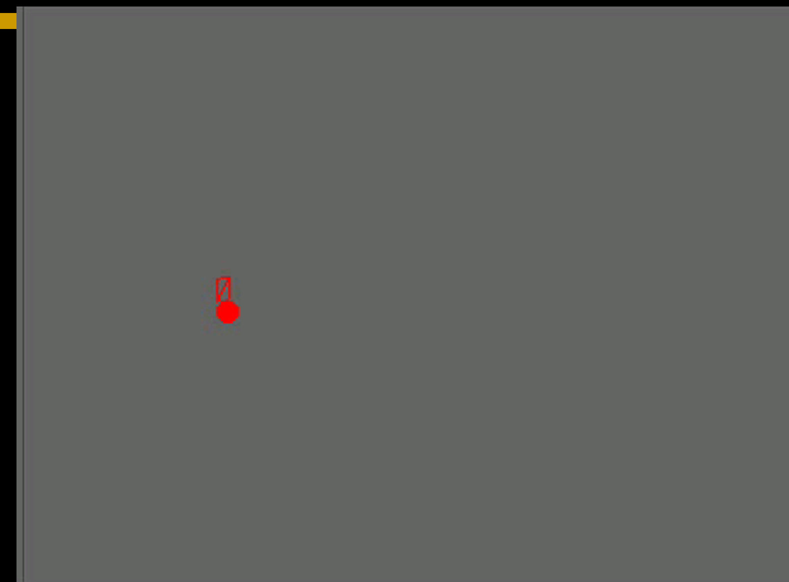


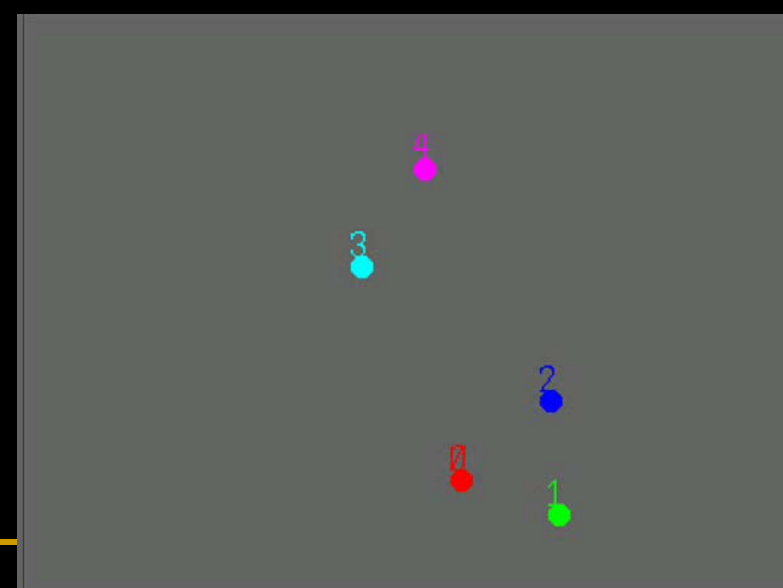
- The topology of the network changes with time
- The connectivity of two nodes depends on the distance of two targets (i.e., if they are close enough)

Tracking Multiple Targets

Click to play video

Tracking Multiple Targets

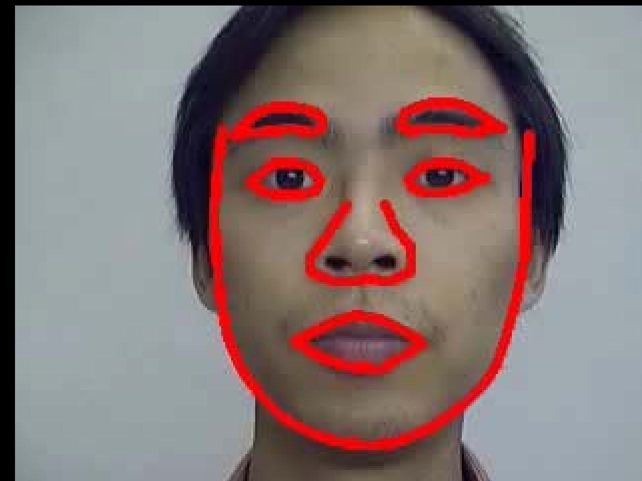


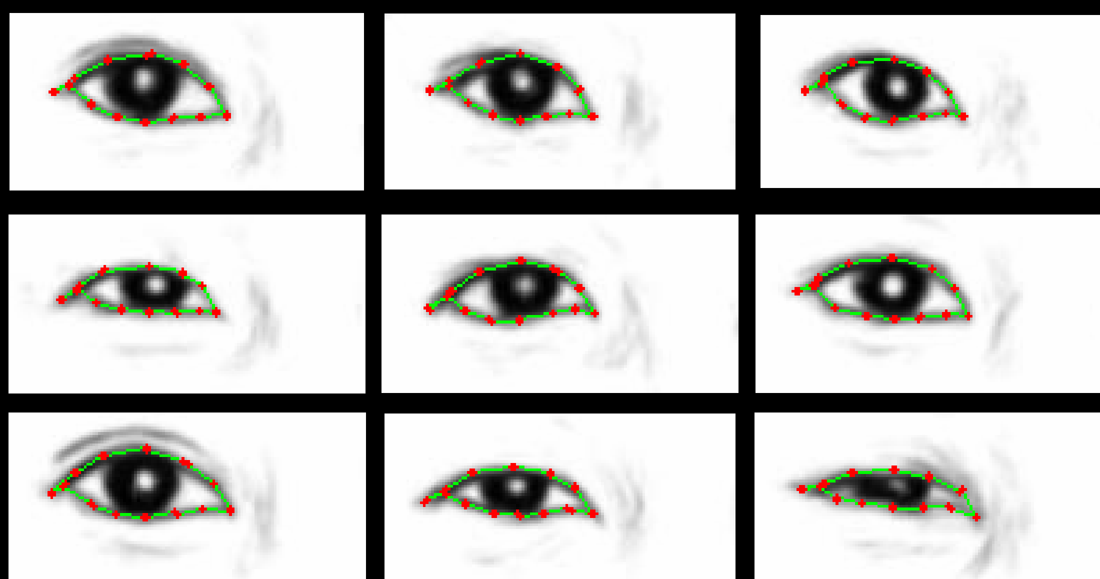


Collaboration and Deformation

- Collaboration
  - A combination of cooperation and competition
- Deformable motion
  - Structured
  - non-structured

Initial Results





Conclusions

- Visual motion capturing of non-rigid targets is challenging due to the high-dimensionality of the motion.
- Existing methods (e.g., differential-based and sampling-based methods) can not scale to complex high-dim motion tracking, due to the curse of dimensionality.
- The new approach:
  - A decentralized representation → A Markov network
  - Motion capturing → Bayesian inference of the network
  - A variational analysis → a new computational diagram
  - Implementation → Mean Field Monte Carlo (MFMC)
  - A collaborative particle network → an efficient solution
- This new approach aims at an efficient and effective solution to this challenging task.