Beyond Trees: Common-Factor Model for 2D Human Pose Recovery

Xiangyang Lan Dan Huttenlocher

Computer Science Department, Cornell University

Outline

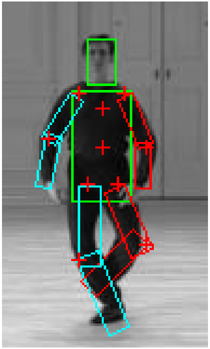
1. Problem: Spatial Constraints Modeling for Human
   - Build spatial constraint(prior) model for human body, to determine 2D human pose from single image.
2. Previous Work: Tree-Structured Models
   - Conditional independence assumption on the limbs can be problematic
3. Our Approach: Common-Factor Models
   - Use factor analysis to identify latent (control) variable, capture correlation among limbs in a simple way, and permit efficient inference.
   - Experimental results

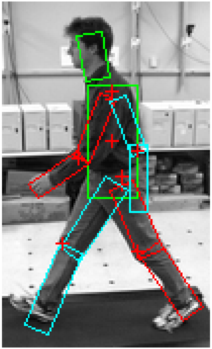
2D Human Pose Recovery from Single Image

- Learning: spatial constraints (prior) modeling for human body.
- Inference: MAP estimate (maximum a posteriori) for the most probable pose.
- Example:

2D Human Pose Recovery from Single Image

- Learning: spatial constraints (prior) modeling for human body.
- Inference: MAP estimate (maximum a posteriori) for the most probable pose.
- Example:





Previous Work: Tree-Structured Models

- P.F. Felzenszwalb and D.P. Huttenlocher.
  Efficient Matching of Pictorial Structures. CVPR 2000.
- D. Ramanan and D.A. Forsyth.
  Finding and Tracking People from the Bottom Up. CVPR03.
- L. Sigal, S. Bhatia, S. Roth, M.J. Black, and M. Isard.
  Tracking Loose-Limbed People. CVPR04.
- Many others ...

Most earlier work uses Tree-Structured Model

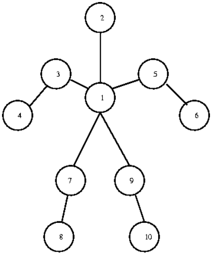
- Captures kinematic structure of human body joints
- Efficient inference algorithms exist.

Previous Work: Tree-Structured Model

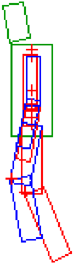
Undirected Graphical Models:

- nodes : random variables (location for body parts)
- edges : explicit dependency among random variables.

One Example of Tree-Structured Model:



tree-structured model

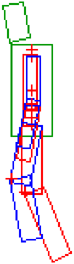


parameterized for side-viewed walking

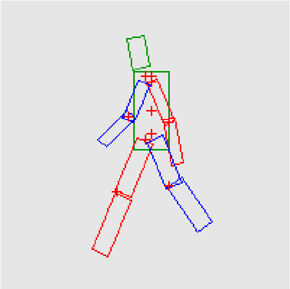
estimate (x, y, s, o) *for each node*

However . . .

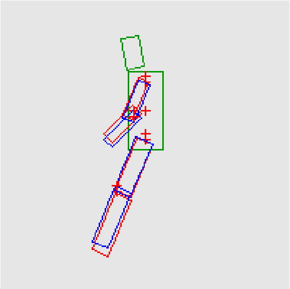
- Certain poses that are valid under the kinematic constraints of a tree model are highly unnatural.



model



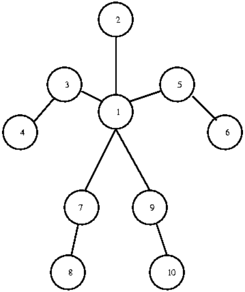
natural



unnatural...

Why?

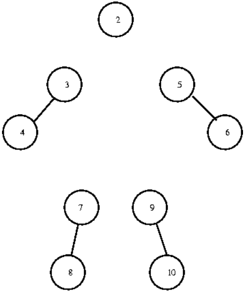
- Limbs may violate the assumption of conditional independence.
- For instance, balancing and coordinating are important in many activities.



Previous Work: Tree-Structured Model

Why?

- Limbs may violate the assumption of conditional independence.
- For instance, balancing and coordinating are important in many activities.



Our Method: Common-Factor Models

Build a spatial model with richer structure

Possible Solution?

- Sample highly probable poses.
- Build a richer model !

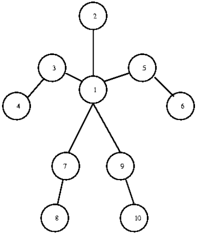
We take the second approach here, and introduce common-factor models:

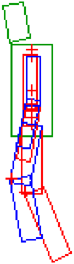
- augment tree models using factor analysis,
- captures correlation among limbs beyond the kinematic tree,
- while still allowing efficient inference.

Starting Point : Learn a Tree-Structured Model

- Training Data:
  - 240 frames of a side-view of a person walking, marked with the location of each body part, from the CMU HumanID database.

- Learned Pictorial Structure:





Residual Covariance Analysis

Identify the limbs that violate conditional independence assumption

- Node : parameterized as (x, y, s, o).
- Residual Y_{ij} : difference between the expected and observed part j location wrt. its parent part i.
- Edge : modeling the kinematic joint as: $\phi(l_i, l_j) = N(Y_{ij}, 0, \sigma_{ij})$.

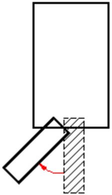
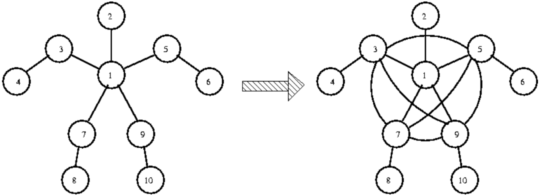


Table: Correlation Coefficient Table in Orientation Dimension

	Head	Lf. Arm	Lf. Leg	Rt. Arm	Rt. Leg
Head	1.00	0.00	-0.00	-0.06	0.00
Lf. Arm	0.00	1.00	-0.58	-0.83	0.67
Lf. Leg	-0.00	-0.58	1.00	0.61	-0.43
Rt. Arm	-0.06	-0.83	0.61	1.00	-0.59
Rt. Leg	0.00	0.67	-0.43	-0.59	1.00

Building Clique: Augment The Tree Structure

Graphical Structure:

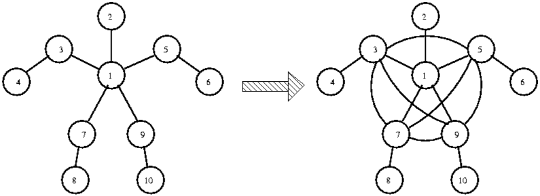


Modeling each joint independently $\Rightarrow$ Modeling set of correlated joints collectively (Clique Potential function):

- $\phi_C(L_C) = \mathcal{N}((Y_{r1}, \ldots, Y_{rk}), 0, \Sigma_C) = \mathcal{N}(Y_C, 0, \Sigma_C)$
- Y_{ij}: residue over the joint for part i and j.
- L_C: the locations for all parts within clique C.

Building Clique: Augment The Tree Structure

Graphical Structure:



Inference algorithm:

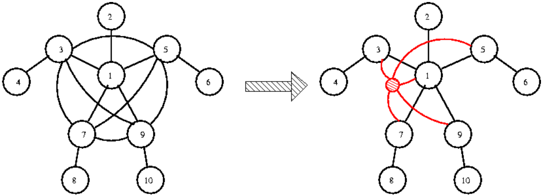
- Intractable. Searching space is exponential in the clique size (i.e. h^5, h is the possible locations a part can take).

Solution:

- Factor Analysis to determine the intrinsic dimensionality.

Final Product : Common-Factor Model

Graphical Structure:



$$Y = \mathcal{N}(AX, \Lambda)$$

- $X \sim \mathcal{N}(0,1)$: the common factor, 1D random variable in the orientation dimension.
- Y: the upper limbs' orientation.
- A: factor loading vector, capturing how X controls the upper limbs' orientation Y.

General Algorithm for Learning

- Learn a tree model from the training data.
- Identify the parts that violate the conditional independence assumption of the tree.
- Use factor analysis to discover the latent control variable(s) that causes the correlations between those parts.
- Introduce that latent control variable(s) into the graphical structure to capture that additional correlations.

General Algorithm for Inference (MAP Estimate)

A variant of Viterbi algorithm:

- For each fixed value of X, the common-factor model degrades into a tree. Apply dynamic programming to get the MAP estimate for that X.
- Return the global optimal answer over all values of X as the MAP estimate.

With distance transform, running time equals $O(nh_x h)$, versus $O(nh^4)$ for clique model.

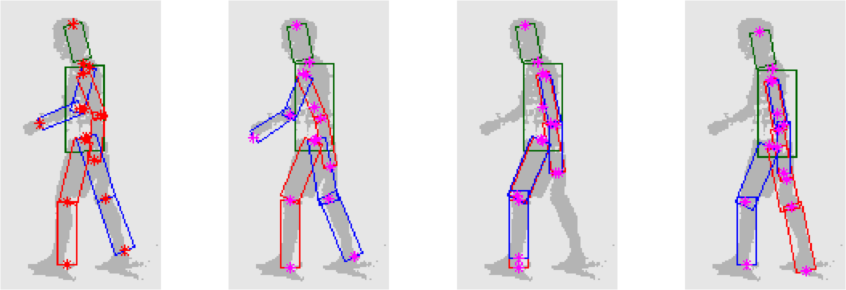
- n: number of parts.
- h_x: number of possible values (bins) for common factor X.
- h: number of possible locations (bins) for each part.

Experimental Results: Brown Walking Data

Testing Data:

- 50 frames from the Brown sequence, with ground truth joint locations obtained from MOCAP system.

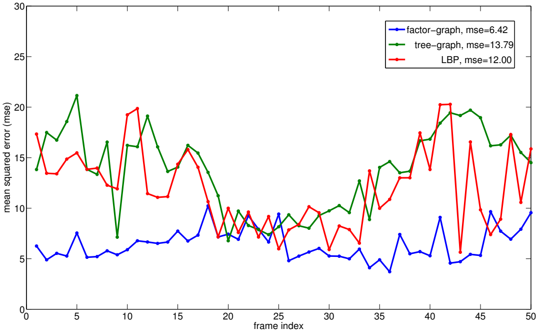
MAP (maximum a posterior) Estimate:



ground truth | *common-factor model* | *tree-structured model* | *clique model with LBP*

Experimental Results: Brown Walking Data

Per-frame error, average over the joints.

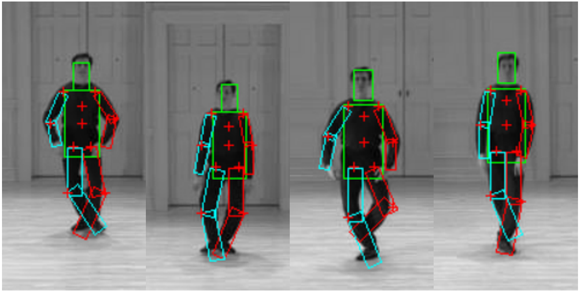


Per-joint error, average over the frames.

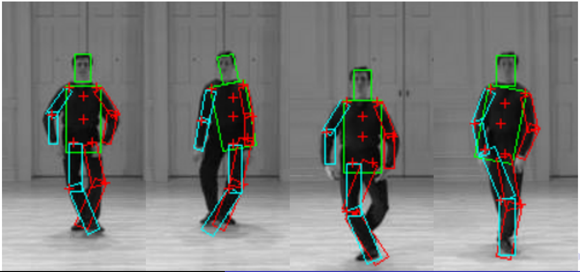
	shoulder	elbow	wrist	hip	knee	ankle
Factor	4.8	5.5	8.6	4.2	4.4	5.4
Tree	9.1	11.1	19.4	6.4	6.6	28.6
LBP	9.9	11.9	20.5	6.4	5.3	20.5

More Experiment Results: Dancing Snapshots

Common-Factor model:



Tree-Structured model:



Conclusion

Contribution:

- Develop common-factor models, that augment the tree models using factor analysis.
- It captures the spatial correlations (dependencies) among limbs as well as the kinematic structure in a simple way, and still allows efficient inference algorithm.
- Enable us to do 2D human pose recovery using a graphical model beyond trees efficiently.