Pictorial Structures for Articulated Pose Estimation

Ankit Gupta

CSE 590V: Vision Seminar

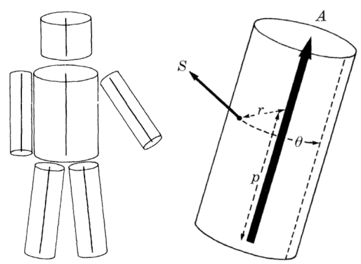
Goal

Articulated pose estimation ( )

recovers the pose of an articulated object which consists of joints and rigid parts

Slide taken from authors, Yang et al.

Classic Approach



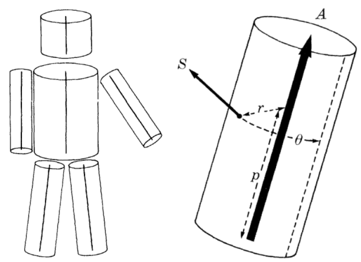
Marr & Nishihara 1978

Part Representation

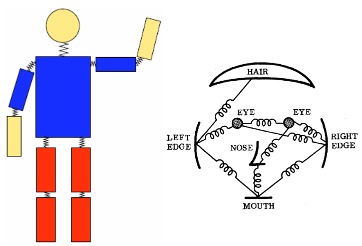
- Head, Torso, Arm, Leg
- Location, Rotation, Scale

Slide taken from authors, Yang et al.

Classic Approach



Marr & Nishihara 1978



Fischler & Elschlager 1973
Felzenszwalb & Huttenlocher 2005

Part Representation

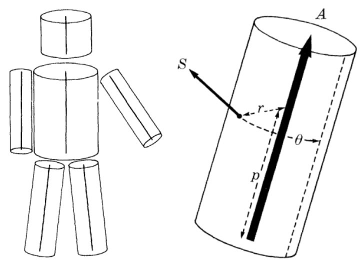
- Head, Torso, Arm, Leg
- Location, Rotation, Scale

Pictorial Structure

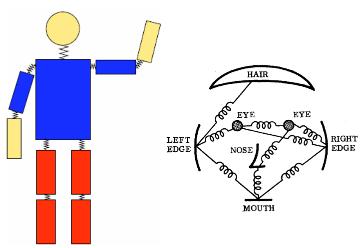
- Unary Templates
- Pairwise Springs

Slide taken from authors, Yang et al.

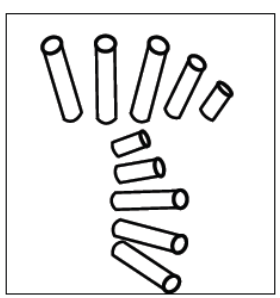
Classic Approach



Marr & Nishihara 1978



Fischler & Elschlager 1973
Felzenszwalb & Huttenlocher 2005



Part Representation

- Head, Torso, Arm, Leg
- Location, Rotation, Scale

Pictorial Structure

- Unary Templates
- Pairwise Springs

Lan & Huttenlocher 2005
Sigal & Black 2006
Ramanan 2007
Epshteian & Ullman 2007
Wang & Mori 2008
Ferrari etc. 2008

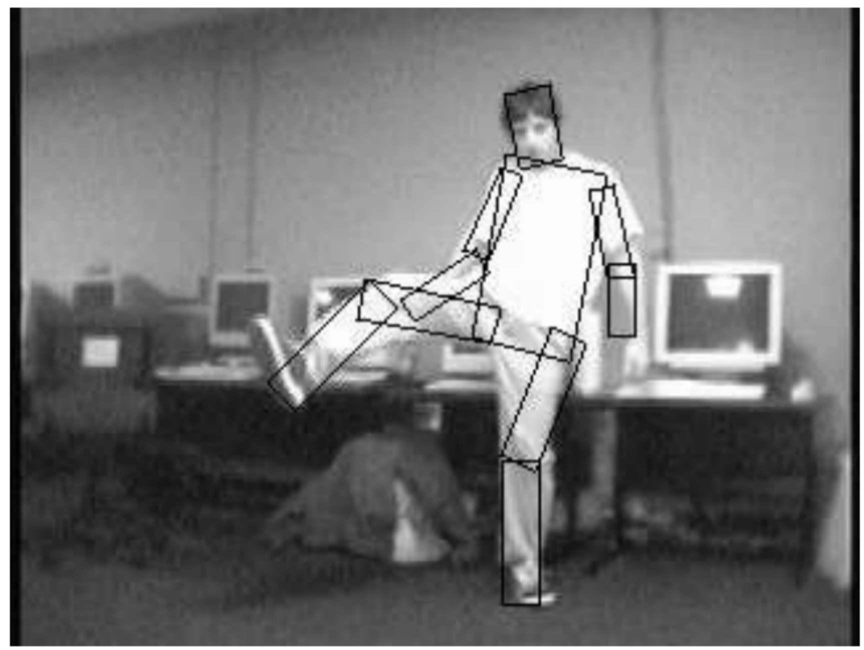
Andriluka etc. 2009
Eichner etc. 2009
Singh etc. 2010
Johnson & Everingham 2010
Sapp etc. 2010
Tran & Forsyth 2010

Slide taken from authors, Yang et al.

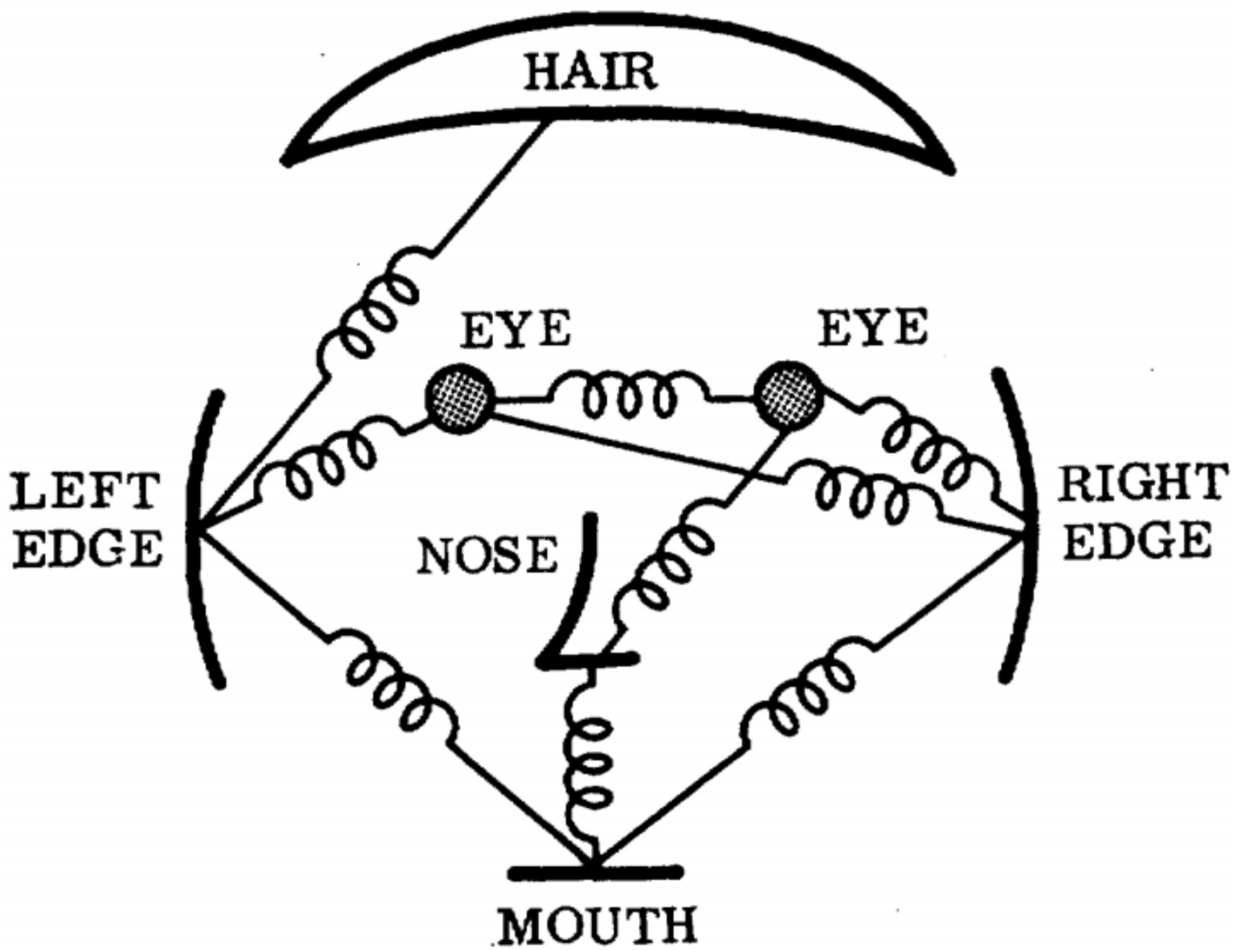
Pictorial Structures for Object Recognition

Pedro F. Felzenszwalb, Daniel P. Huttenlocher

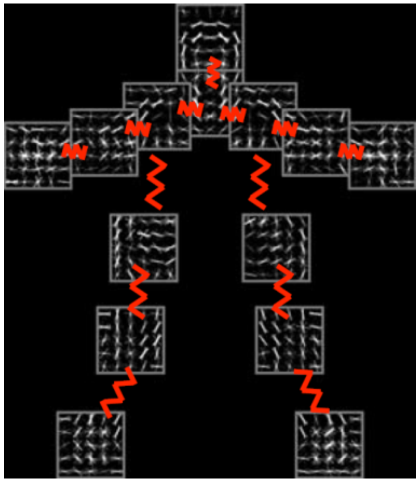
IJCV, 2005



Pictorial structure for Face



Pictorial Structure Model

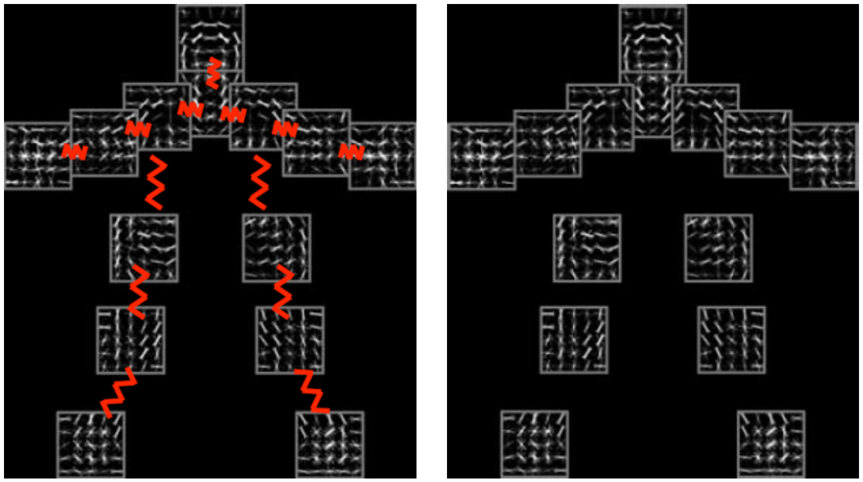
- 

$S(I, L)$

- I: Image
- l_i: Location of part i

Slide taken from authors, Yang et al.

Pictorial Structure Model

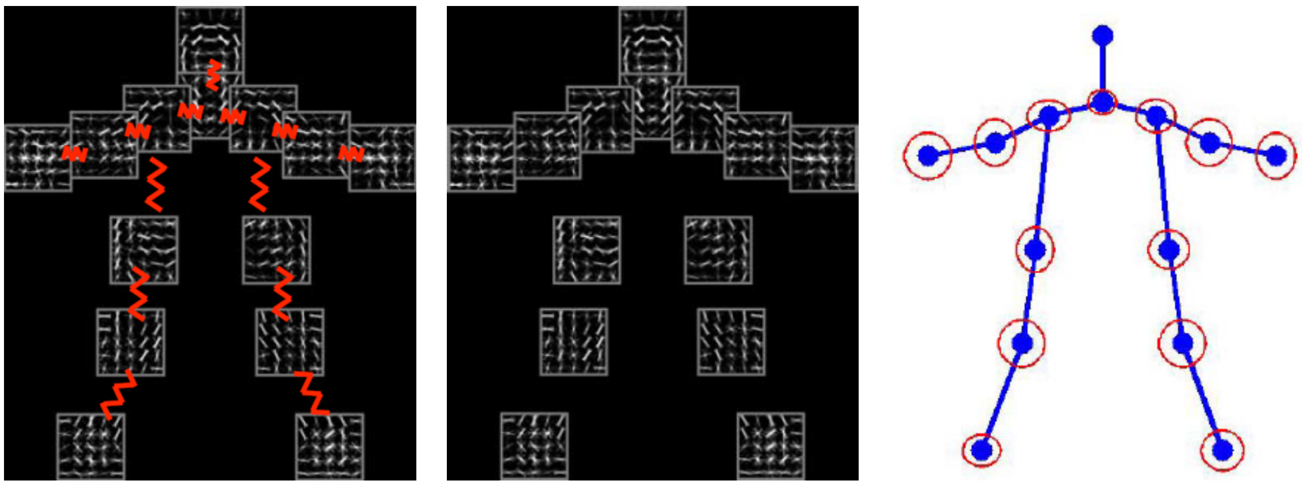
- 

$$\underbrace{S(I,L)}\; = \underbrace{\sum_{i\in V} \alpha_i \cdot \phi(I,l_i)}$$

- α_i : Unary template for part i
- $\phi(I,l_i)$: Local image features at location l_i

Slide taken from authors, Yang et al.

Pictorial Structure Model

- 

$$S(I,L) = \sum_{i \in V} \alpha_i \cdot \phi(I, l_i) + \sum_{ij \in E} \beta_{ij} \cdot \psi(l_i, l_j)$$

- $\psi(l_i, l_j)$: Spatial features between l_i and l_j
- β_{ij}: Pairwise springs between part i and part j

Slide taken from authors, Yang et al.

Using this Model

Train phase

Test phase

Train phase

$$S(I,L) = \sum_{i \in V} \alpha_i \cdot \phi(I, l_i) + \sum_{ij \in E} \beta_{ij} \cdot \psi(l_i, l_j)$$

Given:
- Images (I)
- Known locations of the parts (L)

Need to learn
- Unary templates α_i
- Spatial features β_{ij}

Test phase

Train phase

Test phase

$$S(I,L) = \sum_{i \in V} \alpha_i \cdot \phi(I, l_i) + \sum_{ij \in E} \beta_{ij} \cdot \psi(l_i, l_j)$$

Given:
- Images (I)
- Known locations of the parts (L)

Need to learn
- Unary templates α_i
- Spatial features β_{ij}

Standard Structural SVM formulation
- Standard solvers available (SVMStruct)

Train phase

$$S(I,L) = \sum_{i \in V} \alpha_i \cdot \phi(I,l_i) + \sum_{ij \in E} \beta_{ij} \cdot \psi(l_i,l_j)$$

Given:
- Images (I)
- Known locations of the parts (L)

Need to learn
- Unary templates α_i
- Spatial features β_{ij}

Standard Structural SVM formulation
- Standard solvers available (SVMStruct)

Test phase

$$S(I,L) = \sum_{i \in V} \alpha_i \cdot \phi(I,l_i) + \sum_{ij \in E} \beta_{ij} \cdot \psi(l_i,l_j)$$

Given:
- Image (I)

Need to compute
- Part locations (L)

Algorithm
- L* = arg max (S(I,L))

Train phase

$$S(I,L) = \sum_{i \in V} \alpha_i \cdot \phi(I,l_i) + \sum_{ij \in E} \beta_{ij} \cdot \psi(l_i,l_j)$$

Given:
- Images (I)
- Known locations of the parts (L)

Need to learn
- Unary templates α_i
- Spatial features β_{ij}

Standard Structural SVM formulation
- Standard solvers available (SVMStruct)

Test phase

$$S(I,L) = \sum_{i \in V} \alpha_i \cdot \phi(I,l_i) + \sum_{ij \in E} \beta_{ij} \cdot \psi(l_i,l_j)$$

Given:
- Image (I)

Need to compute
- Part locations (L)

Algorithm
- L* = arg max (S(I,L))

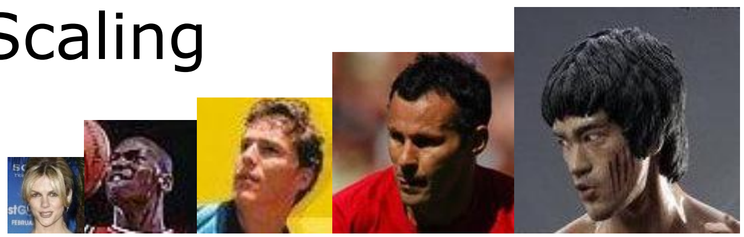
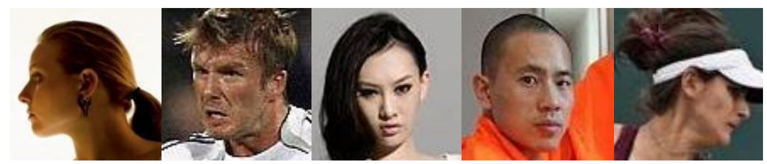
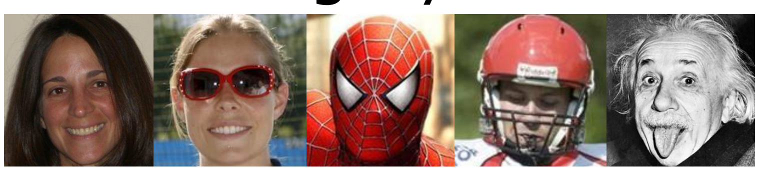
Standard inference problem
- For tree graphs, can be exactly computed using belief propagation

Articulated Pose Estimation with Flexible Mixtures of Parts

Yi Yang & Deva Ramanan

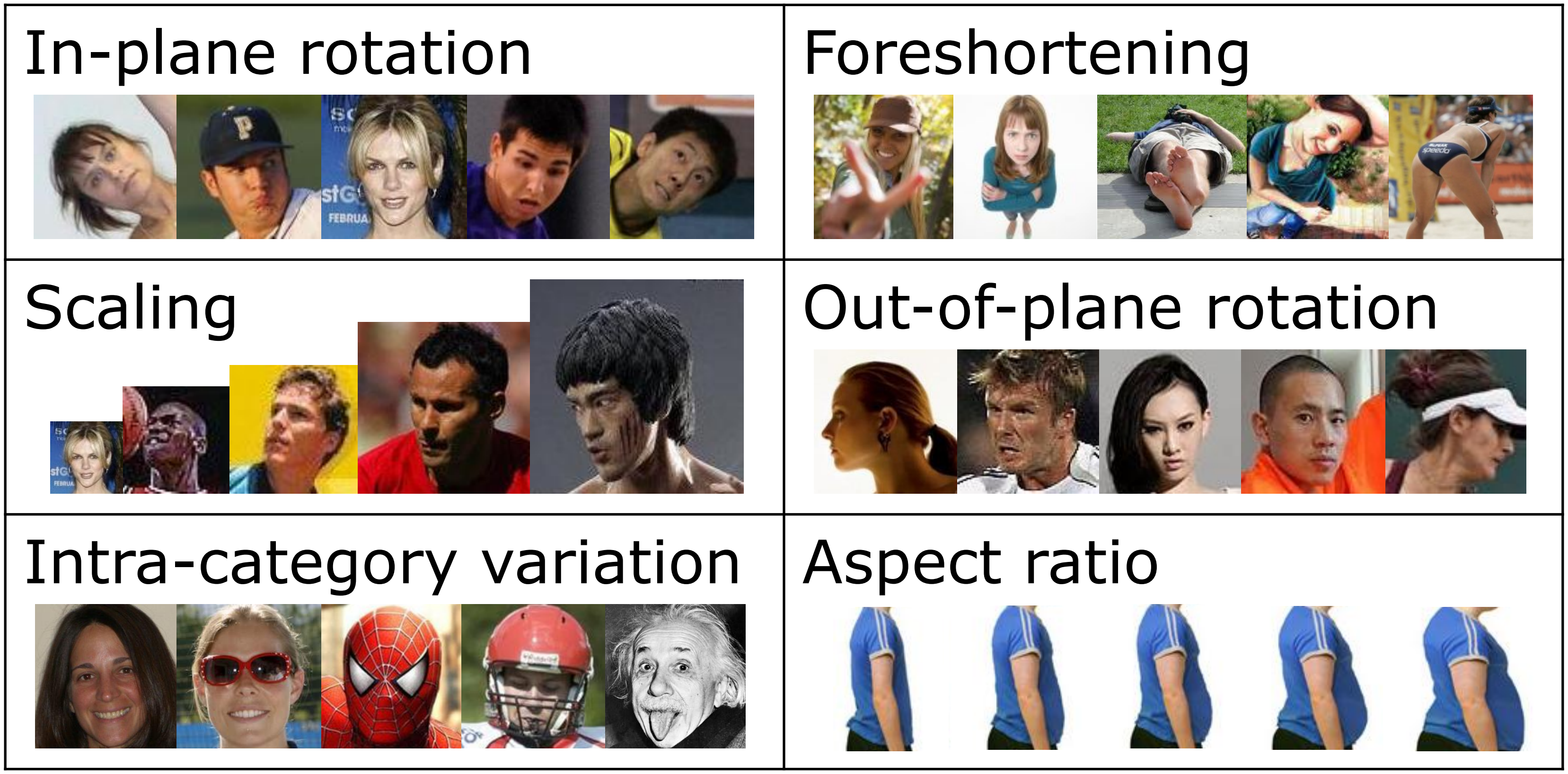
University of California, Irvine

Problems with previous methods: Wide Variations

In-plane rotation 	Foreshortening 
Scaling 	Out-of-plane rotation 
Intra-category variation 	Aspect ratio 

Slide taken from authors, Yang et al.

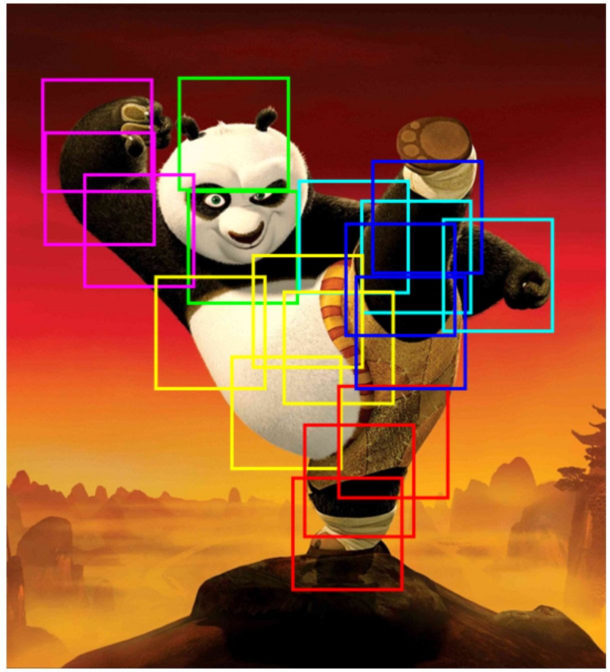
Problems with previous methods: Wide Variations



Naïve brute-force evaluation is expensive

Slide taken from authors, Yang et al.

Our Method – “Mini-Parts”

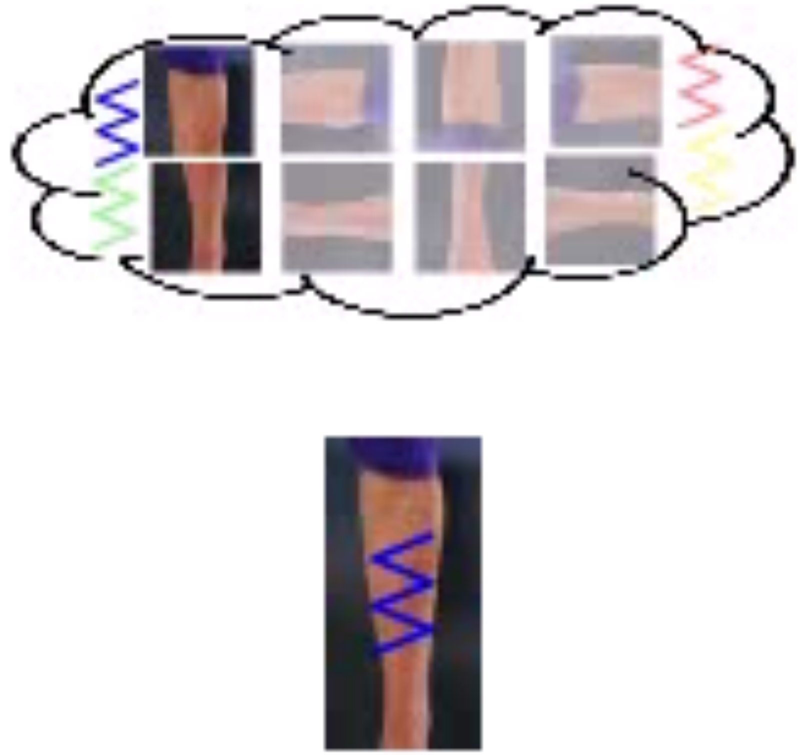


Key idea:

“mini part” model can approximate deformations

Slide taken from authors, Yang et al.

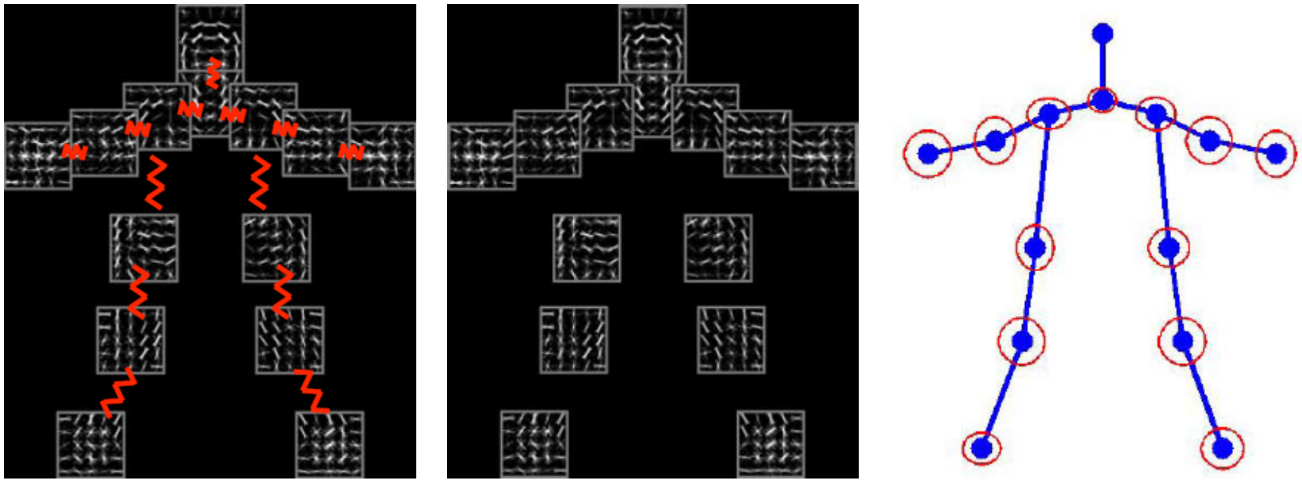
Example: Arm Approximation



Slide taken from authors, Yang et al.

Pictorial Structure Model

•

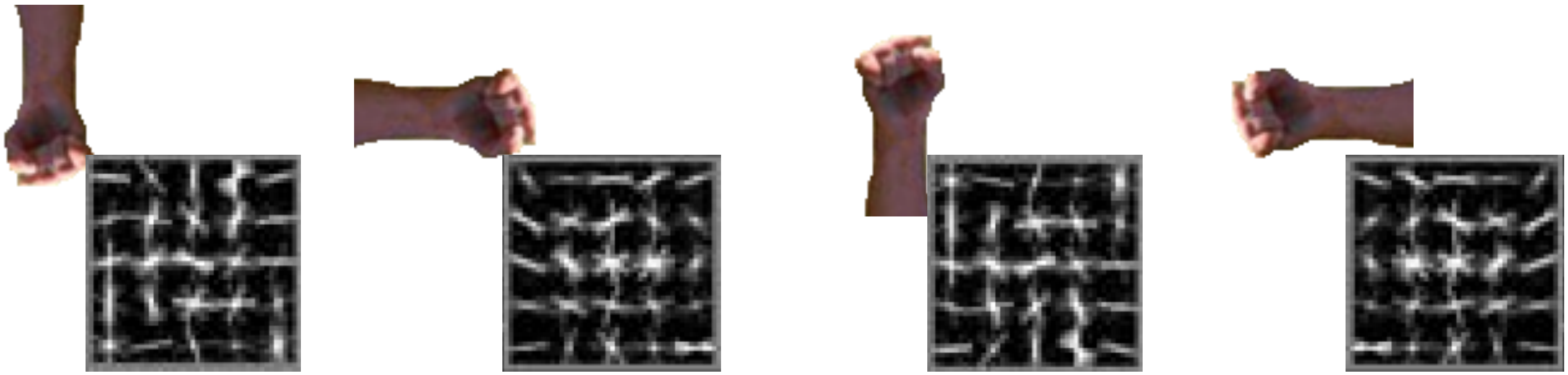


$$S(I,L) = \sum_{i \in V} \alpha_i \cdot \phi(I, l_i) + \sum_{ij \in E} \beta_{ij} \cdot \psi(l_i, l_j)$$

- $\psi(l_i, l_j)$: Spatial features between l_i and l_j
- β_{ij}: Pairwise springs between part i and part j

Slide taken from authors, Yang et al.

The Flexible Mixture Model

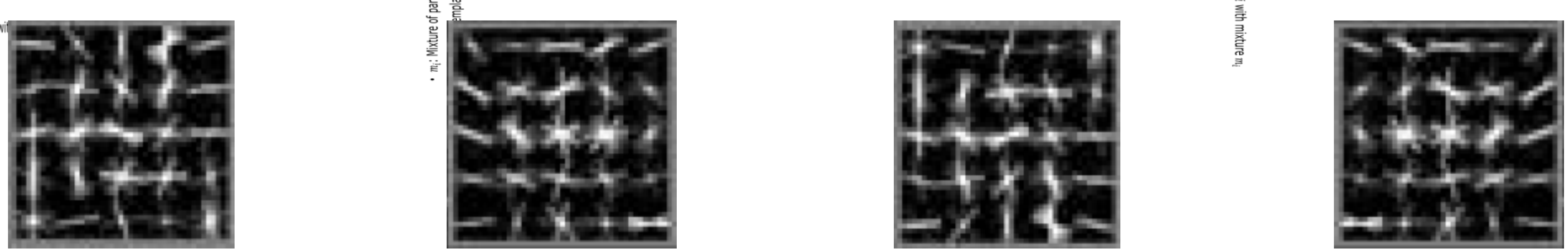


$S(I, L, M)$

- m_i: Mixture of part i

Slide taken from authors, Yang et al.

Our Flexible Mixture Model

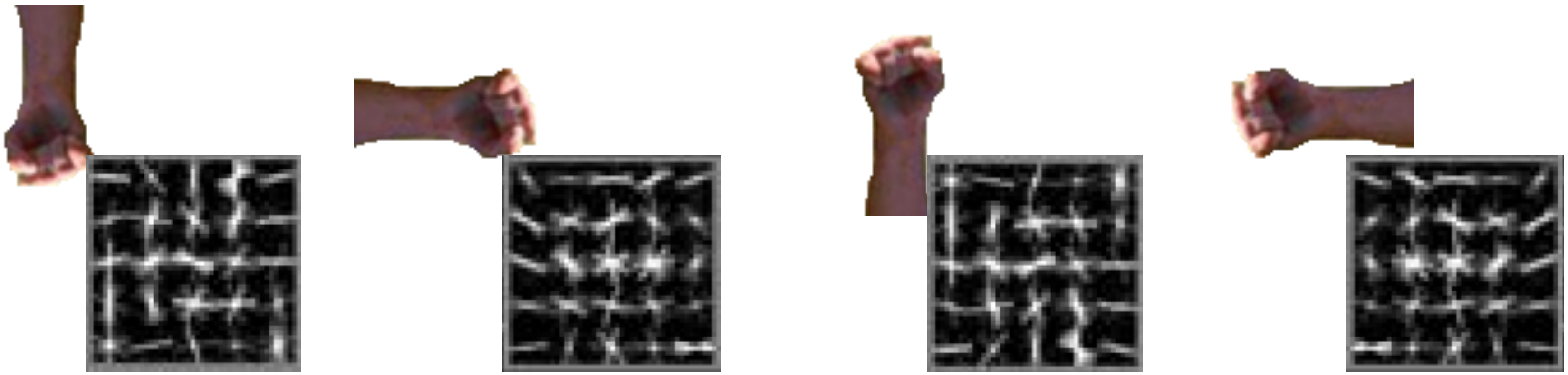


$$S(I, L, M) = \sum_{i \in V} \alpha_i^{m_i} \cdot \phi(I, l_i)$$

- m_i: Mixture of part i
- $\alpha_i^{m_i}$: Unary template for part i with mixture m_i

Slide taken from authors, Yang et al.

Our Flexible Mixture Model

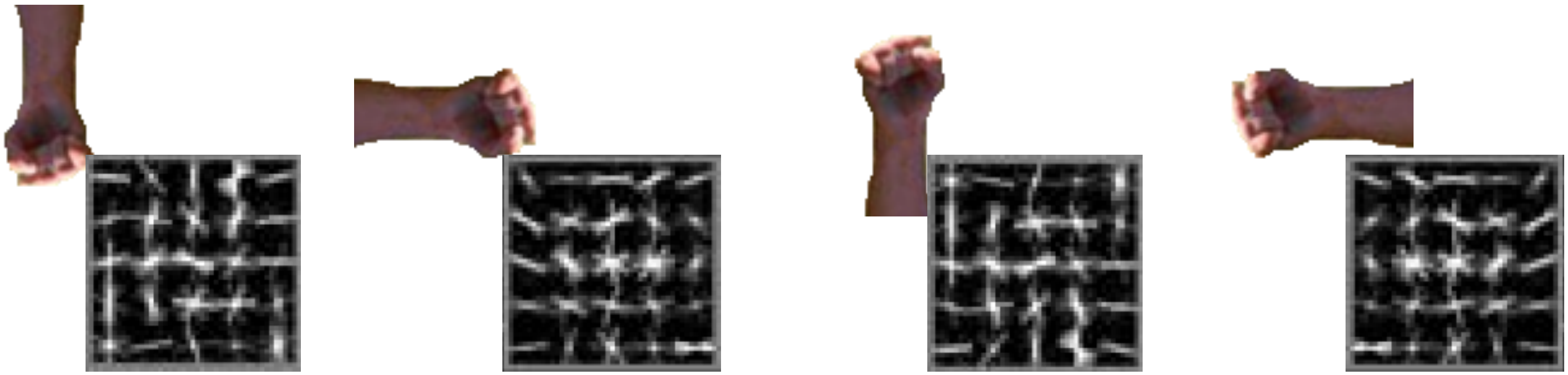


$$S(I, L, M) = \sum_{i \in V} \alpha_i^{m_i} \cdot \phi(I, l_i) + \sum_{ij \in E} \beta_{ij}^{m_i m_j} \cdot \psi(l_i, l_j)$$

- m_i: Mixture of part i
- $\alpha_i^{m_i}$: Unary template for part i with mixture m_i
- $\beta_{ij}^{m_i m_j}$: Pairwise springs between part i with mixture m_i and part j with mixture m_j

Slide taken from authors, Yang et al.

Our Flexible Mixture Model



$$S(I, L, M) = \sum_{i \in V} \alpha_i^{m_i} \cdot \phi(I, l_i) + \sum_{ij \in E} \beta_{ij}^{m_i m_j} \cdot \psi(l_i, l_j) + S(M)$$

- m_i: Mixture of part i
- $\alpha_i^{m_i}$: Unary template for part i with mixture m_i
- $\beta_{ij}^{m_i m_j}$: Pairwise springs between part i with mixture m_i and part j with mixture m_j

Slide taken from authors, Yang et al.

Co-occurrence "Bias"

- $$S(M) = \sum_{ij \in E} b_{ij}^{m_i m_j}$$

- $b_{ij}^{m_i m_j}$: Pairwise co-occurrence prior between part i with mixture m_i and part j with mixture m_j
- Can also add unary terms $b_i^{m_i}$ to have priors over mixtures of a part

Slide taken from authors, Yang et al.

Co-occurrence “Bias”: Example

Let

part i : eyes, mixture m_i = {open, closed}
part j : mouth, mixture m_j = {smile, frown}

Co-occurrence "Bias": Example

Let

part i : eyes, mixture m_i = {open, closed}
part j : mouth, mixture m_j = {smile, frown}

VS

b (closed eyes, smiling mouth)

b (open eyes, smiling mouth)

Co-occurrence "Bias": Example

Let

part i : eyes, mixture m_i = {open, closed}
part j : mouth, mixture m_j = {smile, frown}

$<$

learnt

b (closed eyes, smiling mouth)

b (open eyes, smiling mouth)

Using this Model

Train phase

Test phase

Train phase

$$S(I,L,M) = \sum_{i \in V} \alpha_i^{m_i} \cdot \phi(I, l_i) + \sum_{ij \in E} \beta_{ij}^{m_i m_j} \cdot \psi(l_i, l_j) + S(M)$$

Given:
- Images (I)
- Known locations of the parts (L)

Need to learn
- Unary templates α_i
- Spatial features β_{ij}
- Co-occurrence $S(M)$

Standard Structural SVM formulation
- Standard solvers available (SVMStruct)

Test phase

Train phase

$$S(I,L,M) = \sum_{i \in V} \alpha_i^{m_i} \cdot \phi(I,l_i) + \sum_{ij \in E} \beta_{ij}^{m_i m_j} \cdot \psi(l_i,l_j) + S(M)$$

Given:
- Images (I)
- Known locations of the parts (L)

Need to learn
- Unary templates α_i
- Spatial features β_{ij}
- Co-occurrence $S(M)$

Standard Structural SVM formulation
- Standard solvers available (SVMStruct)

Test phase

$$S(I,L,M) = \sum_{i \in V} \alpha_i^{m_i} \cdot \phi(I,l_i) + \sum_{ij \in E} \beta_{ij}^{m_i m_j} \cdot \psi(l_i,l_j) + S(M)$$

Given:
- Image (I)

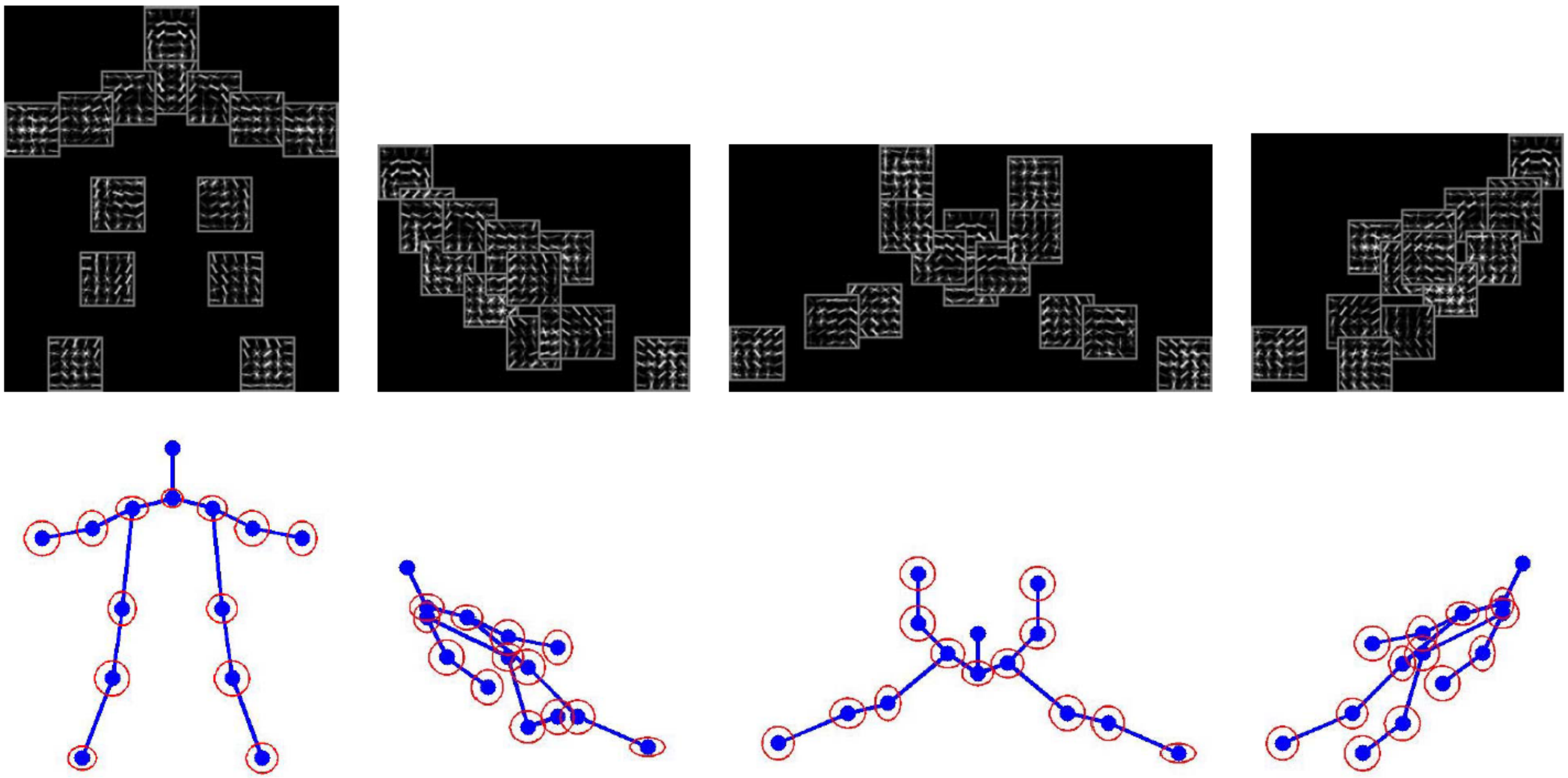
Need to compute
- Part locations (L)
- Part mixtures (M)

Algorithm
- (L*,M*) = arg max (S(I,L,M))

Standard inference problem
- For tree graphs, can be exactly computed using belief propagation

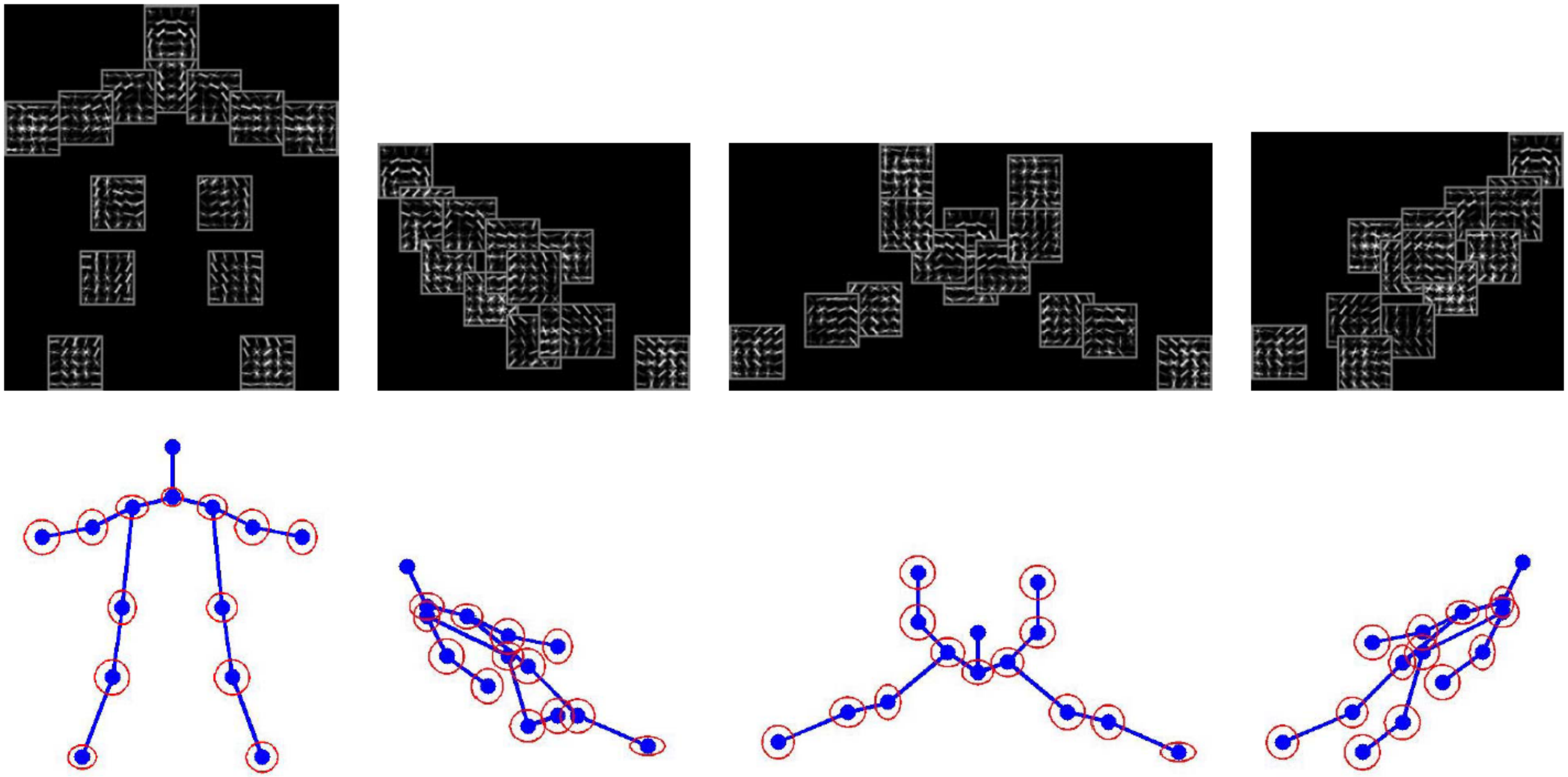
Results

Achieving articulation



Slide taken from authors, Yang et al.

Achieving articulation

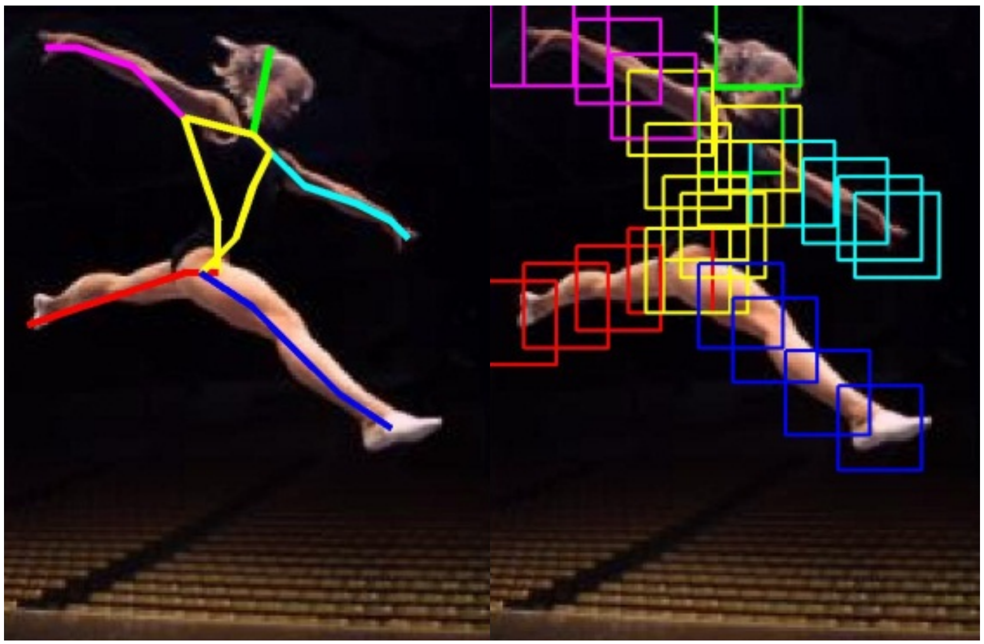


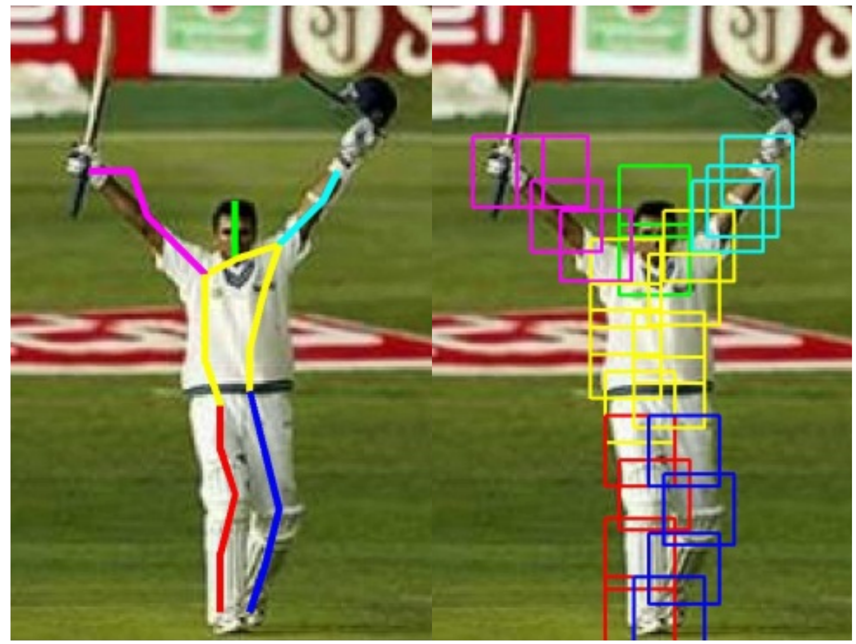
- K parts, M mixtures $\Rightarrow K^M$ unique pictorial structures

Not all are equally likely --- "prior" given by $S(M)$

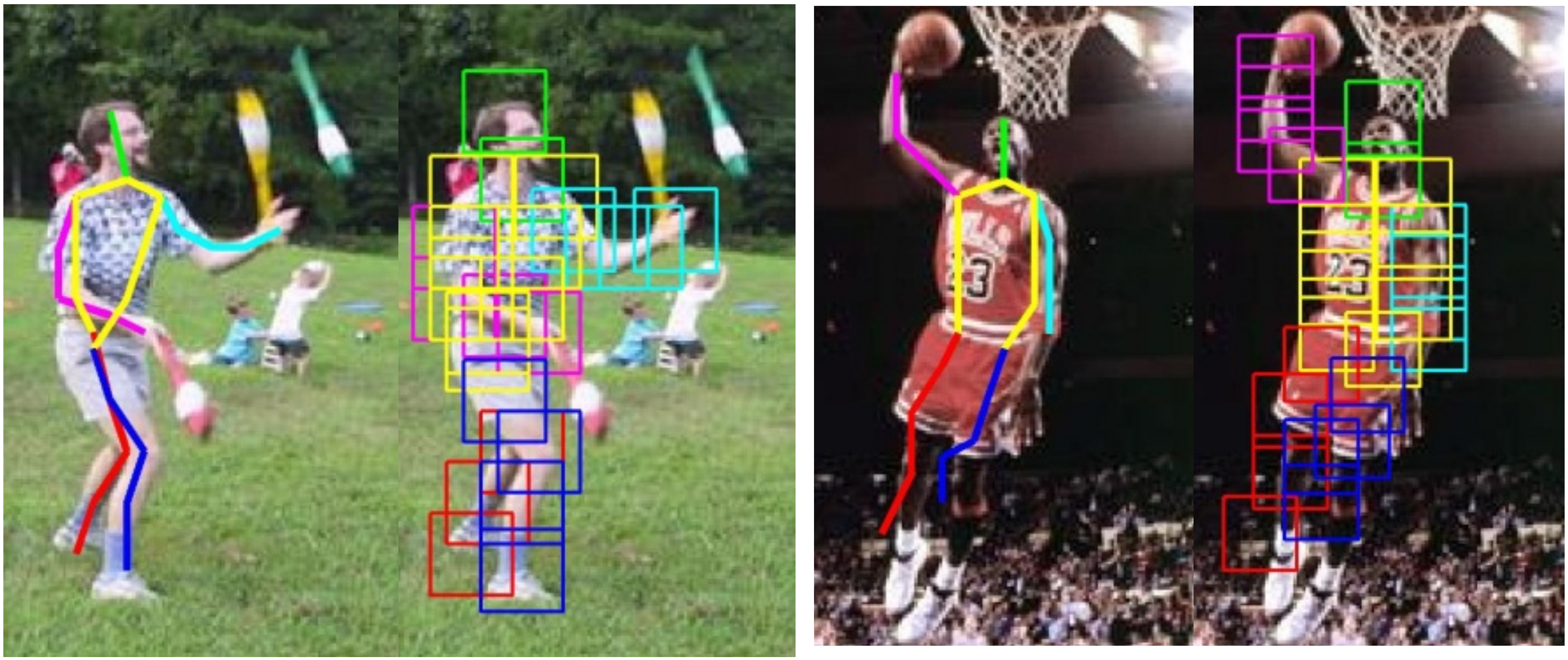
Slide taken from authors, Yang et al.

Qualitative Results

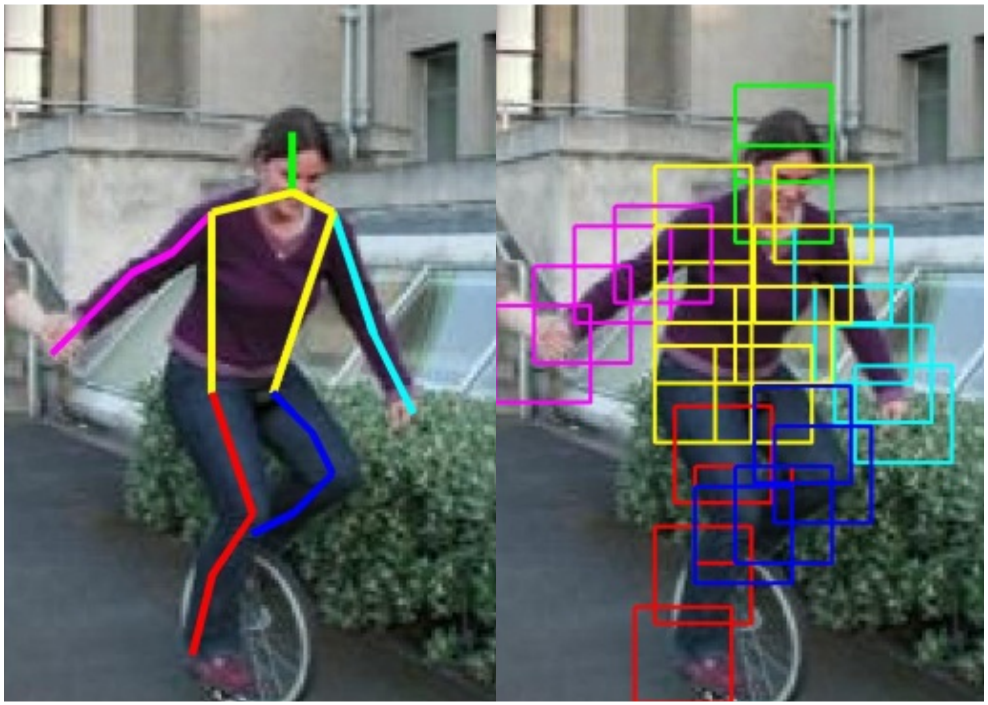


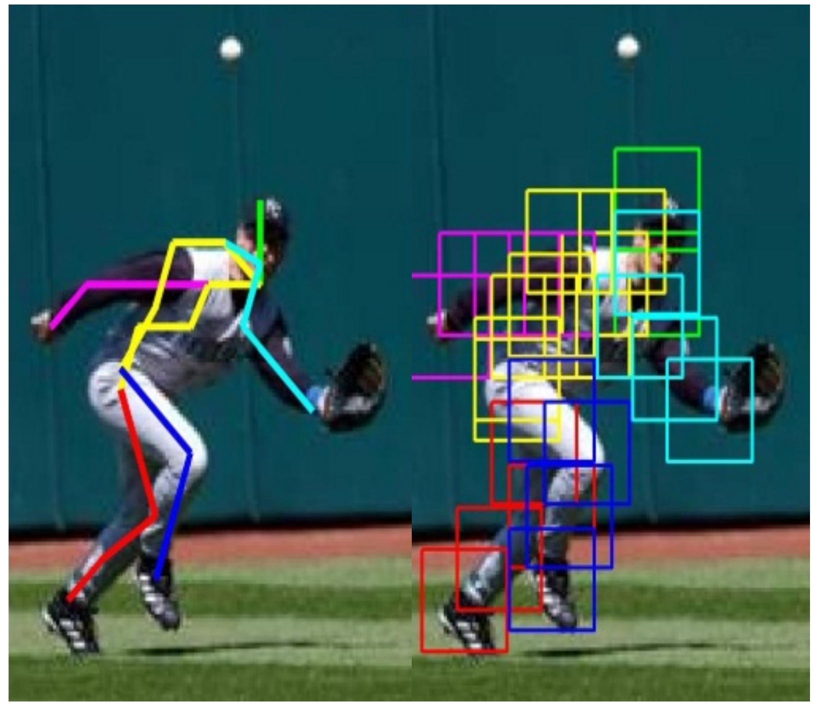


Qualitative Results

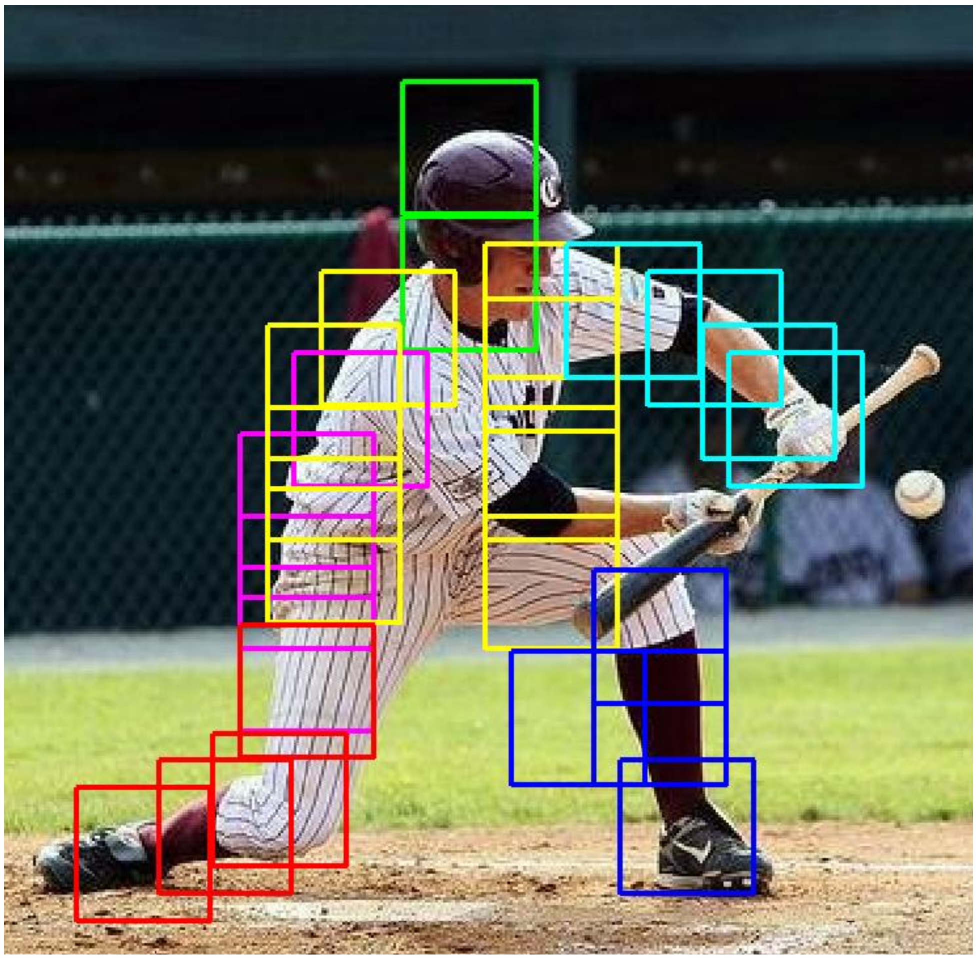


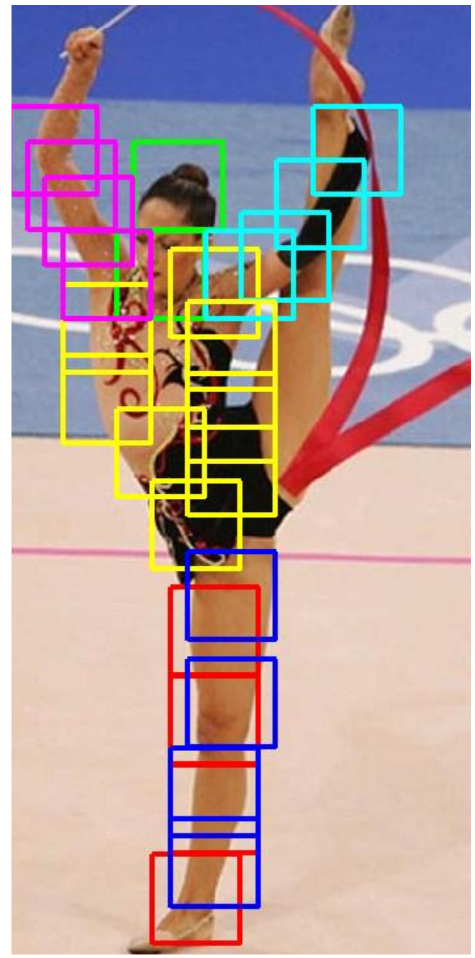
Qualitative Results



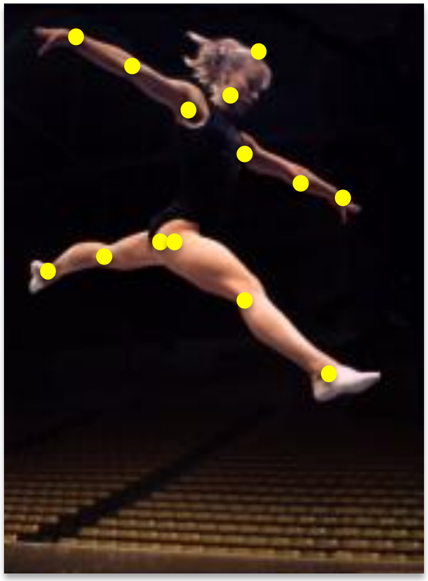


Failure cases





Benchmark Datasets

PARSE Full-body	BUFFY Upper-body
http://www.ics.uci.edu/~dramanan/papers/parse/index.html 	**http://www.robots.ox.ac.uk/~vgg/data/stickmen/index.html** 

Slide taken from authors, Yang et al.

Quantitative Results on PARSE

% of correctly localized limbs

Image Parse Testset

Method	Head	Torso	U. Legs	L. Legs	U. Arms	L. Arms	Total
Ramanan 2007	52.1	37.5	31.0	29.0	17.5	13.6	27.2
Andrikluka 2009	81.4	75.6	63.2	55.1	47.6	31.7	55.2
Johnson 2009	77.6	68.8	61.5	54.9	53.2	39.3	56.4
Singh 2010	91.2	76.6	71.5	64.9	50.0	34.2	60.9
Johnson 2010	85.4	76.1	73.4	65.4	64.7	46.9	66.2
Our Model	**97.6**	**93.2**	**83.9**	**75.1**	**72.0**	**48.3**	**74.9**

1 second per image

Slide taken from authors, Yang et al.

Quantitative Results on BUFFY

% of correctly localized limbs

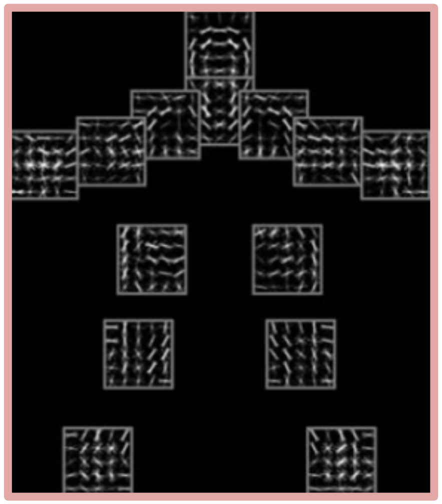
Subset of Buffy Testset

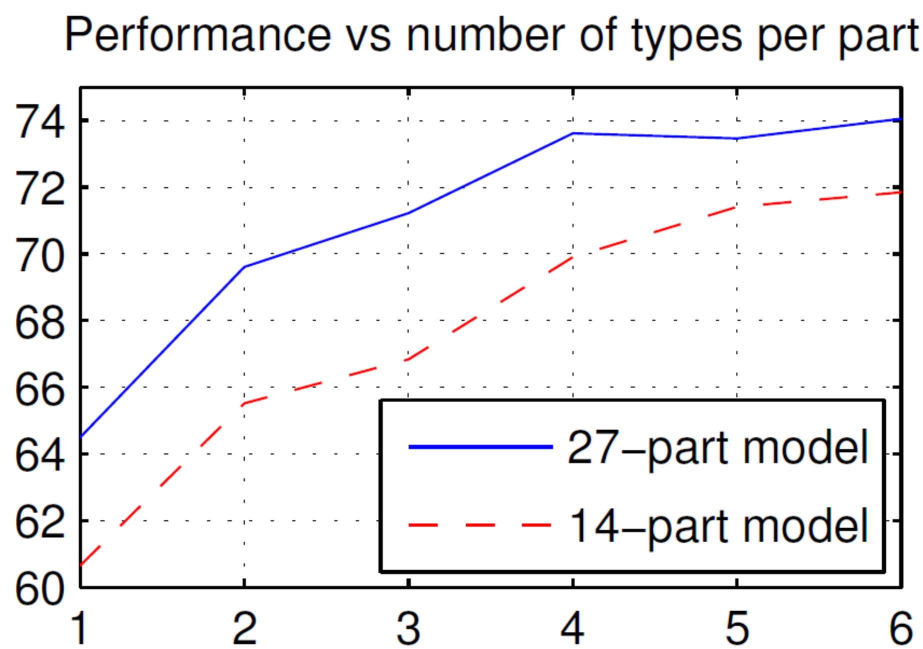
Method	Head	Torso	U. Arms	L. Arms	Total
Tran 2010	---	---	---	---	62.3
Andrikluka 2009	90.7	95.5	79.3	41.2	73.5
Eichner 2009	98.7	97.9	82.8	59.8	80.1
Sapp 2010a	100	**100**	91.1	65.7	85.9
Sapp 2010b	100	96.2	95.3	63.0	85.5
Our Model	**100**	99.6	**96.6**	**70.9**	**89.1**

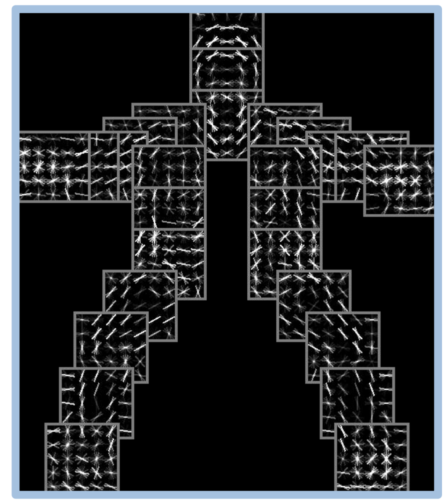
All previous work use explicitly articulated models

Slide taken from authors, Yang et al.

More Parts and Mixtures Help







14 parts (joints)

27 parts (joints + midpoints)

Slide taken from authors, Yang et al.

Discussion

- Possible limitations?
- Something other than human pose estimation?
- Can do more useful things with the Kinect?
- Can this encode occlusions well?

References

- http://phoenix.ics.uci.edu/software/pose/
- Code and benchmark datasets available