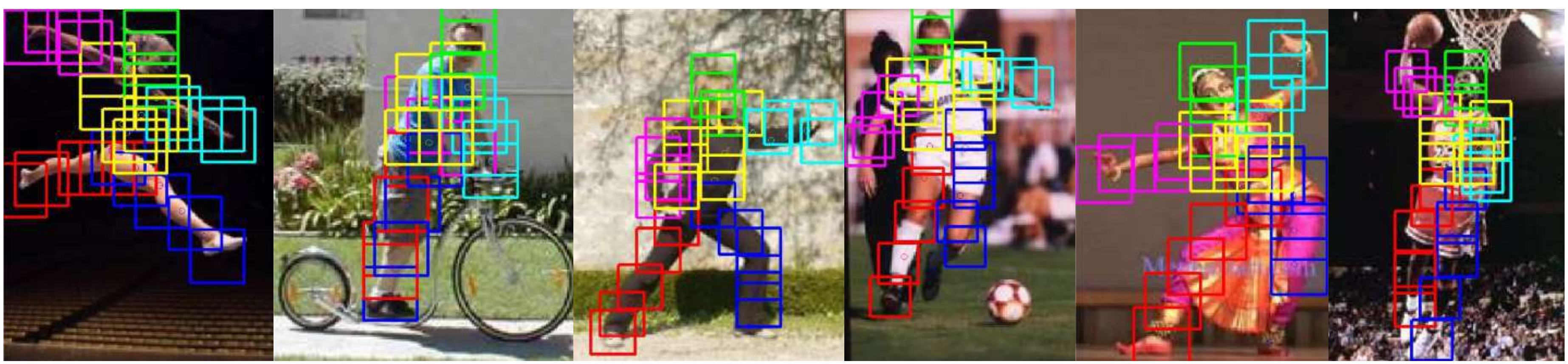
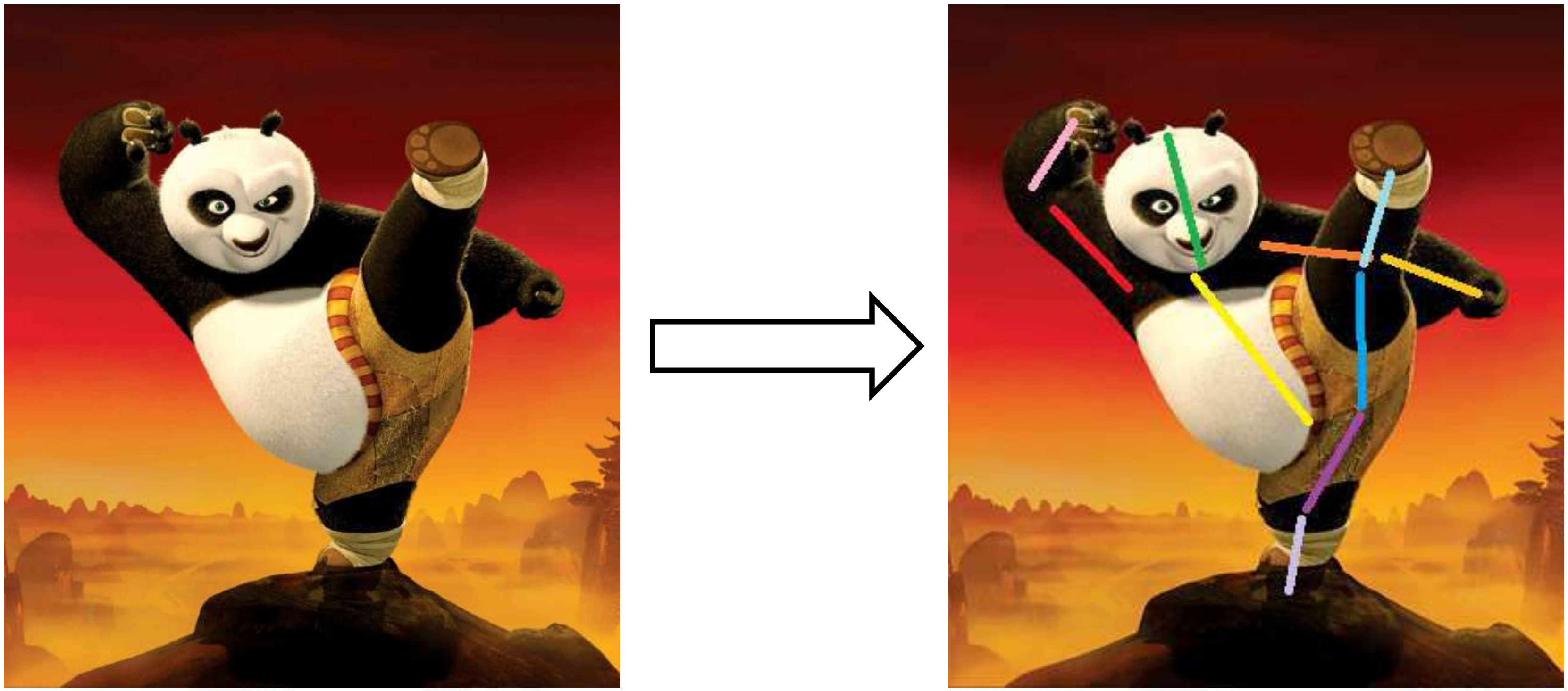
Articulated Pose Estimation with Flexible Mixtures of Parts

Yi Yang & Deva Ramanan

University of California, Irvine



Goal

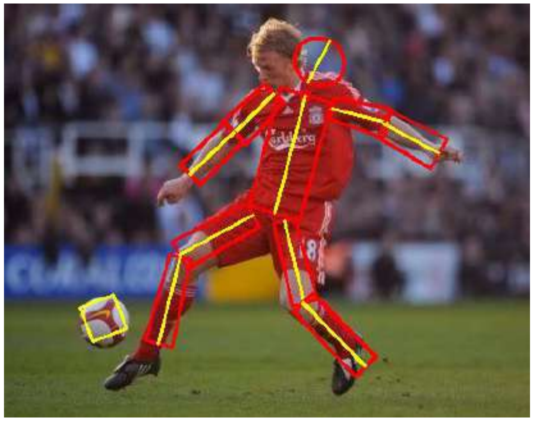


Articulated pose estimation ( )

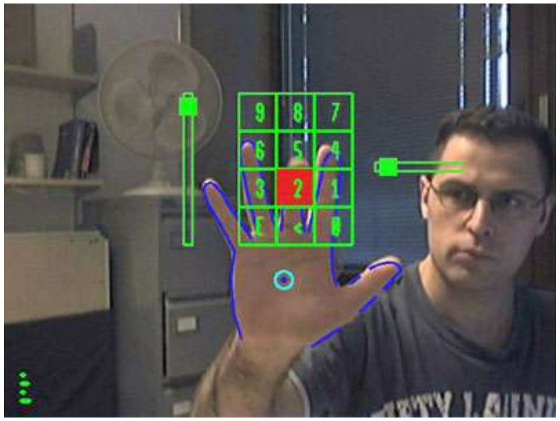
recovers the pose of an articulated object which consists of joints and rigid parts

Applications

Action

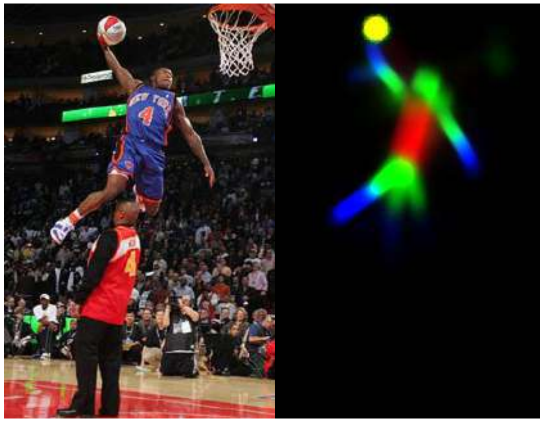


HCI



Gaming

Segmentation

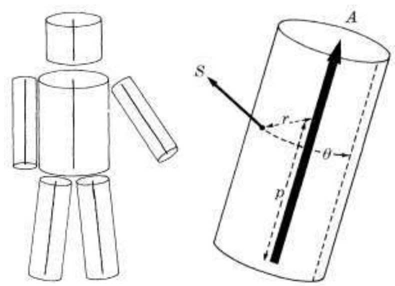


Object

......

Unconstrained Images

Classic Approach

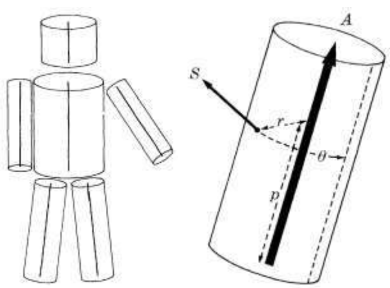


Marr & Nishihara 1978

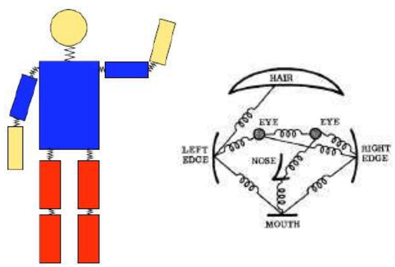
Part Representation

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

Classic Approach



Marr & Nishihara 1978



Fischler & Elschlager 1973
Felzenszwalb & Huttenlocher 2005

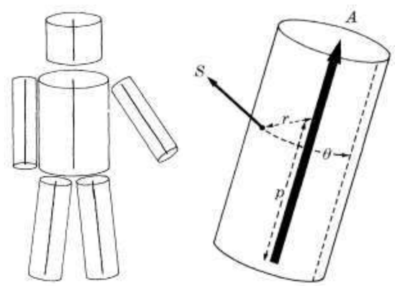
Part Representation

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

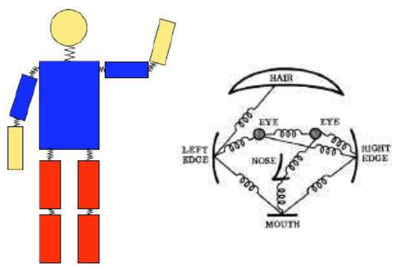
Pictorial Structure

- Unary Templates
- Pairwise Springs

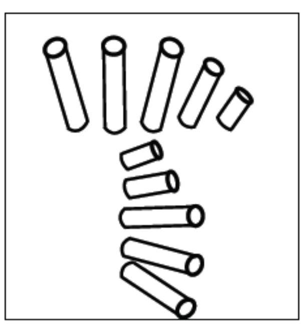
Classic Approach



Marr & Nishihara 1978



Fischler & Elschlager 1973
Felzenszwalb & Huttenlocher 2005



Part Representation

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

Pictorial Structure

- Unary Templates
- Pairwise Springs

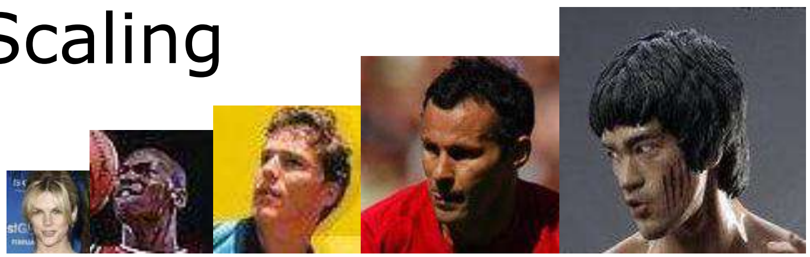
Lan & Huttenlocher 2005
Sigal & Black 2006
Ramanan 2007
Epshteian & Ullman 2007
Wang & Mori 2008
Ferrari etc. 2008

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

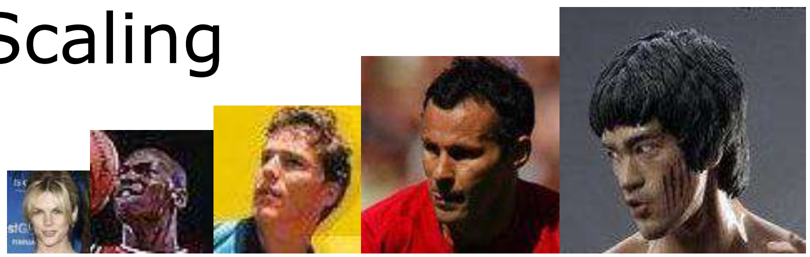
Problem: Wide Variations

In-plane rotation	Foreshortening
	

Problem: Wide Variations

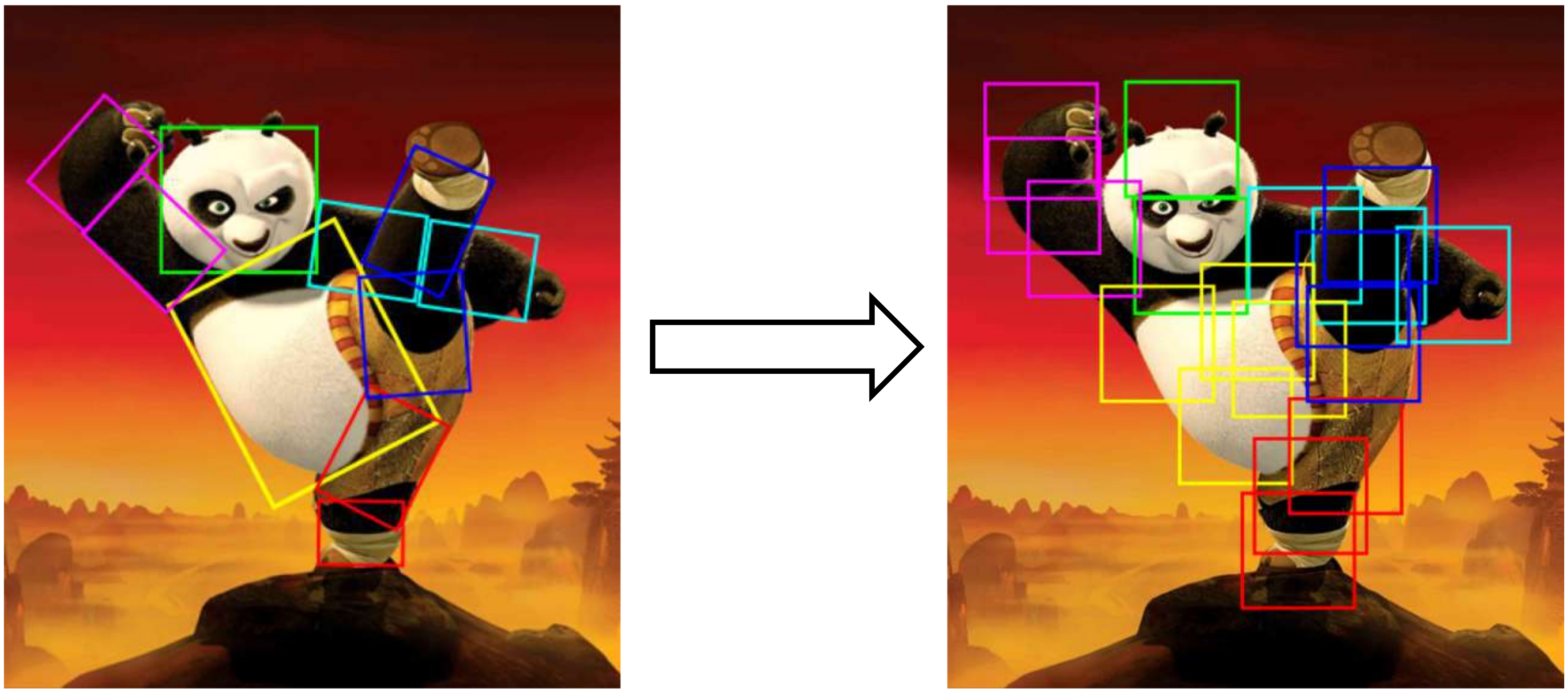
In-plane rotation 	Foreshortening 
Scaling 	Out-of-plane rotation 
Intra-category variation 	Aspect ratio 

Problem: Wide Variations

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

Naïve brute-force evaluation is expensive

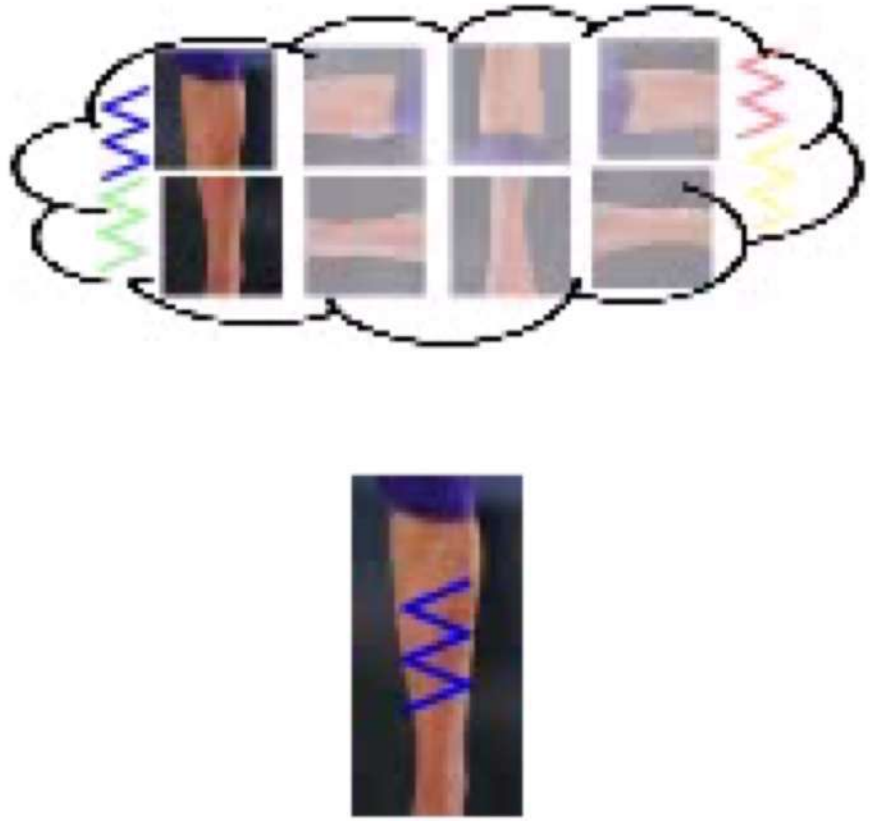
Our Method – “Mini-Parts”



Key idea:

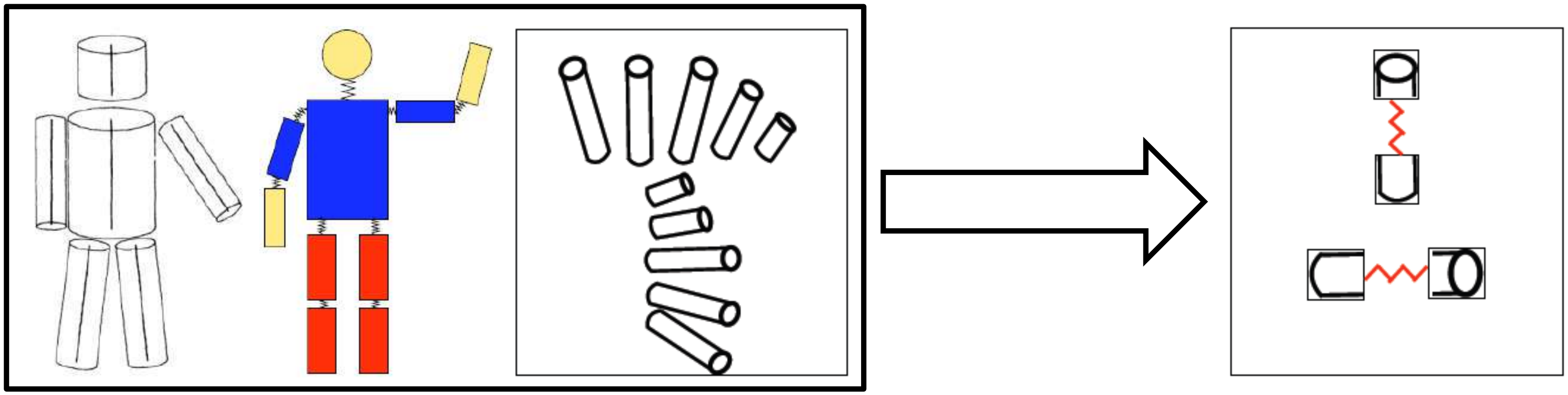
“mini part” model can approximate deformations

Example: Arm Approximation



Example: Torso Approximation

Key Advantages



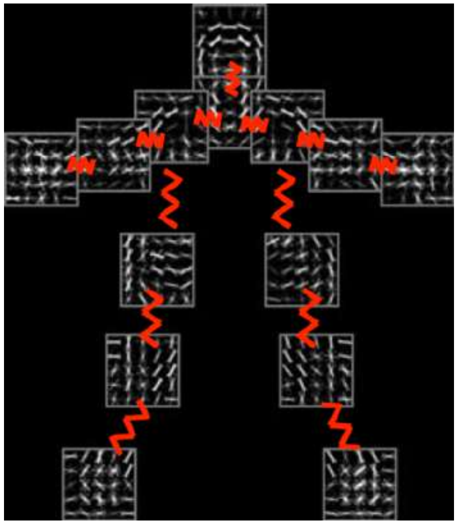
- Flexibility:

General affine warps (orientation, foreshortening, …)

- Speed:

Mixtures of local templates + dynamic programming

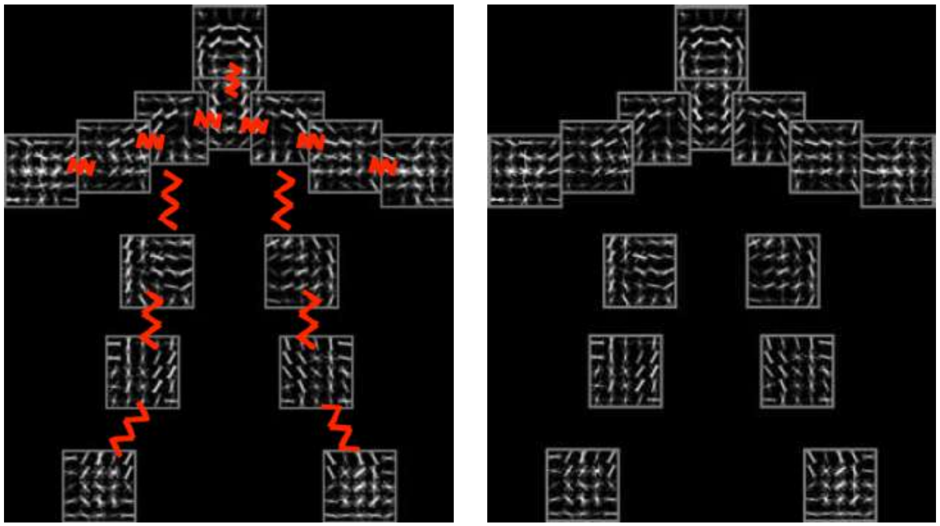
Pictorial Structure Model



$S(I, L)$

- I: Image
- l_i: Location of part i

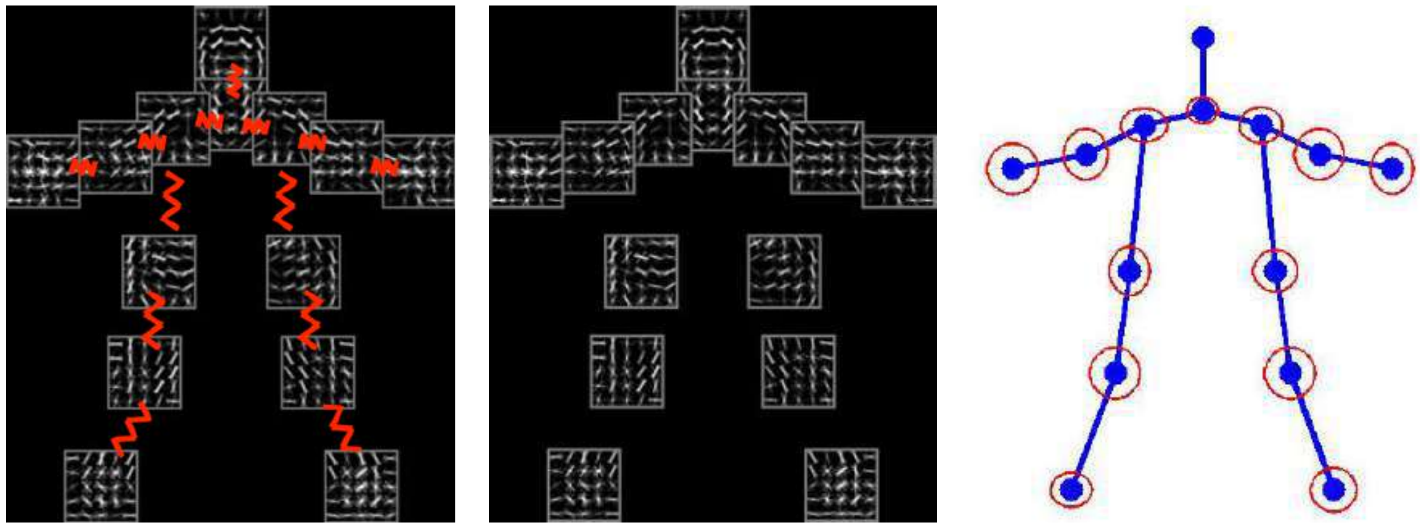
Pictorial Structure Model



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

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

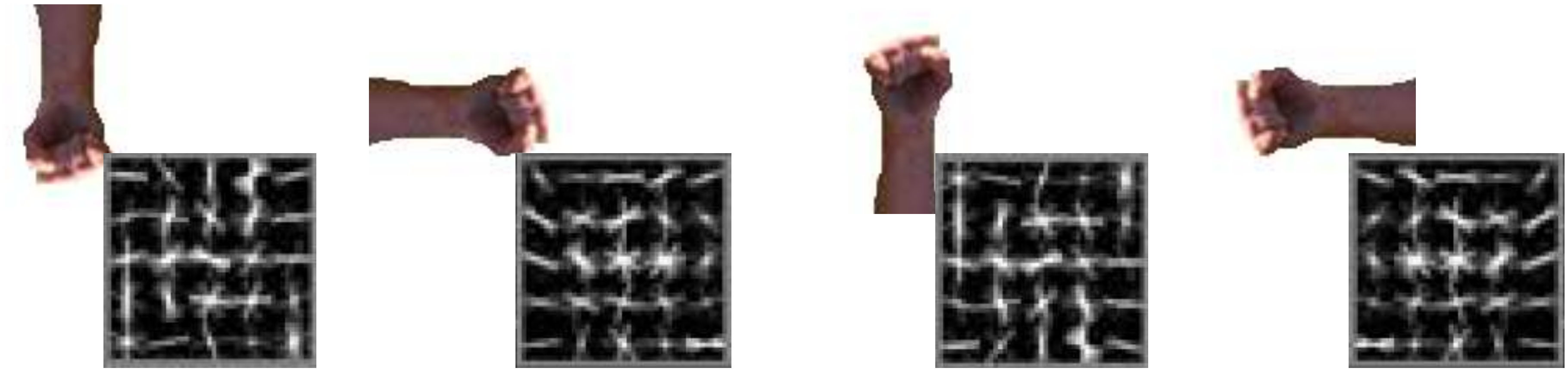
Pictorial Structure Model



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

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

Our Flexible Mixture Model

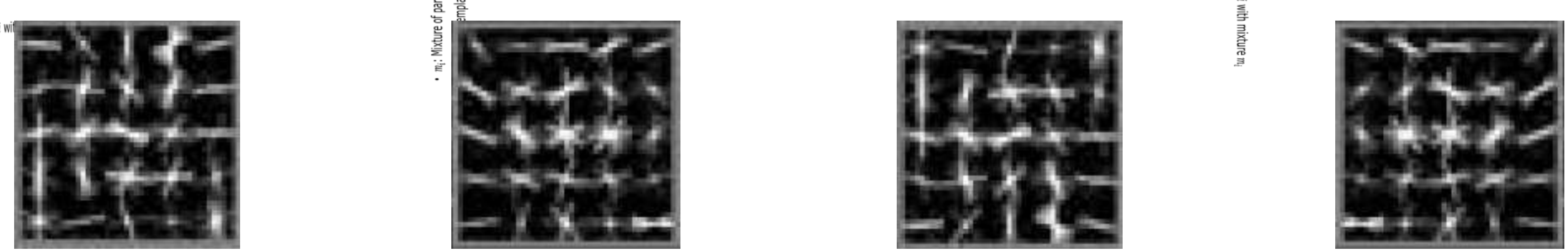


$$S(I,L,M) = \sum_{i \in V} \alpha_i^{m_i} \cdot \phi(I,l_i) + \sum_{ij \in E} \beta_{ij}^{m_i m_j} \cdot \psi(l_i,l_j)$$

- m_i: Mixture of part i

Our Flexible Mixture Model

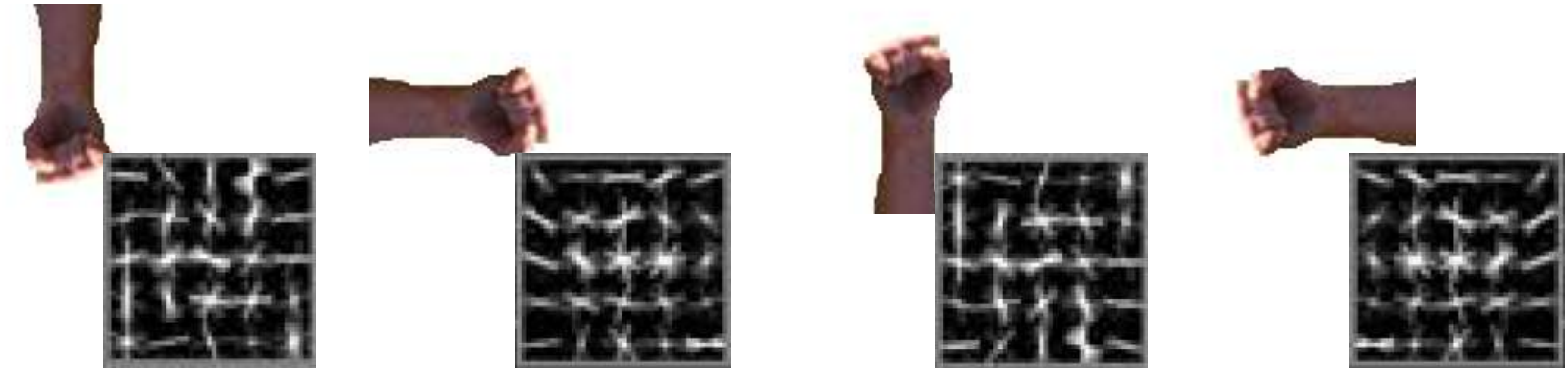
- m_i: Mixture of part i
- $\alpha_i^{m_i}$: Unary template for part i wi



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

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

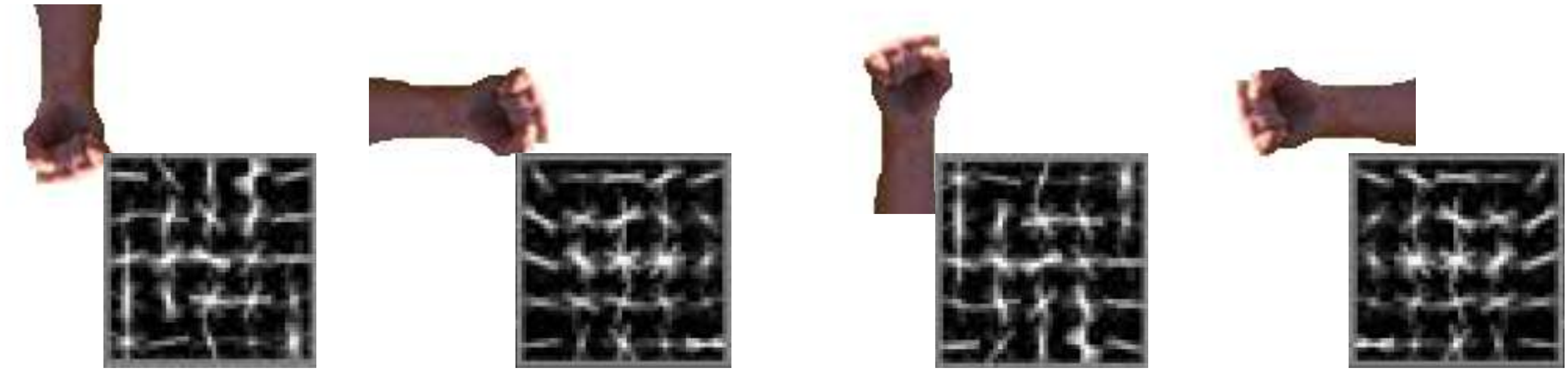
Our Flexible Mixture Model



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

- 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

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

Co-occurrence “Bias”

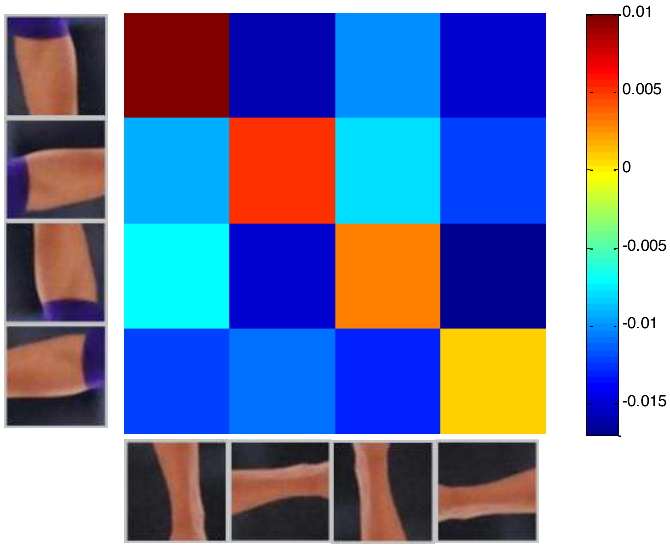
$$S(M) = \sum_{ij \in E} b_{ij}^{m_i m_j}$$

- $b_{ij}^{m_i m_j}$: Pairwise co-occurrence prior between part i with mixture m_i and part j with mixture m_j

Co-occurrence "Bias"

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

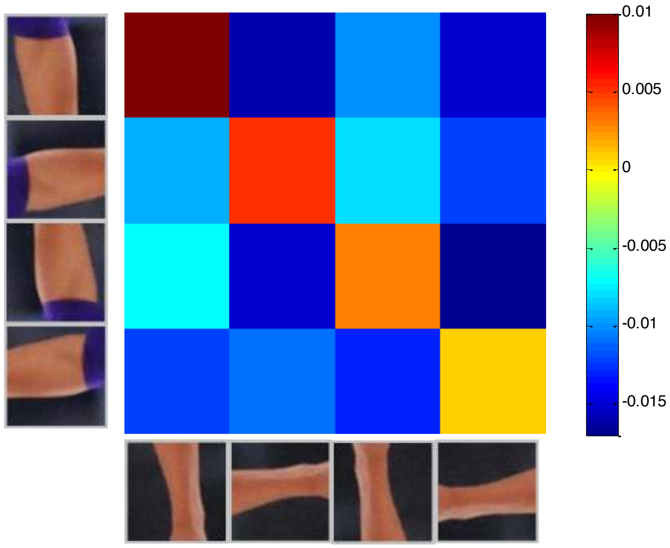
- $b_{ij}^{m_i m_j}$: Pairwise co-occurrence prior between part i with mixture m_i and part j with mixture m_j

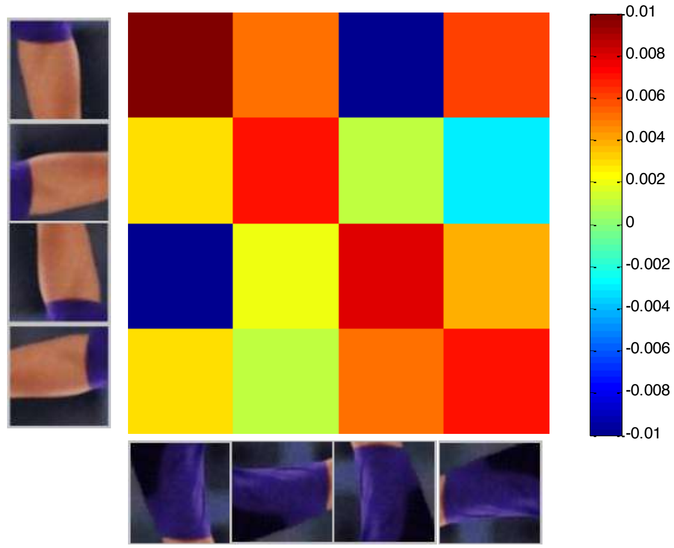


Co-occurrence "Bias"

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

- $b_{ij}^{m_i m_j}$: Pairwise co-occurrence prior between part i with mixture m_i and part j with mixture m_j





Inference & Learning

Inference

$$\max_{L,M} S(I, L, M)$$

For a tree graph (V,E): dynamic programming

Inference & Learning

Inference

$$\max_{L,M} S(I, L, M)$$

For a tree graph (V,E): dynamic programming

Learning

$$\min_{w} \frac{1}{2} \|w\|$$
$$\text{s.t.} \quad \forall n \in \text{pos} \;\; w \cdot \phi(I_n, z_n) \geq 1$$
$$\forall n \in \text{neg}, \forall z \;\; w \cdot \phi(I_n, z) \leq -1$$

Given labeled positive $\{I_n, L_n, M_n\}$ and negative $\{I_n\}$, write $z_n = (L_n, M_n)$, and $S(I, z) = w \cdot \phi(I, z)$

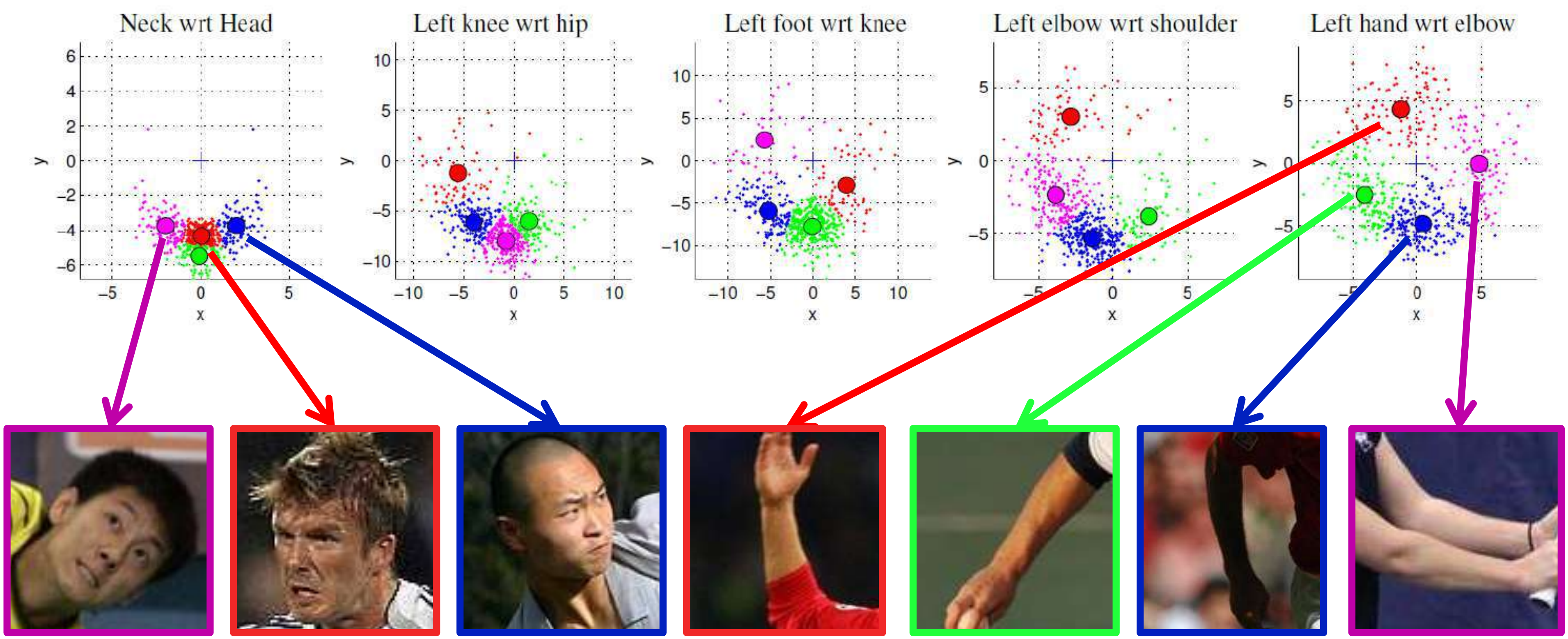
Benchmark Datasets

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

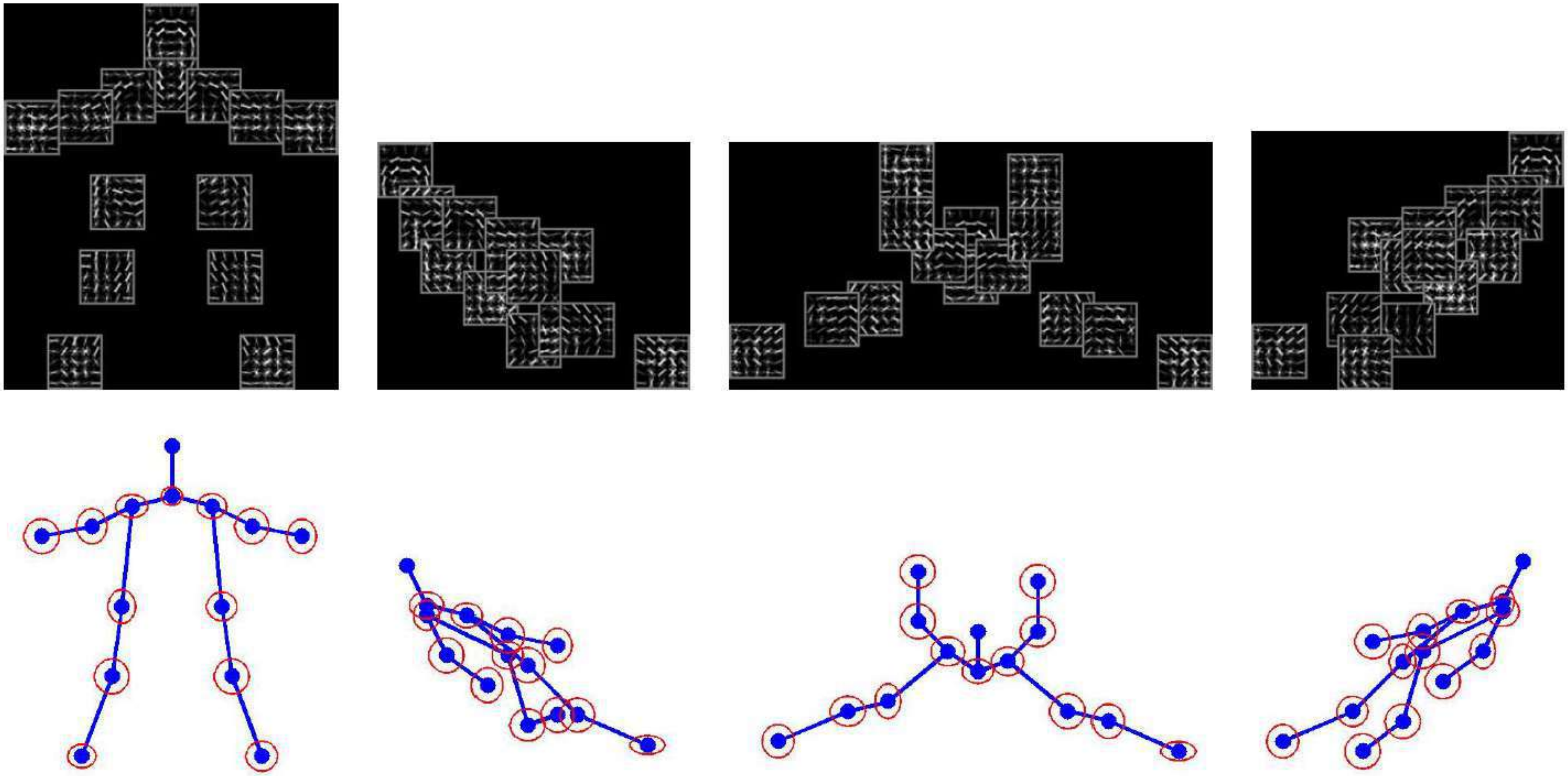
How to Get Part Mixtures?

Solution:

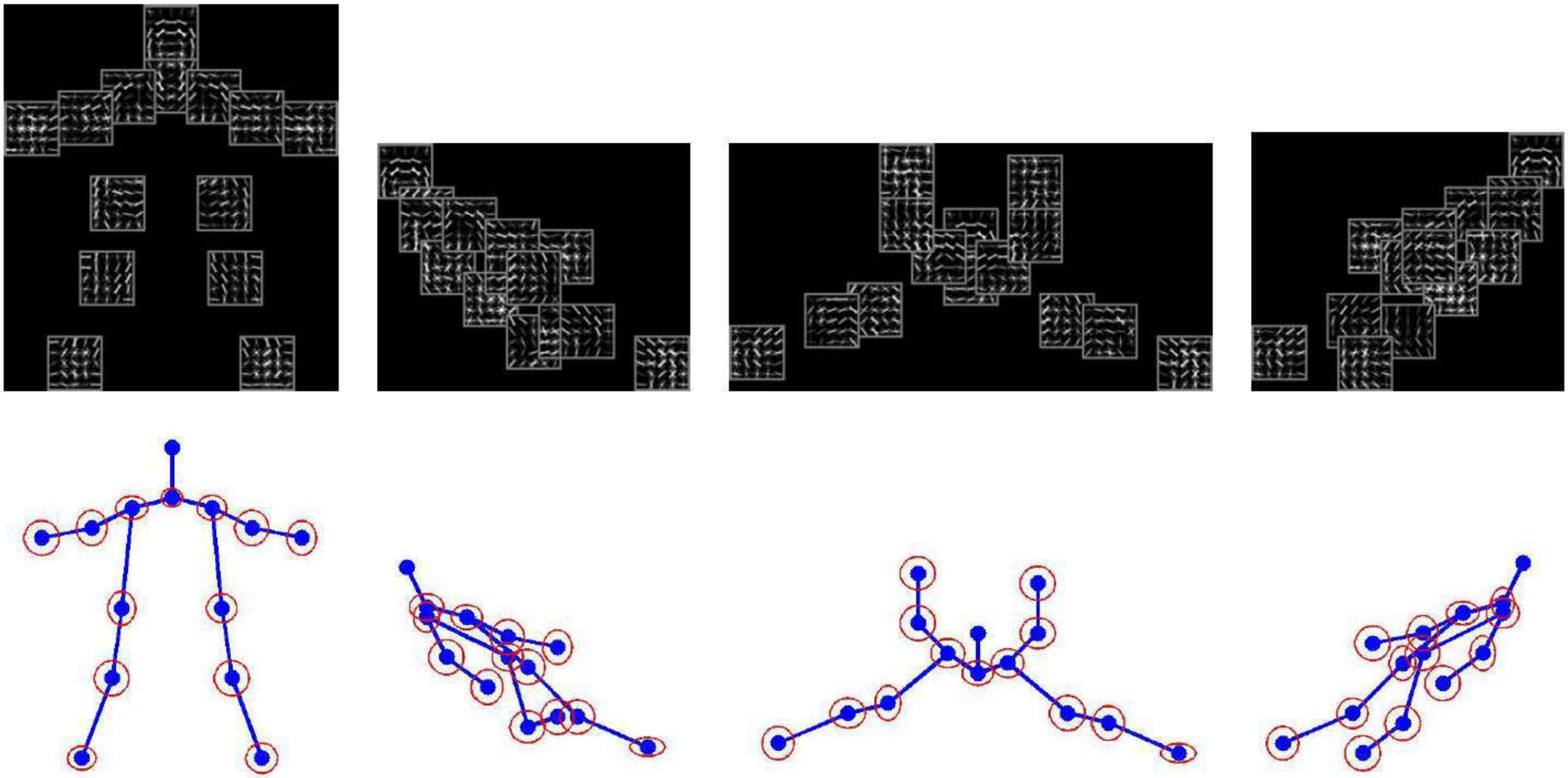
Cluster relative locations of joints w.r.t. parents



Articulation

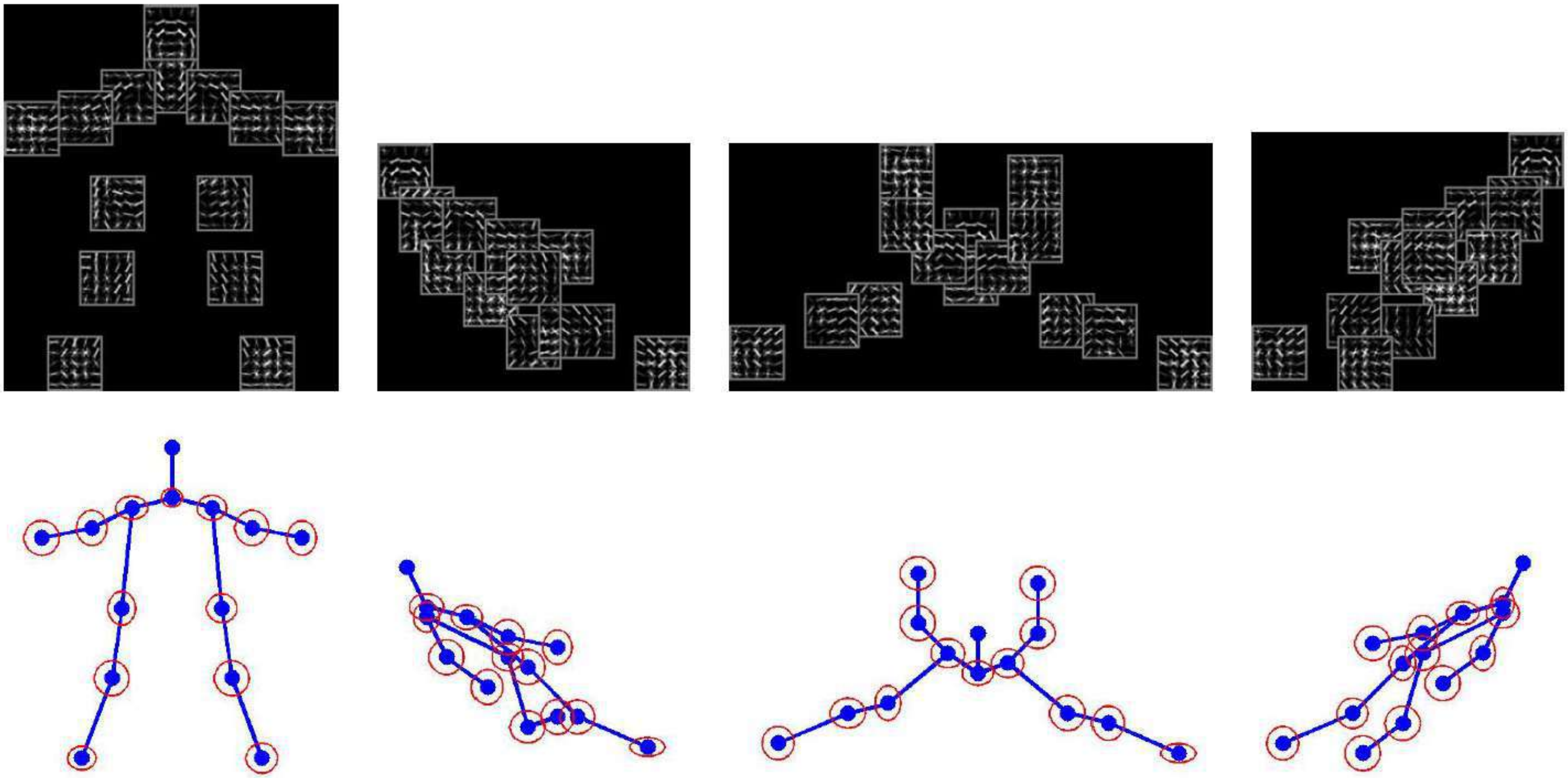


Articulation



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

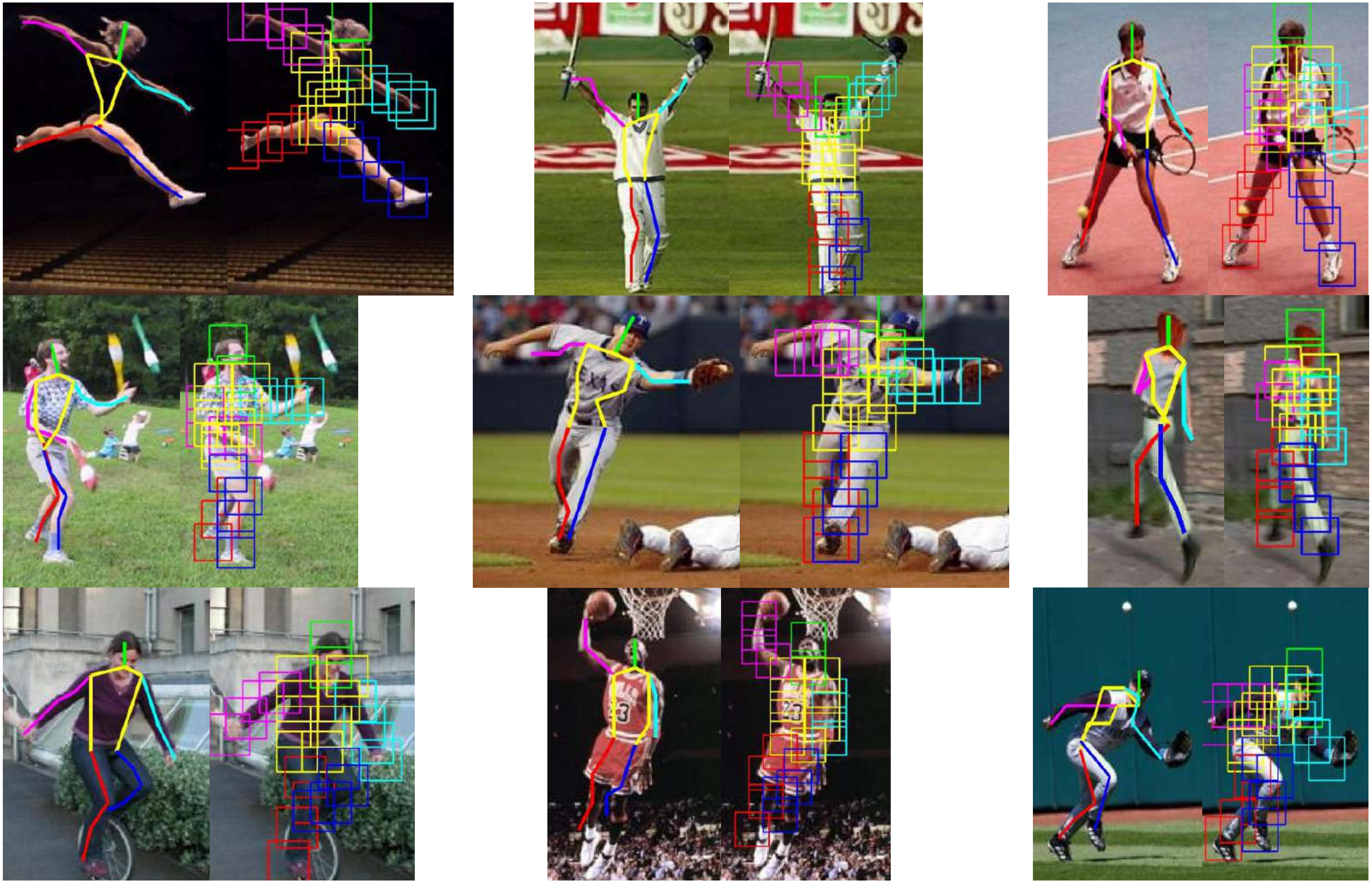
Articulation



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

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

Qualitative Results



Quantitative Results on PARSE

% of correctly localized limbs

Image Parse Testset

Method							Total
Ramanan 2007							27.2
Andrikluka 2009							55.2
Johnson 2009							56.4
Singh 2010							60.9
Johnson 2010							66.2
Our Model							**74.9**

All previous work use explicitly articulated models

Quantitative Results on PARSE

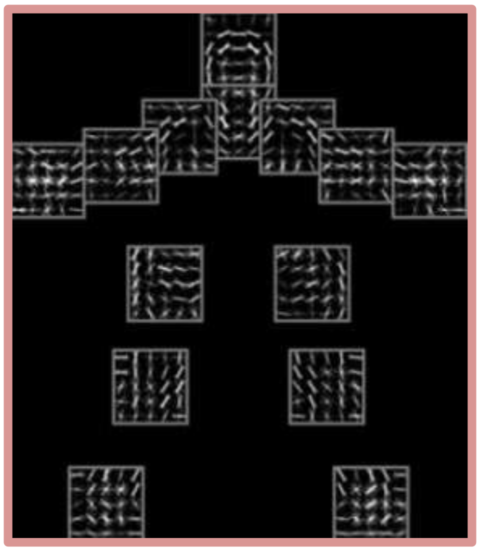
% of correctly localized limbs

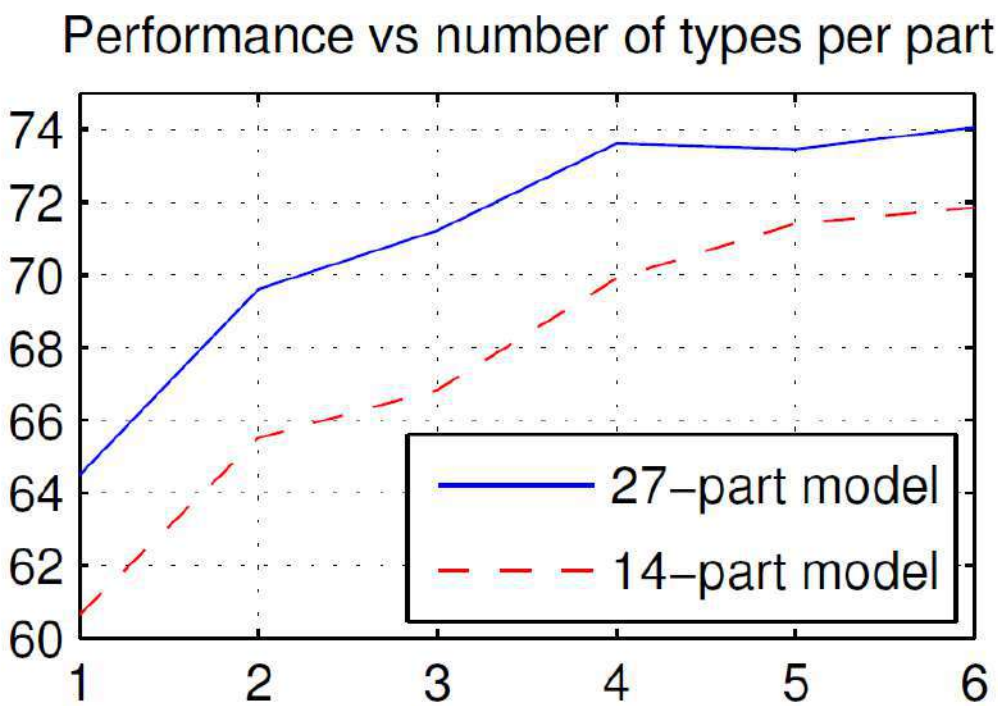
Image Parse Testset

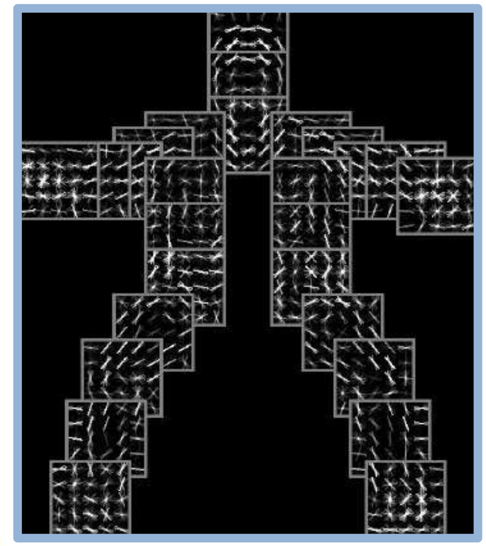
Method	Head	Torso	U. Legs	L. Legs	U. Arms	L. Arms	Total
Ramanan 2007	52.1	37.5	31.0	29.0	17.5	13.6	27.2
Andrikluka 2009	81.4	75.6	63.2	55.1	47.6	31.7	55.2
Johnson 2009	77.6	68.8	61.5	54.9	53.2	39.3	56.4
Singh 2010	91.2	76.6	71.5	64.9	50.0	34.2	60.9
Johnson 2010	85.4	76.1	73.4	65.4	64.7	46.9	66.2
Our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

1 second per image

More Parts and Mixtures Help







14 parts (joints)

27 parts (joints + midpoints)

Quantitative Results on BUFFY

% of correctly localized limbs

Subset of Buffy Testset

Method					Total
Tran 2010					62.3
Andrikluka 2009					73.5
Eichner 2009					80.1
Sapp 2010a					85.9
Sapp 2010b					85.5
Our Model					**89.1**

Our algorithm = 5 seconds -vs- Next best = 5 minutes

Quantitative Results on BUFFY

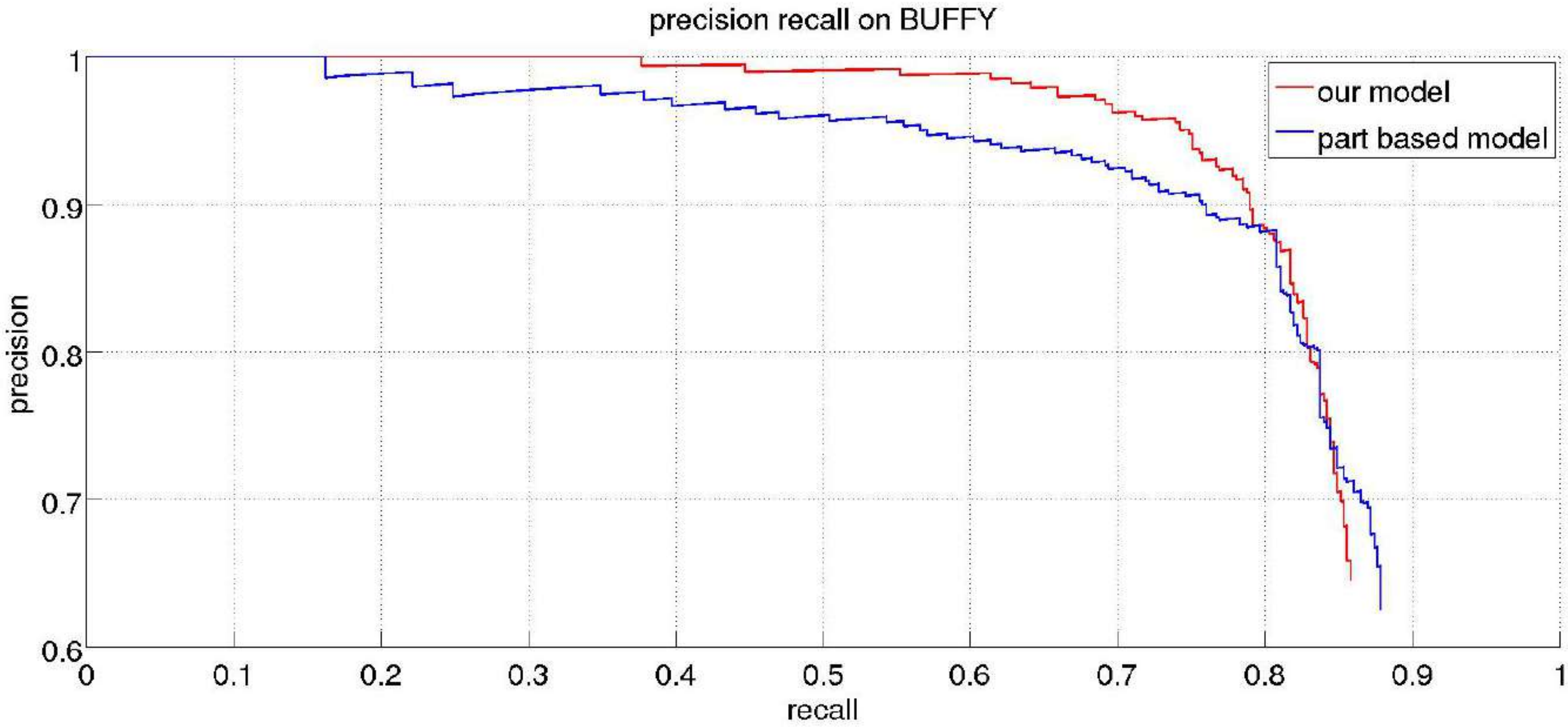
% of correctly localized limbs

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**

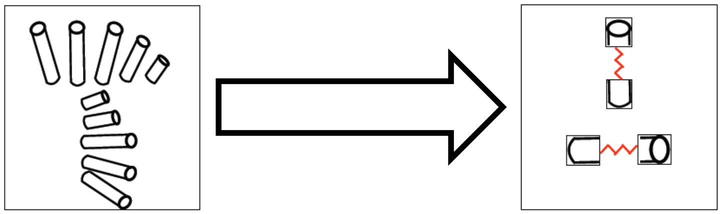
All previous work use explicitly articulated models

Human Detection



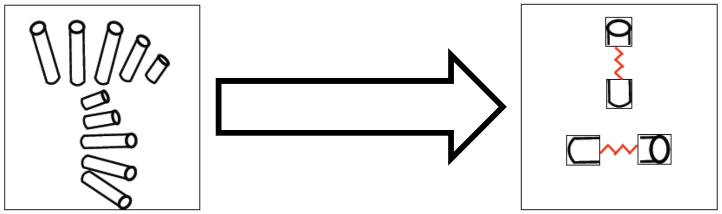
Conclusion

- Model affine warps with a part-based model

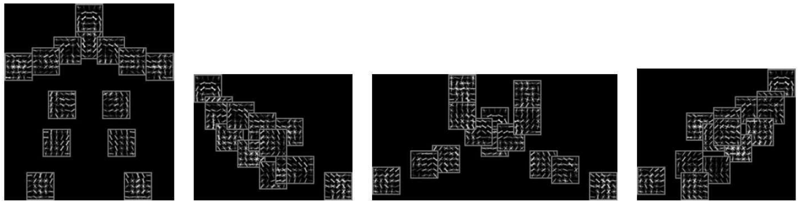


Conclusion

- Model affine warps with a part-based model

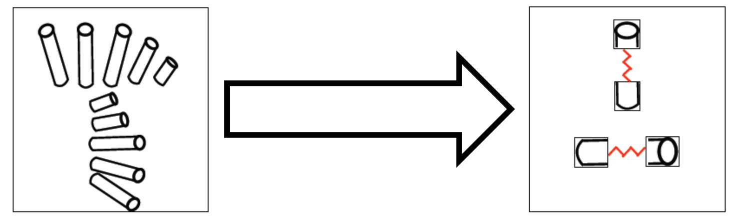


- Exponential set of pictorial structures



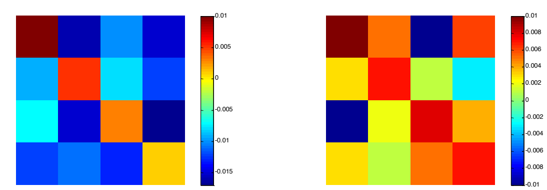
Conclusion

- Model affine warps with a part-based model



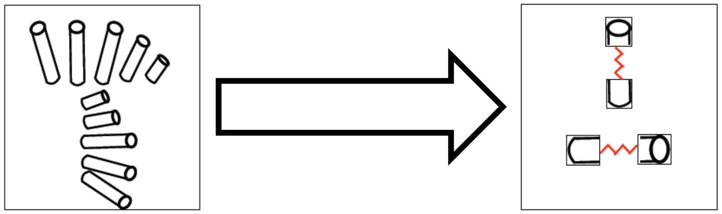
- Exponential set of pictorial structures

- Flexible vs rigid relations

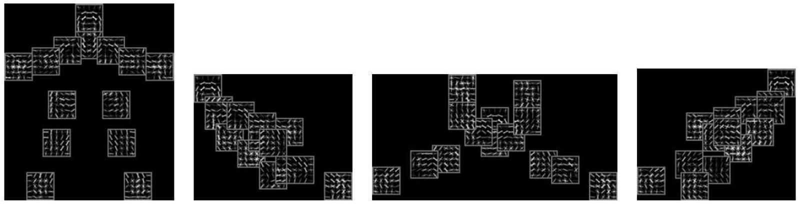


Conclusion

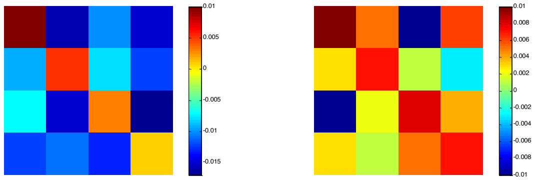
- Model affine warps with a part-based model



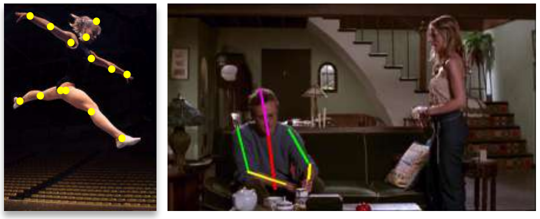
- Exponential set of pictorial structures



- Flexible vs rigid relations



- Supervision helps



Thank you