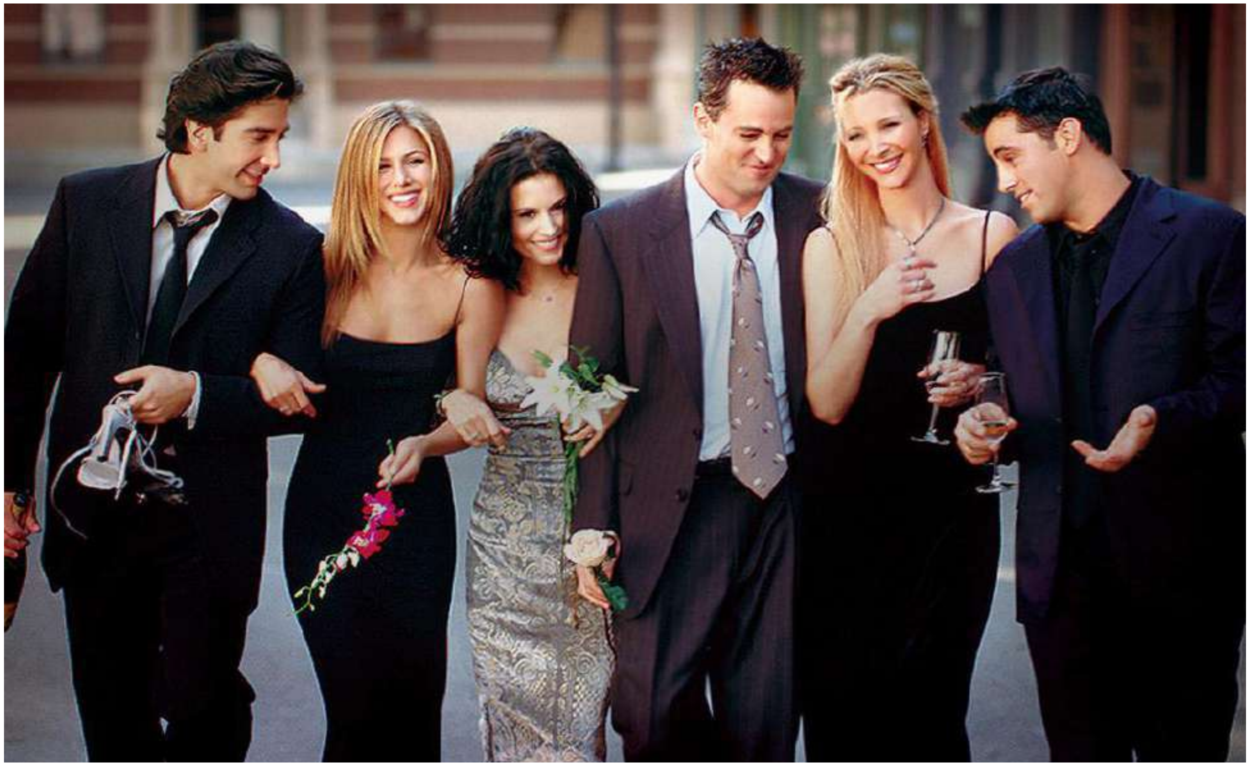
Proxemics Recognition



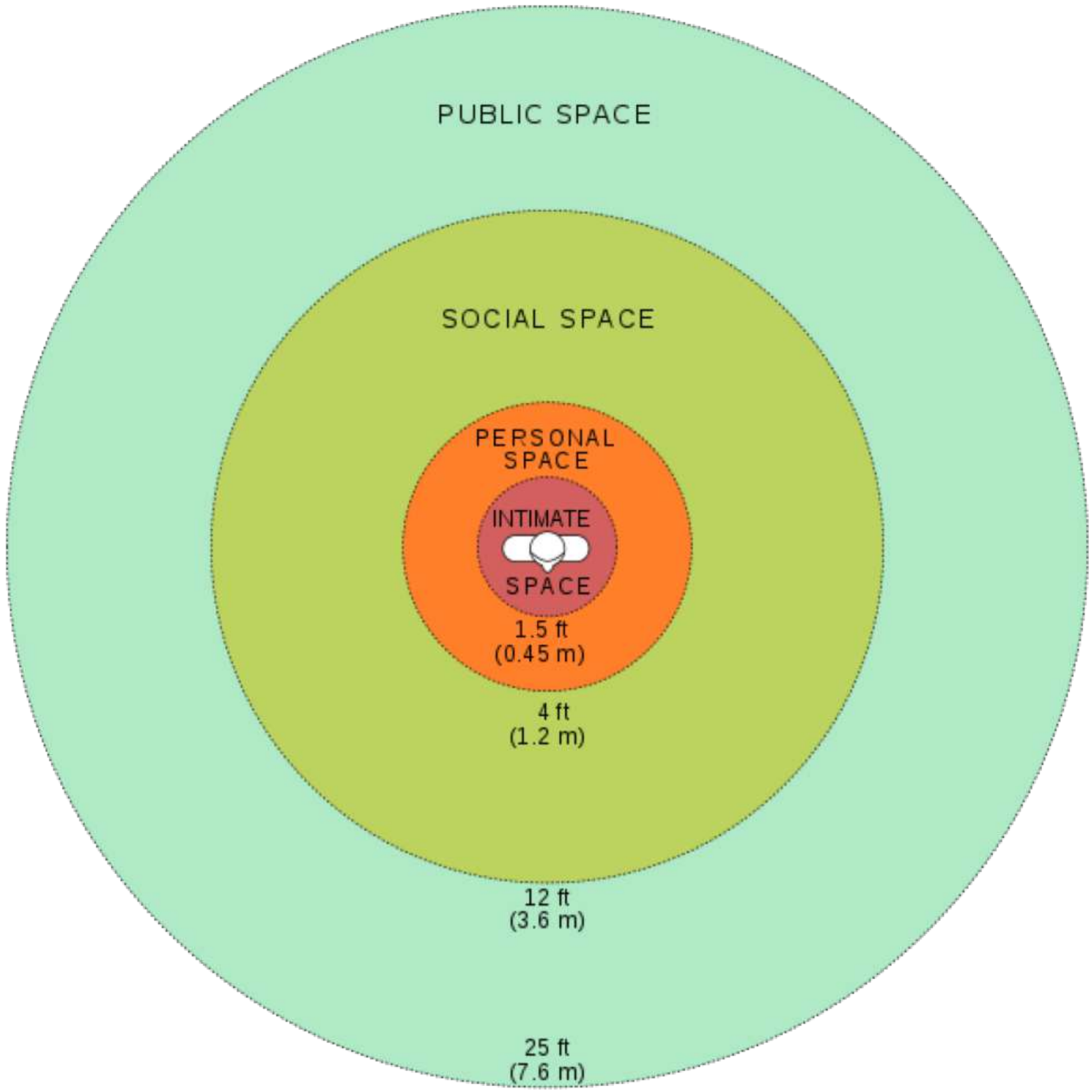
Yi Yang[1], Simon Baker, Anitha Kannan, Deva Ramanan[1]

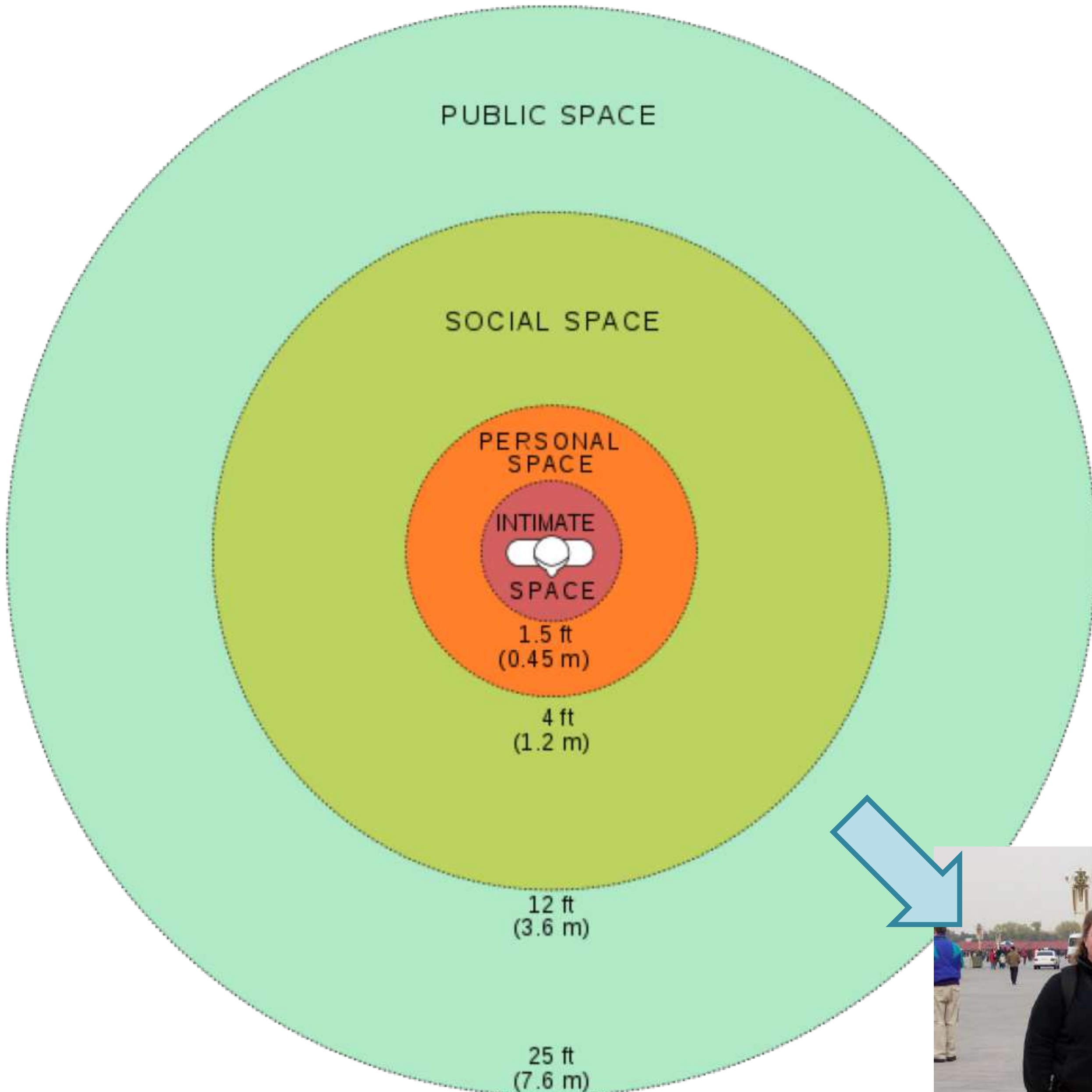
[1]Department of Computer Science, UC Irvine

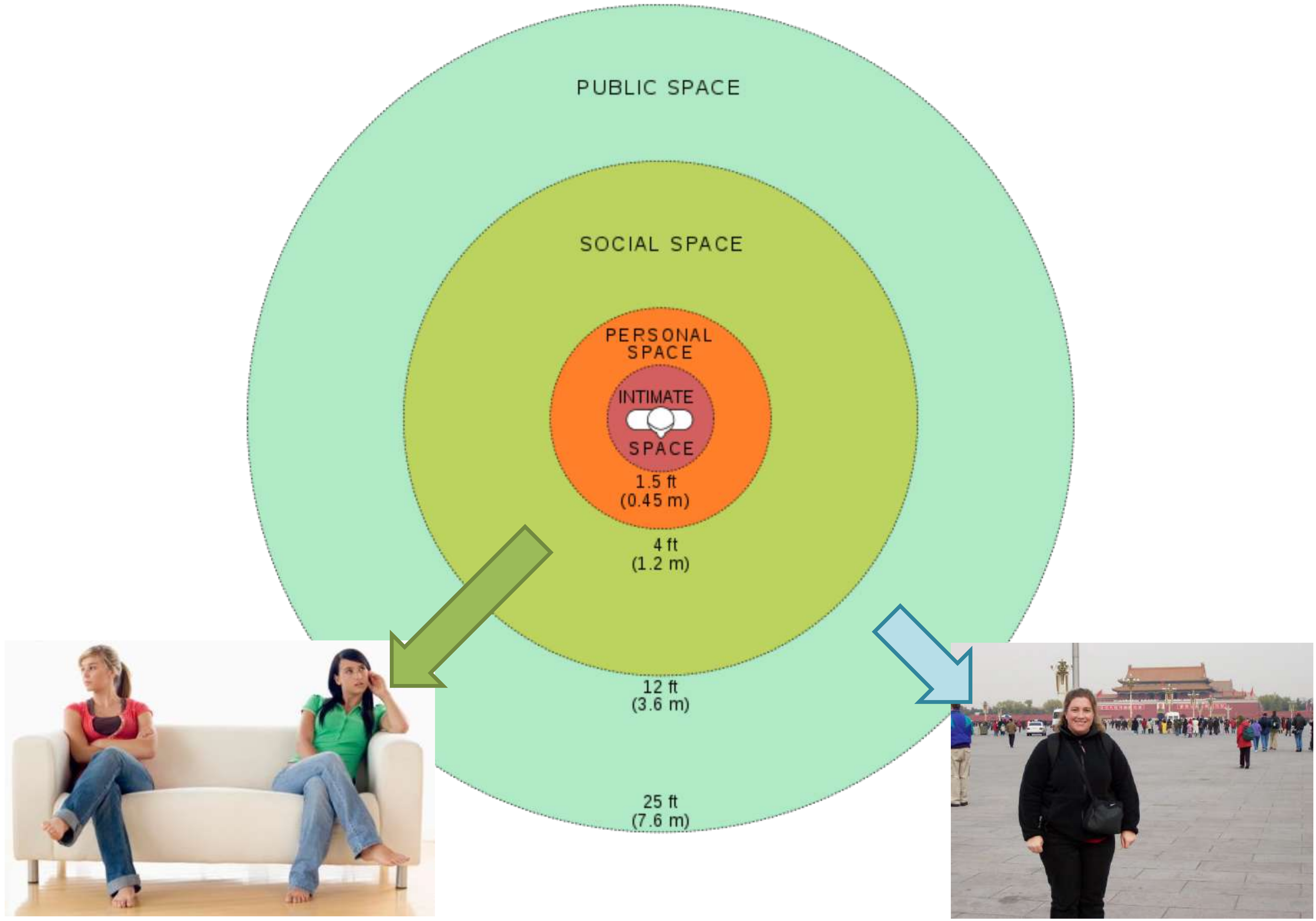
Proxemics

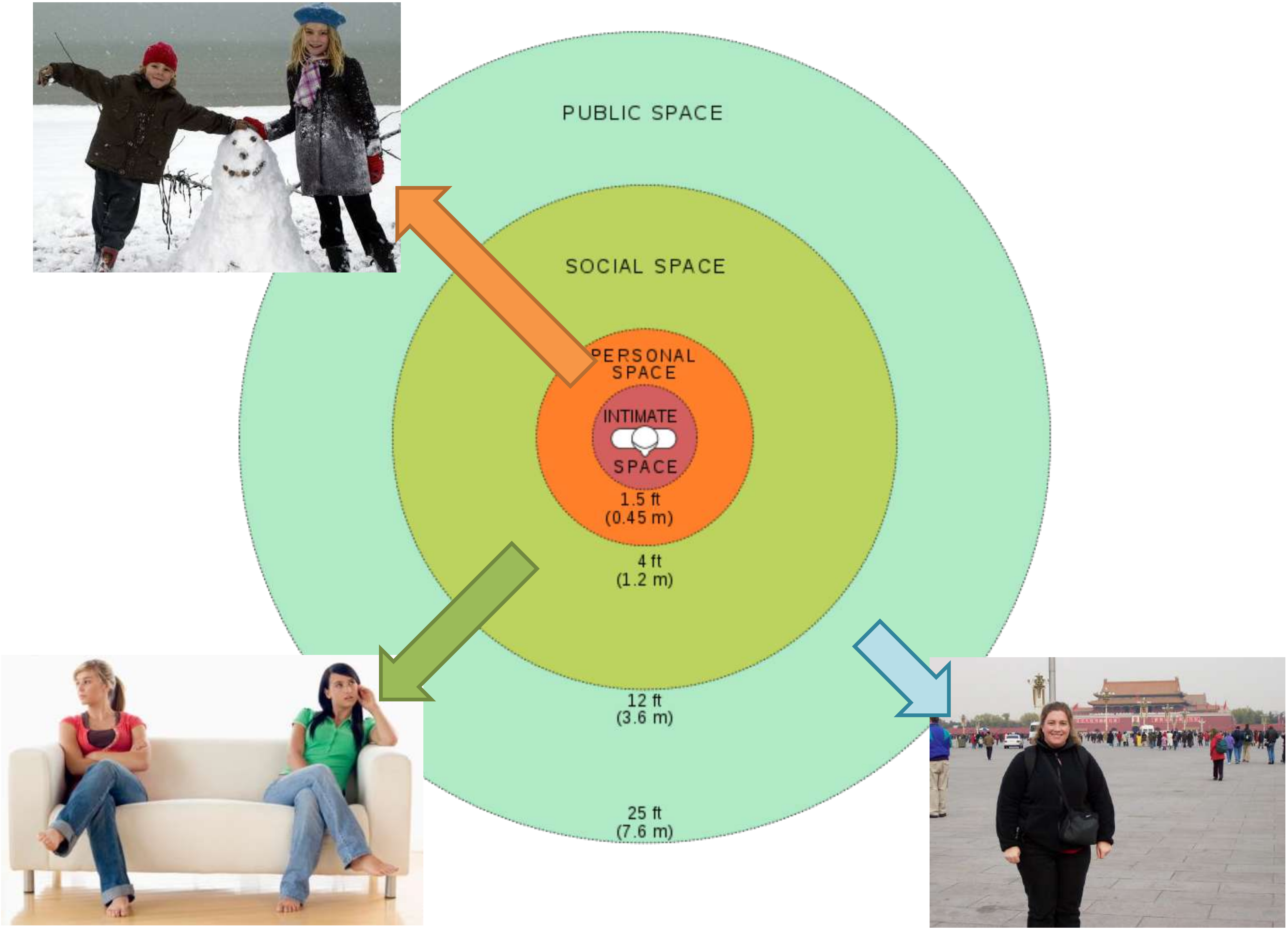
Proxemics: the study of spatial arrangement of people as they interact

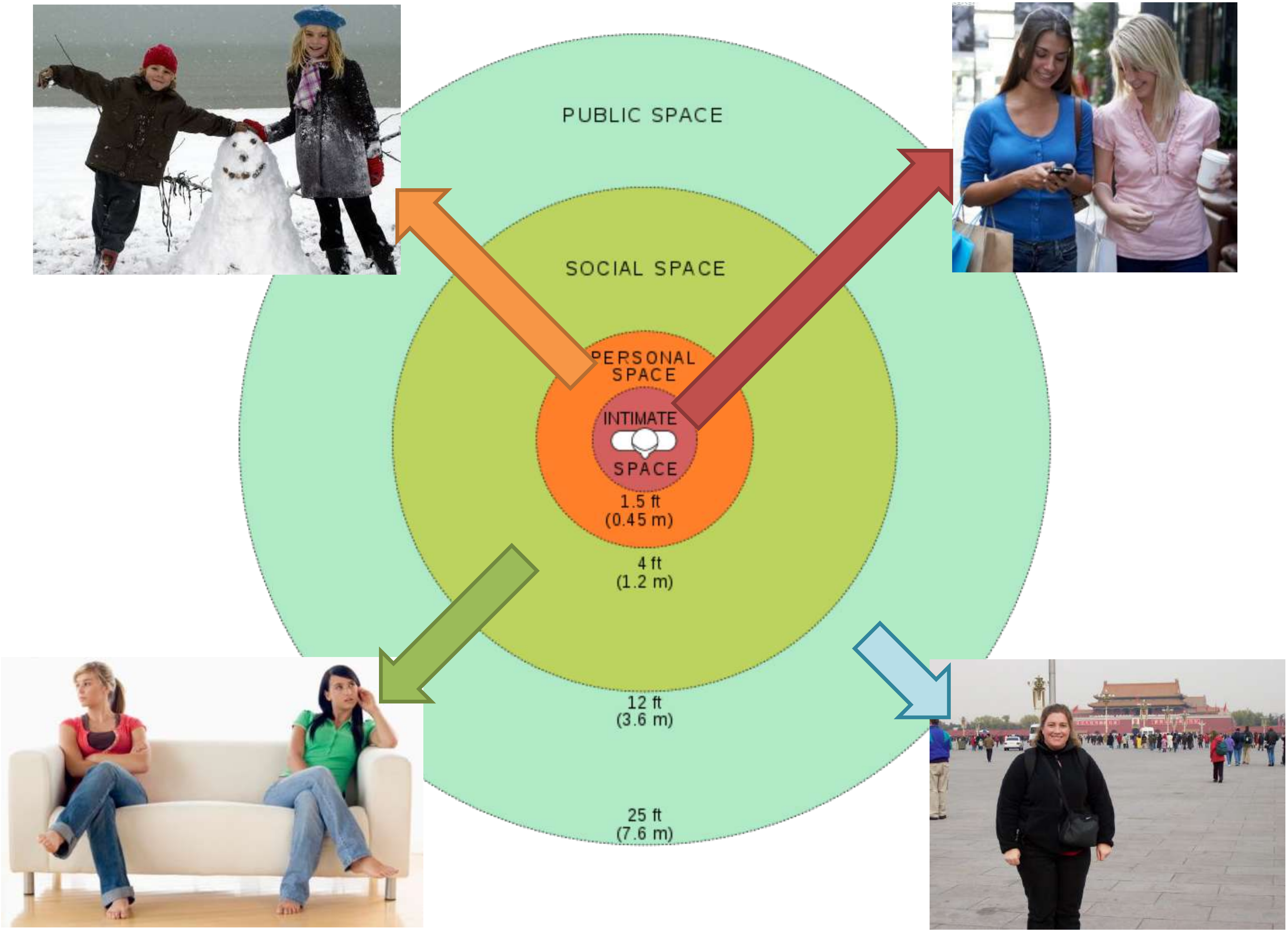
- anthropologist Edward T. Hall in 1963


PUBLIC SPACE
SOCIAL SPACE
PERSONAL SPACE
INTIMATE
SPACE
1.5 ft
(0.45 m)
4 ft
(1.2 m)
12 ft
(3.6 m)
25 ft
(7.6 m)


PUBLIC SPACE
SOCIAL SPACE
PERSONAL SPACE
INTIMATE SPACE
1.5 ft (0.45 m)
4 ft (1.2 m)
12 ft (3.6 m)
25 ft (7.6 m)


PUBLIC SPACE
SOCIAL SPACE
PERSONAL SPACE
INTIMATE SPACE
1.5 ft (0.45 m)
4 ft (1.2 m)
12 ft (3.6 m)
25 ft (7.6 m)

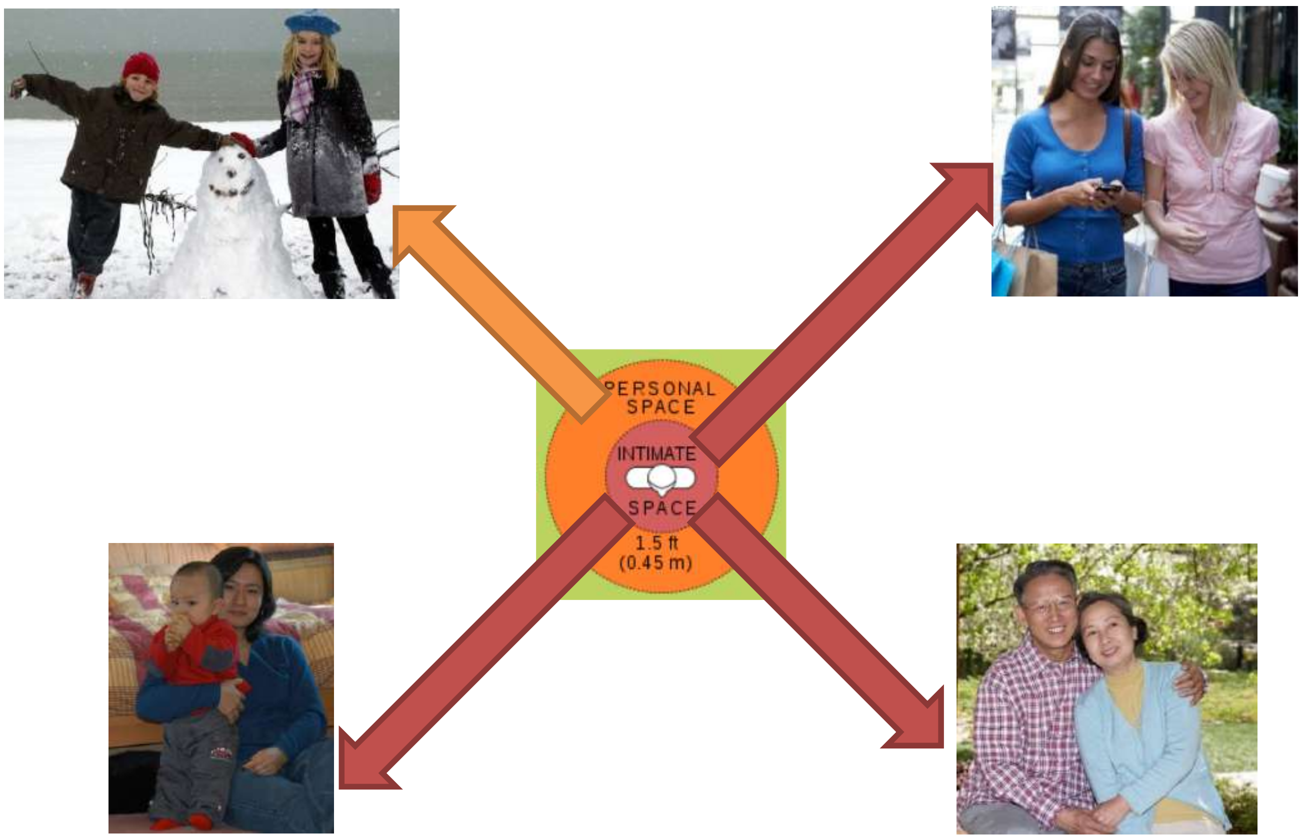

PUBLIC SPACE
SOCIAL SPACE
PERSONAL SPACE
INTIMATE SPACE
1.5 ft (0.45 m)
4 ft (1.2 m)
12 ft (3.6 m)
25 ft (7.6 m)


PUBLIC SPACE
SOCIAL SPACE
PERSONAL SPACE
INTIMATE SPACE
1.5 ft (0.45 m)
4 ft (1.2 m)
12 ft (3.6 m)
25 ft (7.6 m)


Brother and Sister Holding Hands
Friends Walking Side by Side
PERSONAL SPACE
INTIMATE SPACE
1.5 ft (0.45 m)
Mom Holding Baby
Husband Hugging and Holding Wife's Hand

Touch Code

- Hand Touch Hand
- Hand Touch Shoulder
- Shoulder Touch Shoulder
- Arm Touch Torso



Applications

- Personal Photo Search:
  - Find a specific interesting photo
- Analysis of TV shows and movies
- Kinect
- Web Search
- Auto-Movie/Auto-Slideshow
- Locate interesting scenes

Proxemics DataSet

- 200 training images
- 150 testing images

- Collected From
  - Simon, Bing, Google, Gettyimage, Flickr

- No video data
- No Kinect 3d depth data

Proxemics DataSet

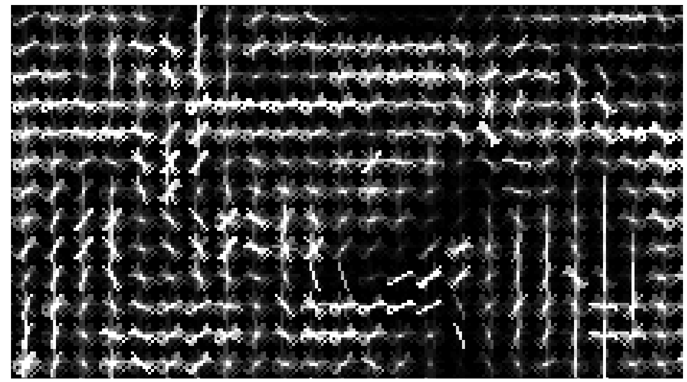
A Naïve Approach

Input Image

A Naïve Approach

Input Image

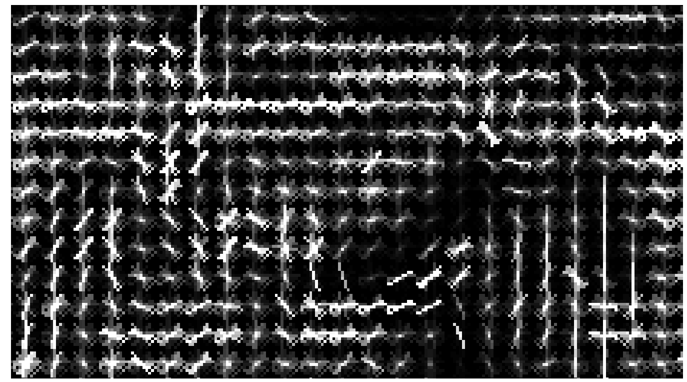
Image Feature

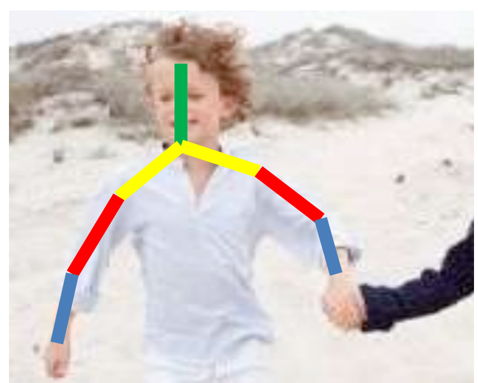


A Naïve Approach

Input Image

Image Feature



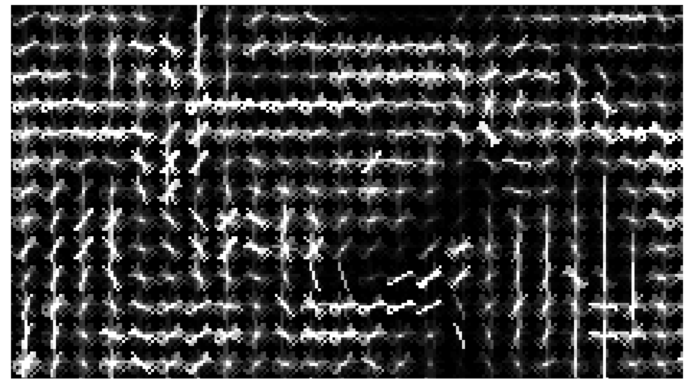


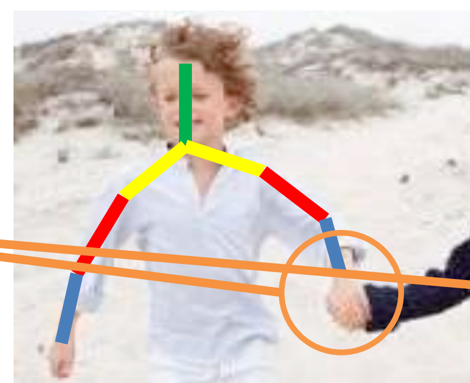
Pose Estimation i.e. Find skeletons

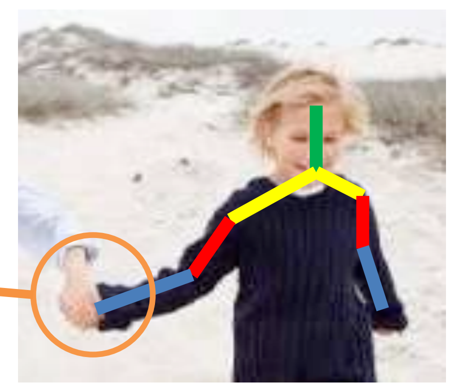
A Naïve Approach

Input Image

Image Feature





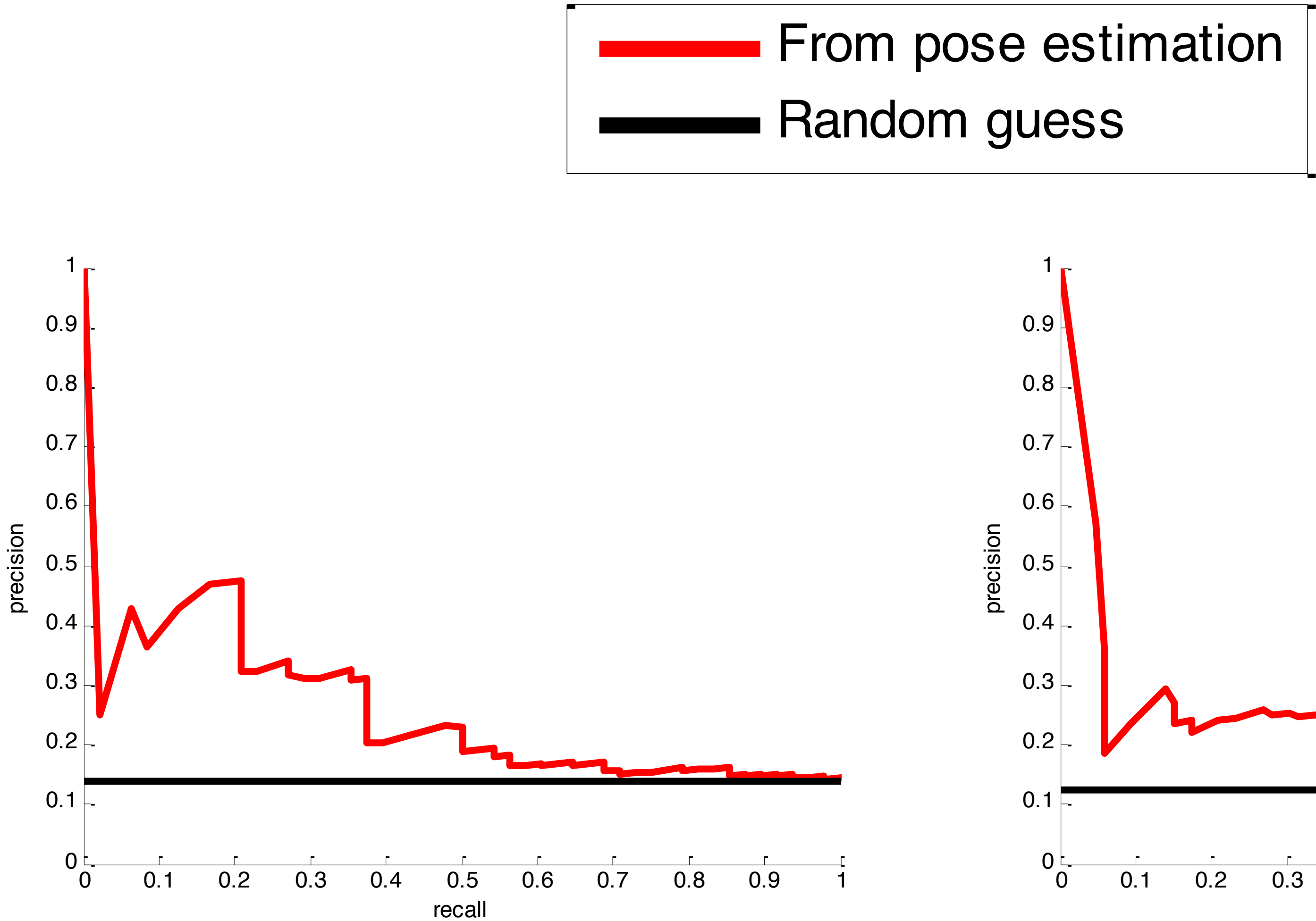


Hand touch Hand

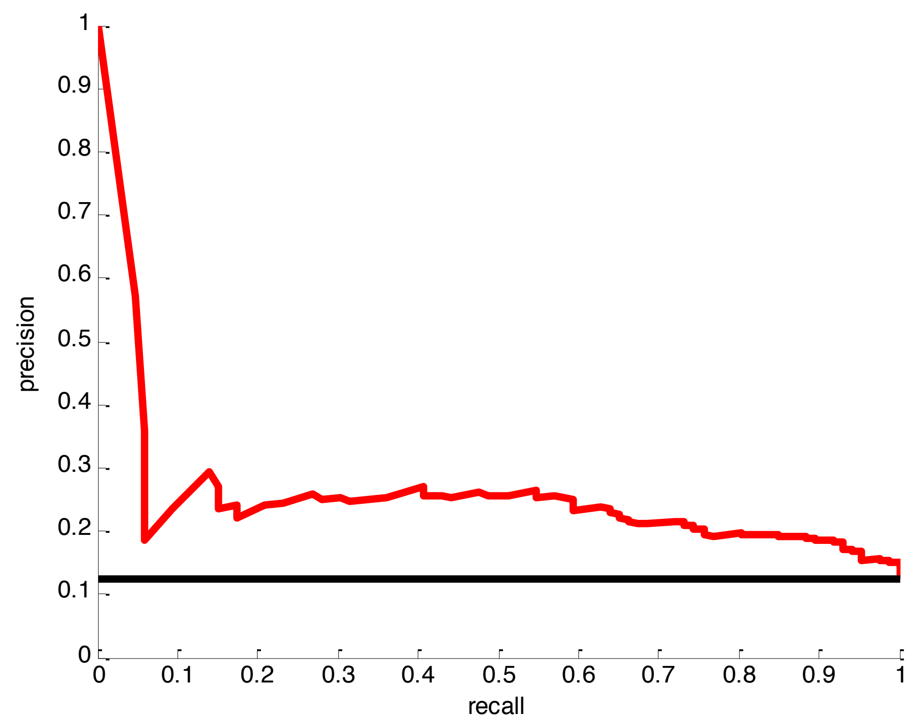
Interaction Recognition

Pose Estimation i.e. Find skeletons

Naïve Approach Results



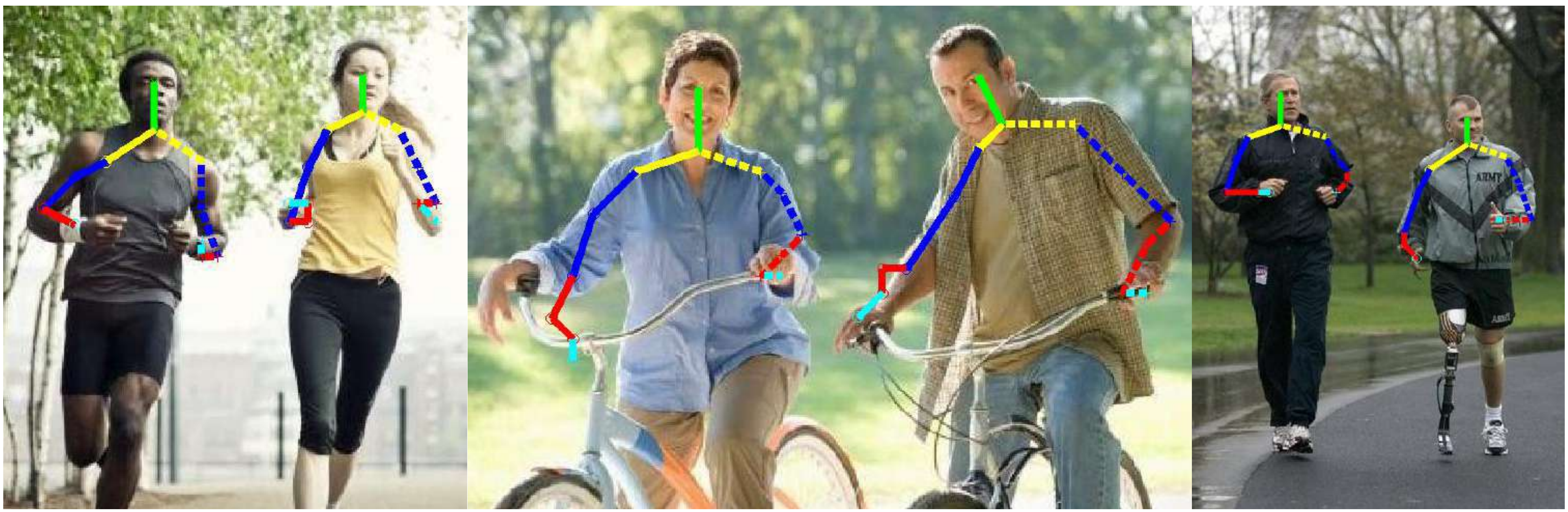
Hand touch Hand



Hand touch Shoulder

Human Pose Estimation

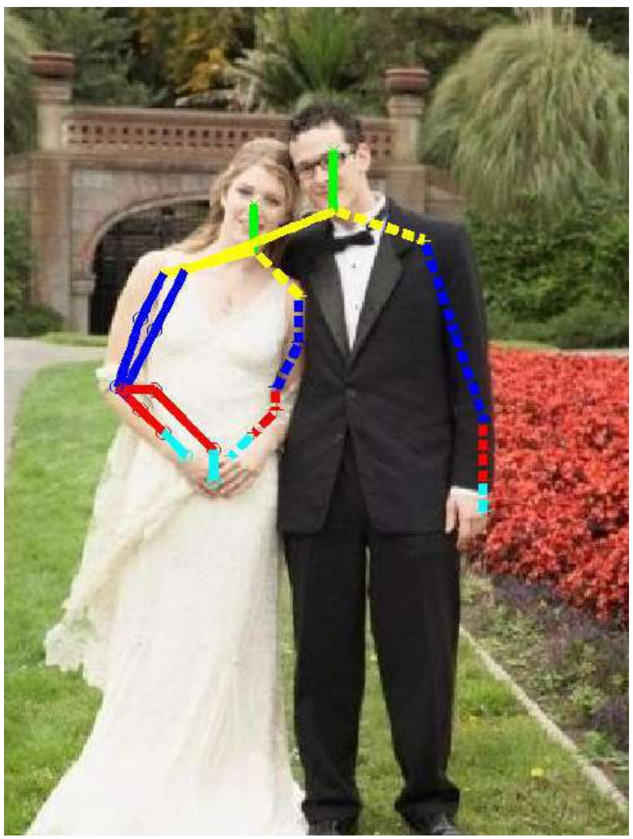
- Not bad when no real interaction between people

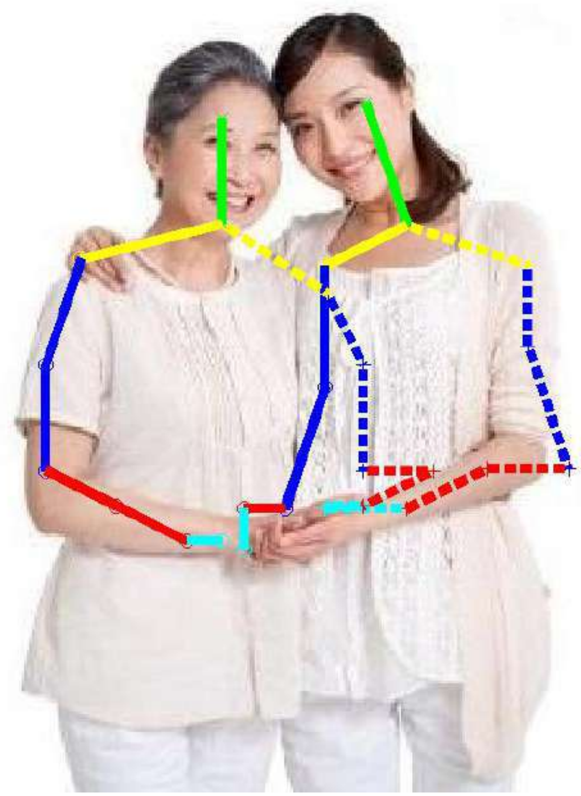


- Y. Yang & D. Ramanan, CVPR 2011

Interactions Hurt Pose Estimation

Occlusion + Ambiguous Parts

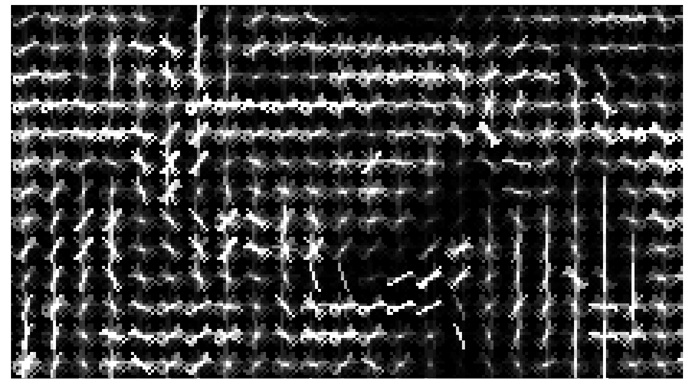




Our Approach
Direct Proxemics Recognition

Input Image

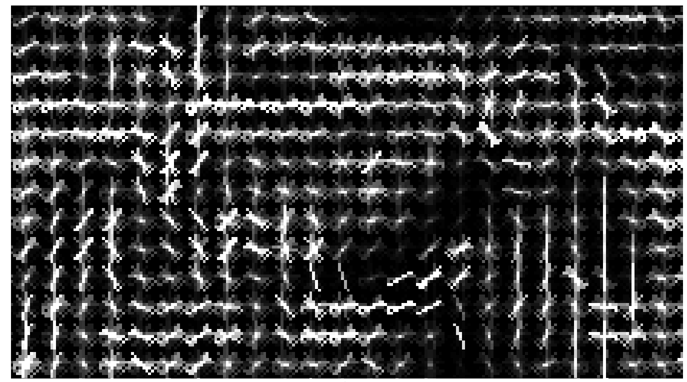
Image Feature

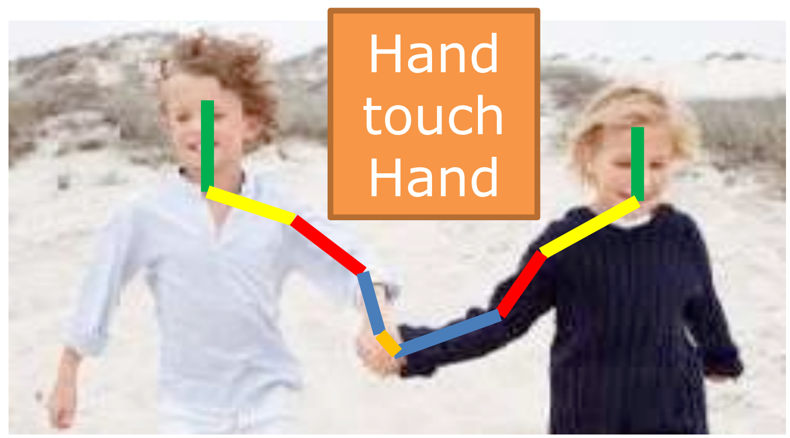


Our Approach
Direct Proxemics Recognition

Input Image

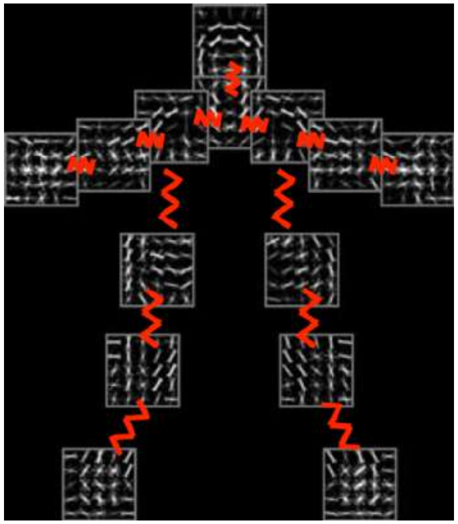
Image Feature





Interaction Recognition

Pictorial Structure Model

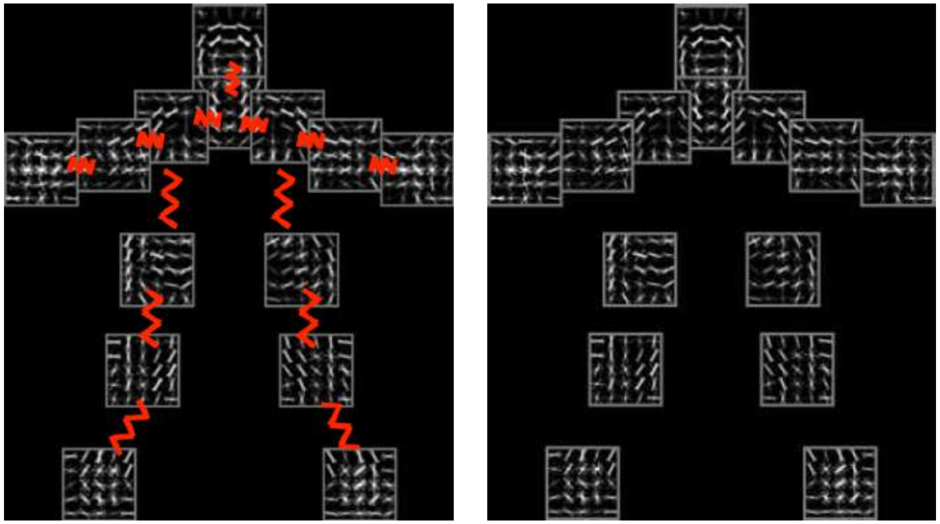


$S(I, L)$

- I: Image
- l_i: Location of part i

- P. Felzenszwalb etc., PAMI 2009

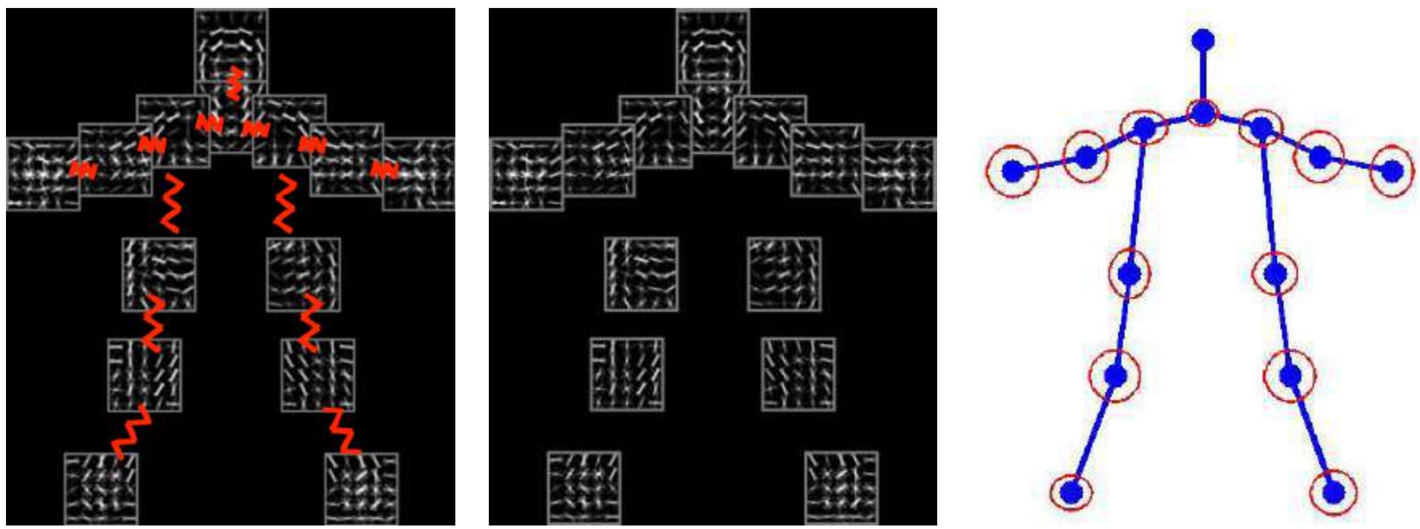
Pictorial Structure Model



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

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

Pictorial Structure Model

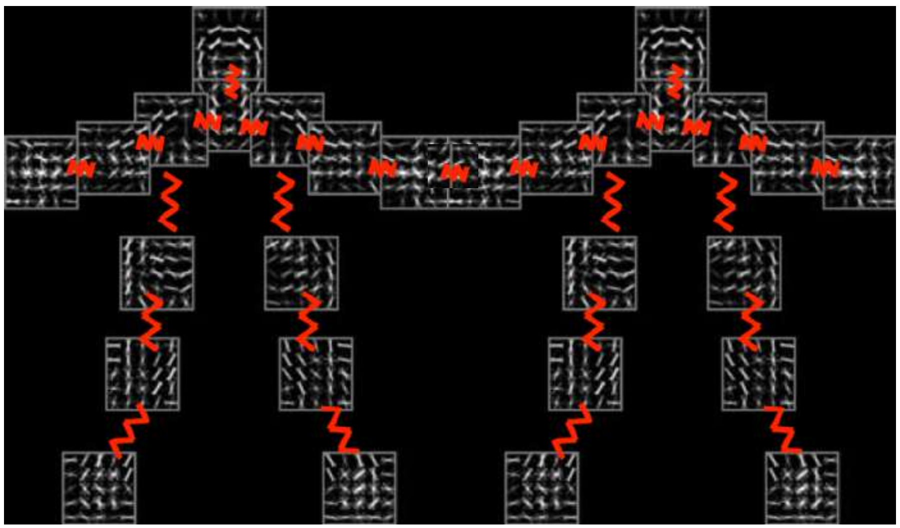


$$S(I, L) = \sum_{i \in V} \alpha_i \cdot \phi(I, l_i) + \sum_{ij \in E} \beta_{ij} \cdot \psi(l_i, l_j)$$

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

"Two Head Monster" Model

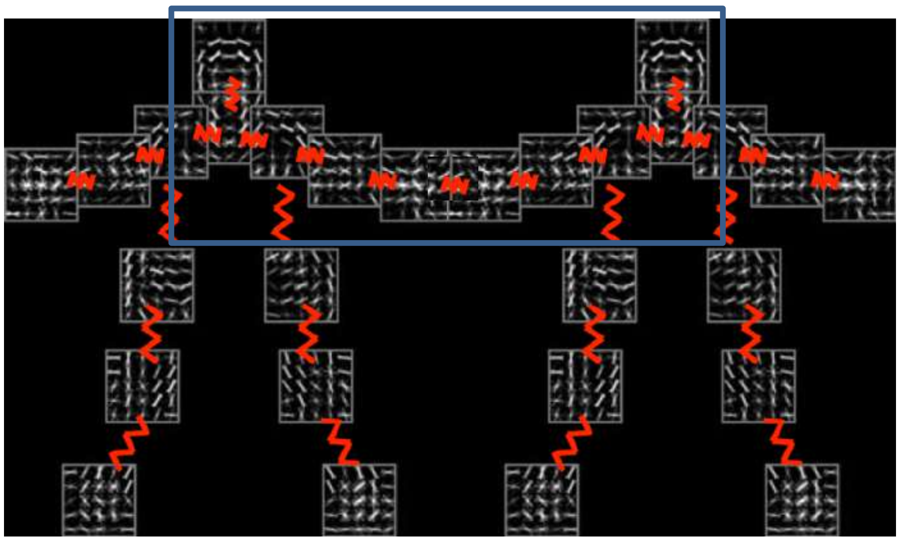
i.e. "Hand-Touching-Hand"

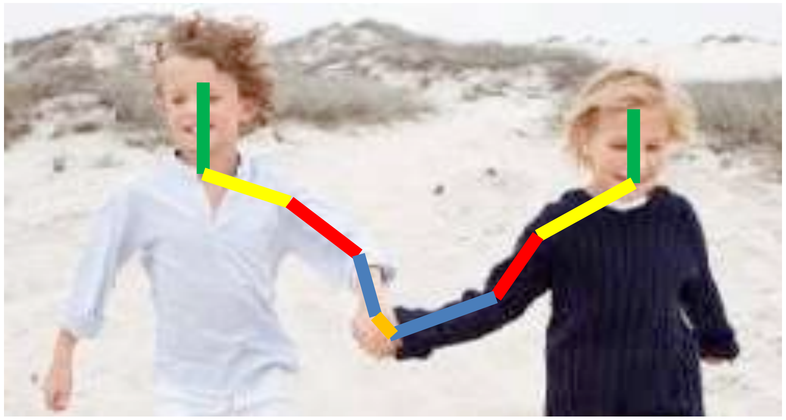


$$S(I, L) = \sum_{i \in V} \alpha_i \cdot \phi(I, l_i) + \sum_{ij \in E} \beta_{ij} \cdot \psi(l_i, l_j)$$

"Two Head Monster" Model

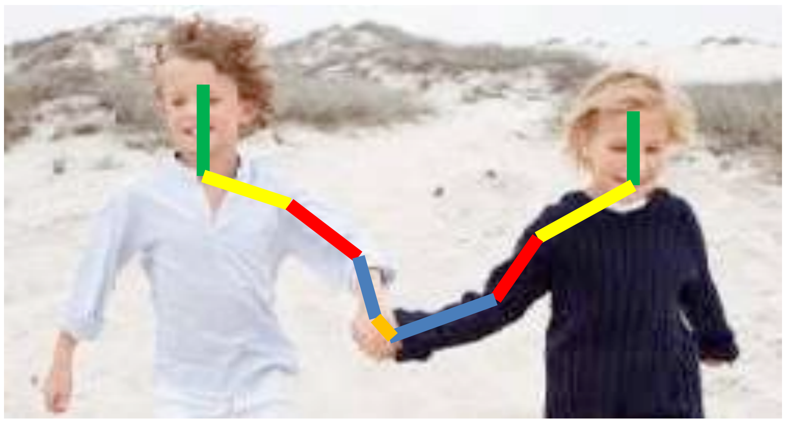
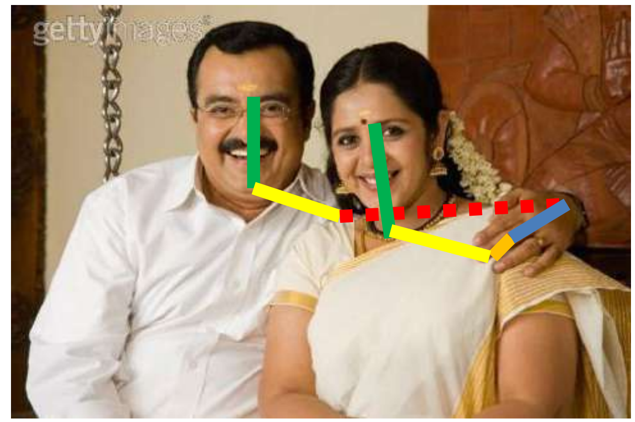
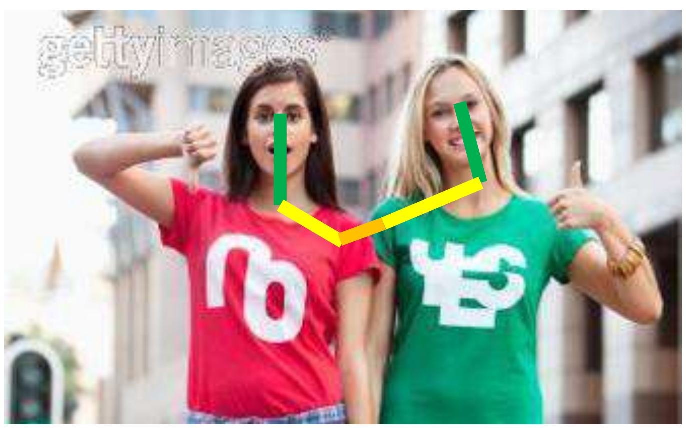
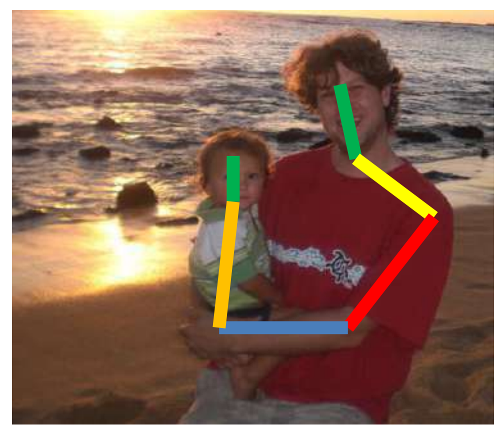
i.e. "Hand-Touching-Hand"





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

The models

Hand touch Hand	Hand touch Shoulder
	
Shoulder touch Shoulder	Arm touch Torso
	

Match Model to Image

- Inference:

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

  - Efficient algorithm:
  - Dynamic programming

- Learning:
  - Structural SVM Solver

Refinements + Extensions

- Sub-categories

  Because of symmetry, 4 models for hand-hand etc

R. Hand
L. Hand

L. Hand
R. Hand

L. Hand
L. Hand

R. Hand
R. Hand

Refinements + Extensions

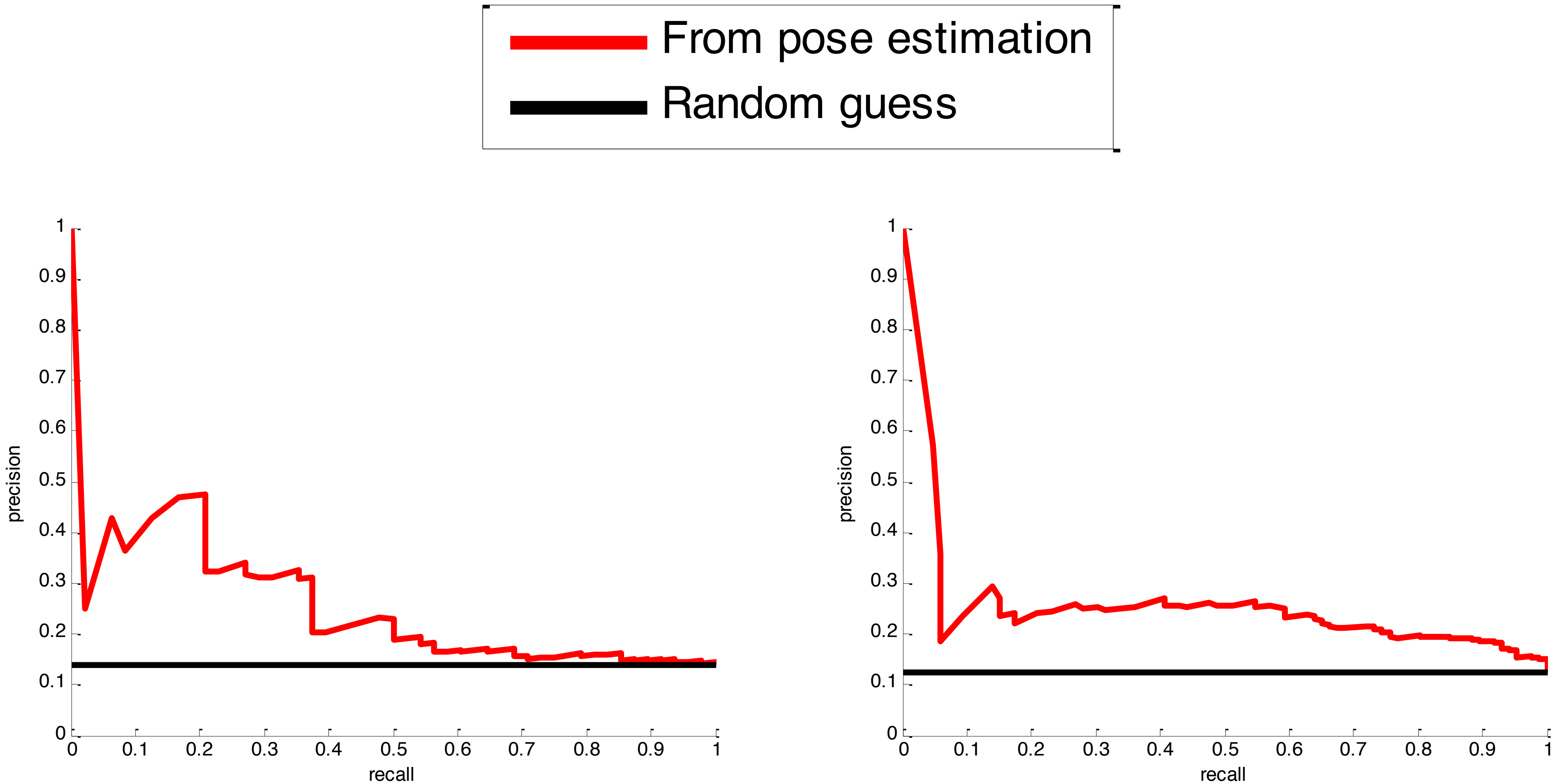
- Sub-categories

  Because of symmetry, 4 models for hand-hand etc

- Co-occurrence of proxemics:

  Reduce redundancy, map Multi-Label -> Multi-Class

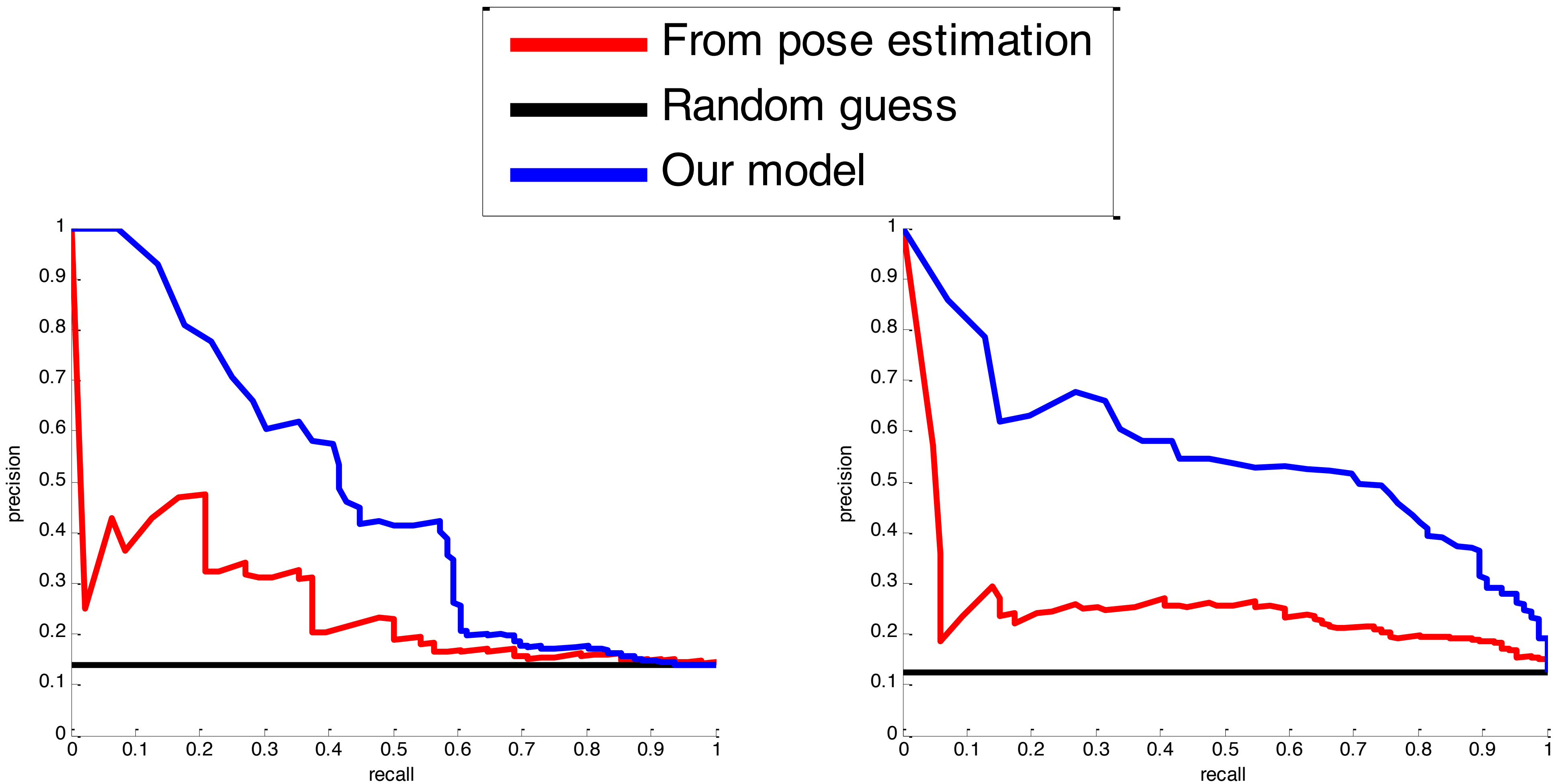
Naïve Approach Results



Hand touch Hand

Hand touch Shoulder

Direct Approach Results



Hand touch Hand

Hand touch Shoulder

Quantitative Results

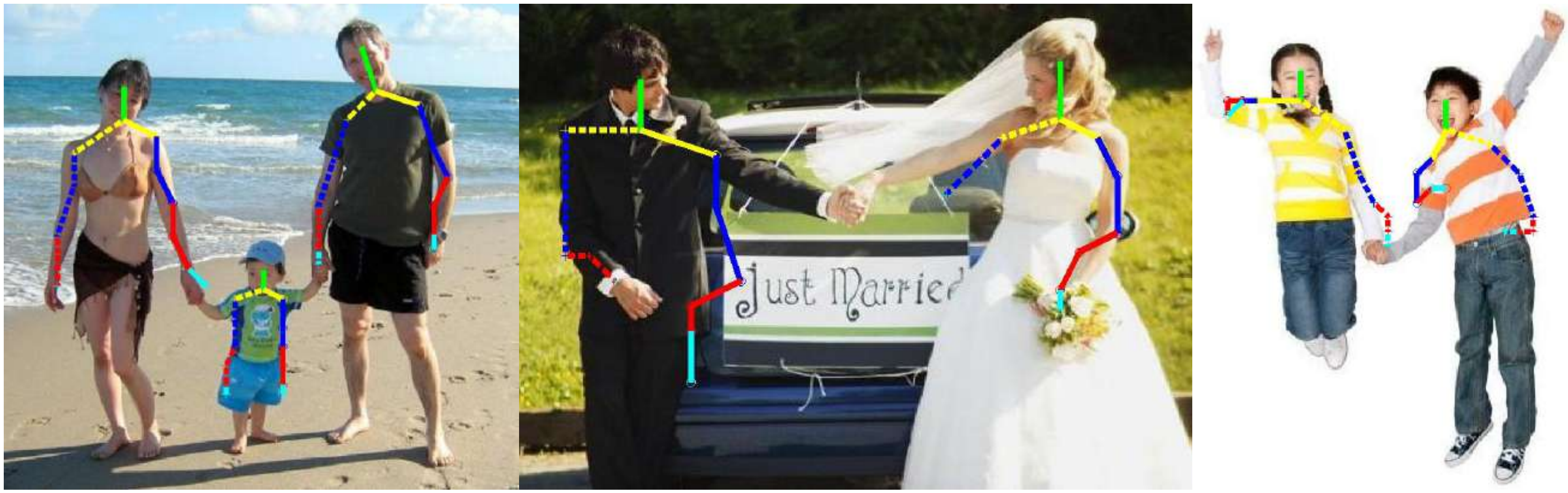
Proxemic Recognition Average Precision

Proxemics	Hand Hand	Hand Shoulder	Shoulder Shoulder	Hand Torso	Average
Random guess	14.0	12.6	24.6	9.9	15.3
From pose estimation [1]	26.5	25.6	71.7	18.7	35.6
Our direct model	**46.9**	**55.2**	**72.0**	**87.3**	**65.4**

[1] Y. Yang & D. Ramanan, CVPR 2011

Improves Pose Estimation

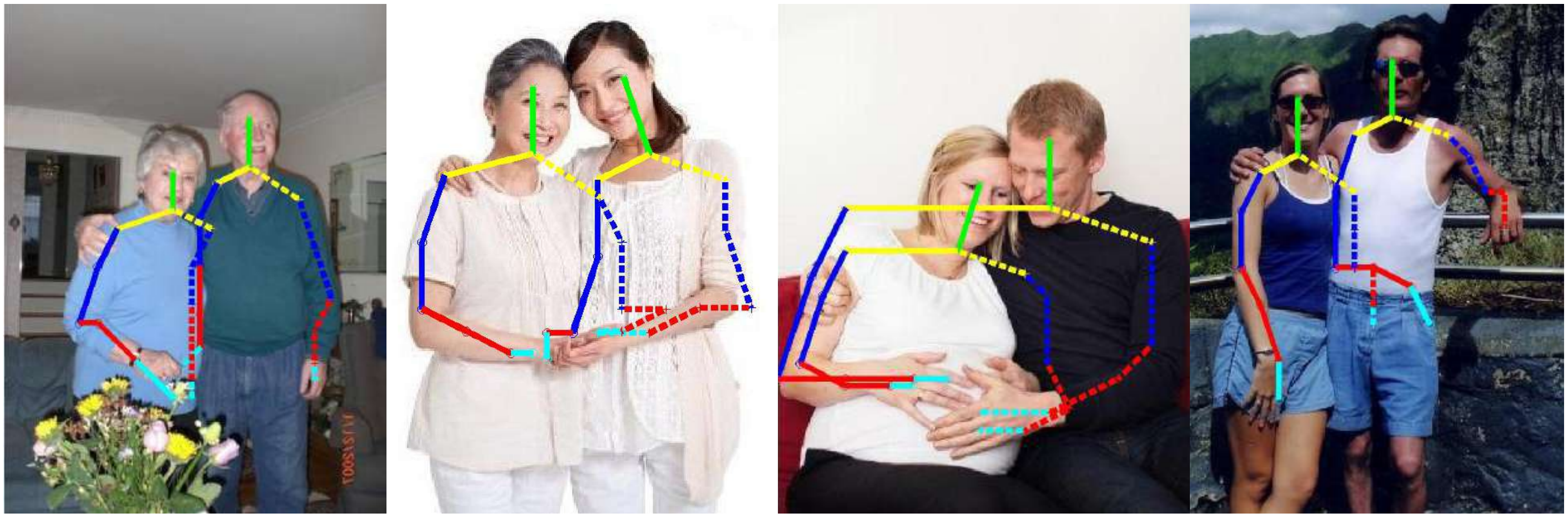
Y & D
CVPR
2011



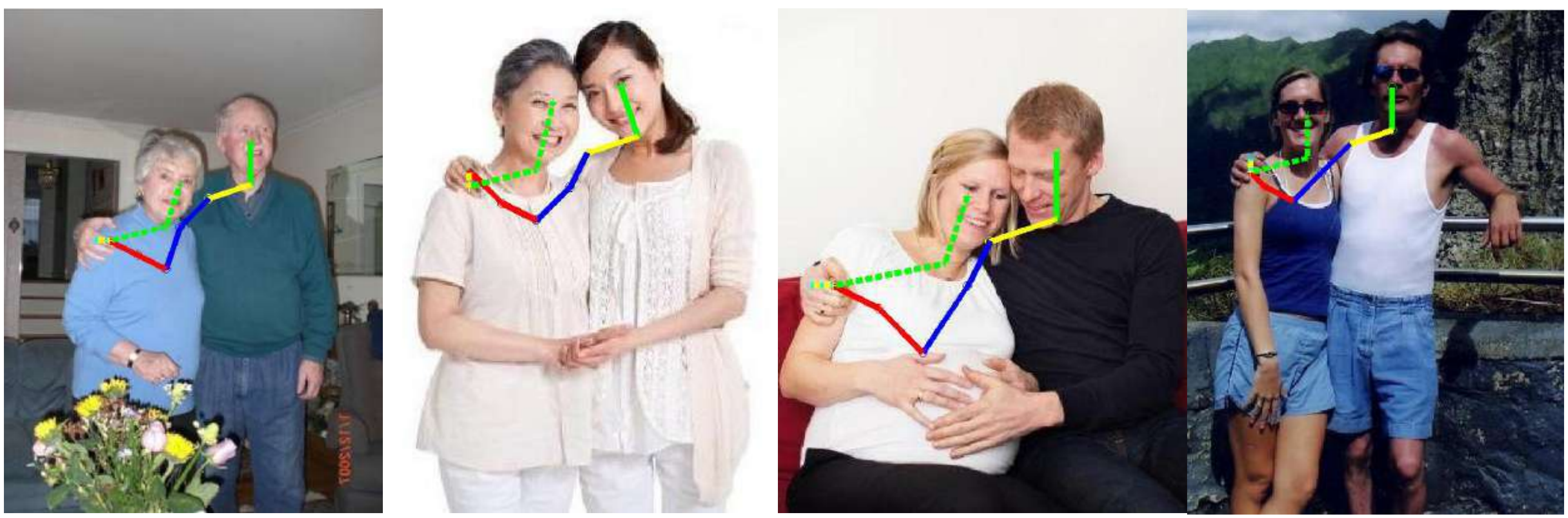
Our
Model

Improves Pose Estimation

Y & D
CVPR
2011



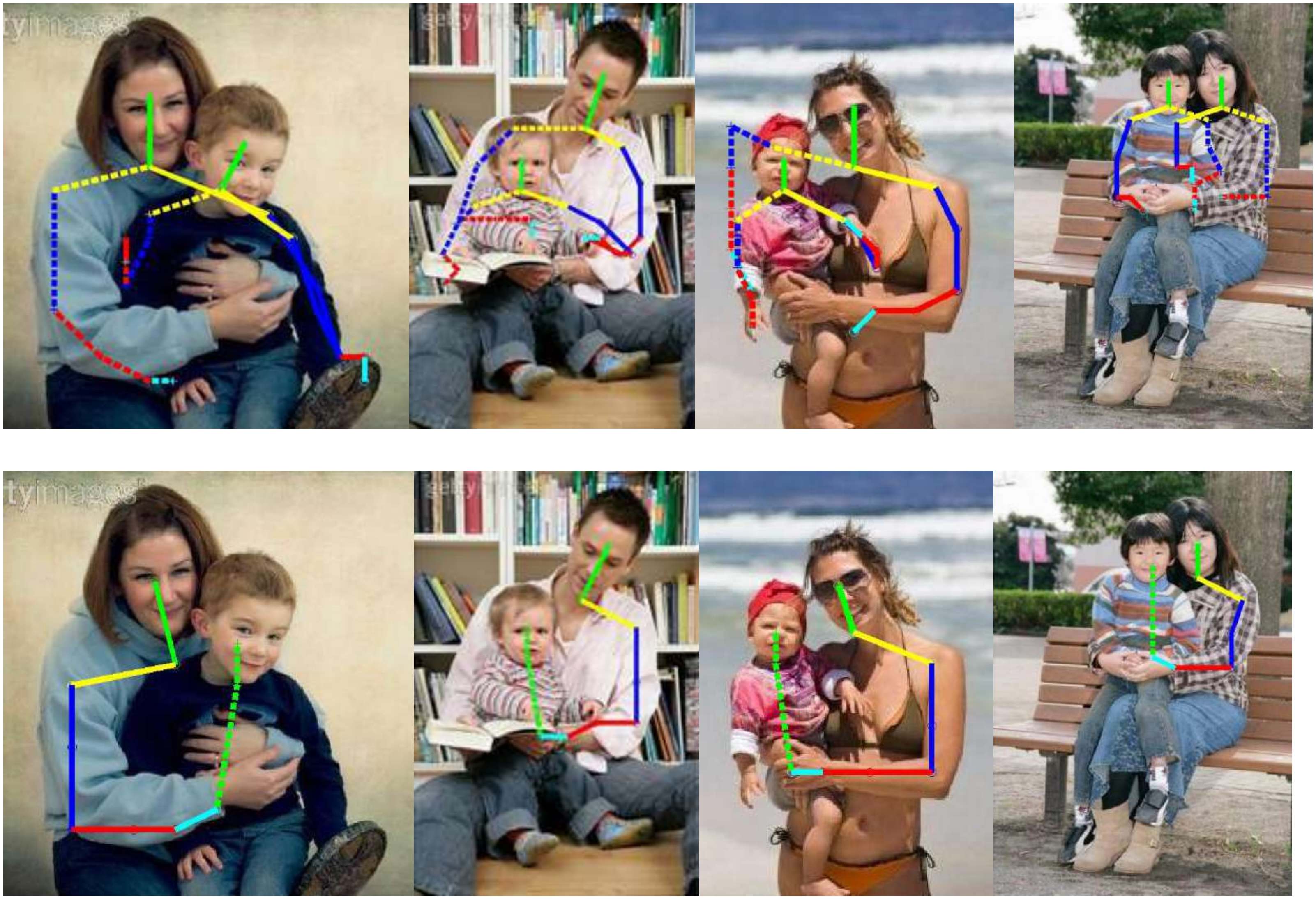
Our
Model



Improves Pose Estimation

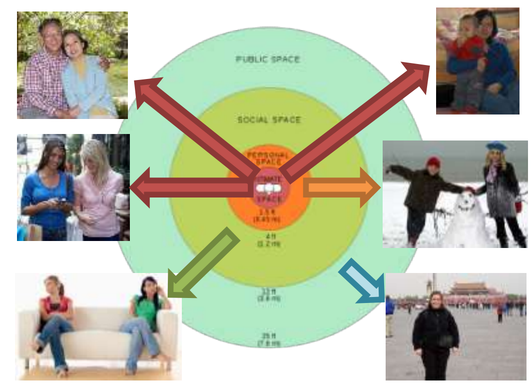
Y & D CVPR 2011

Our Model



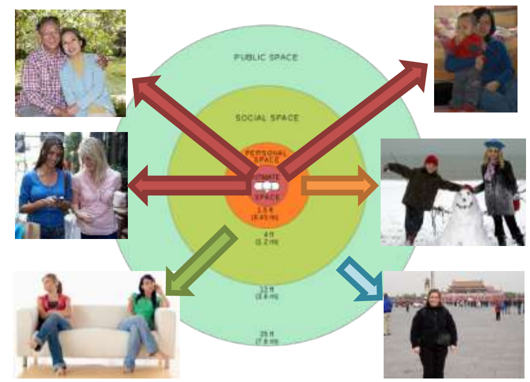
Conclusion

- Proxemics and touch codes for human interaction

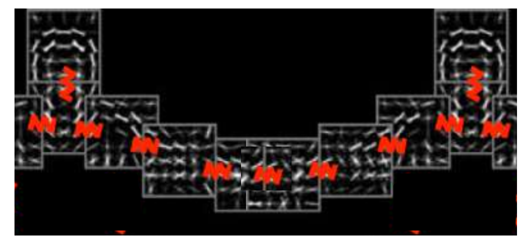


Conclusion

- Proxemics and touch codes for human interaction

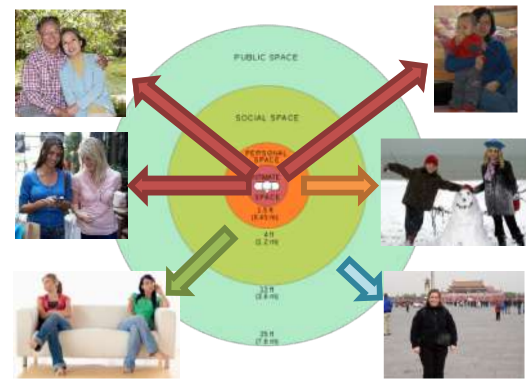


- Directly recognizing proxemics significantly outperforms

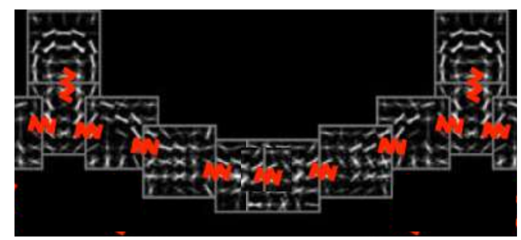


Conclusion

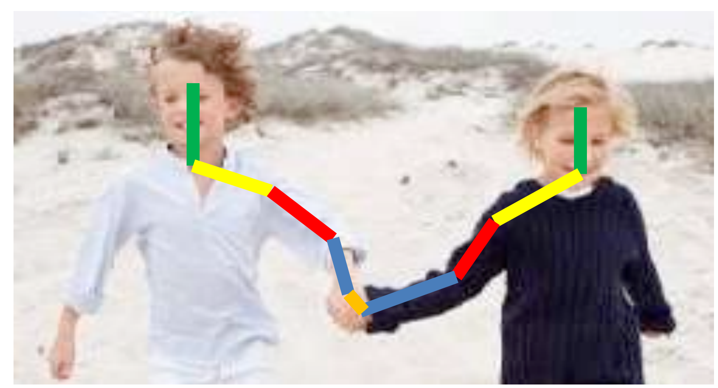
- Proxemics and touch codes for human interaction



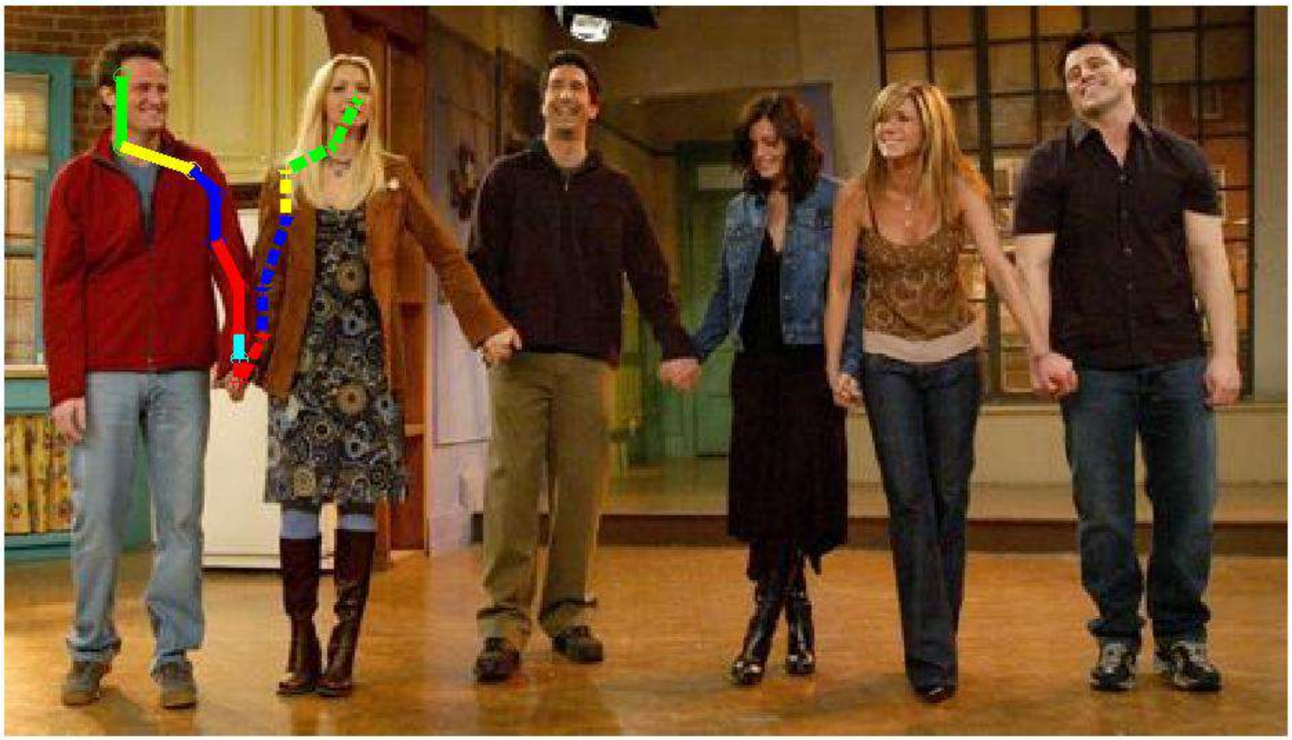
- Directly recognizing proxemics significantly outperforms



- Recognizing proxemics helps pose estimation

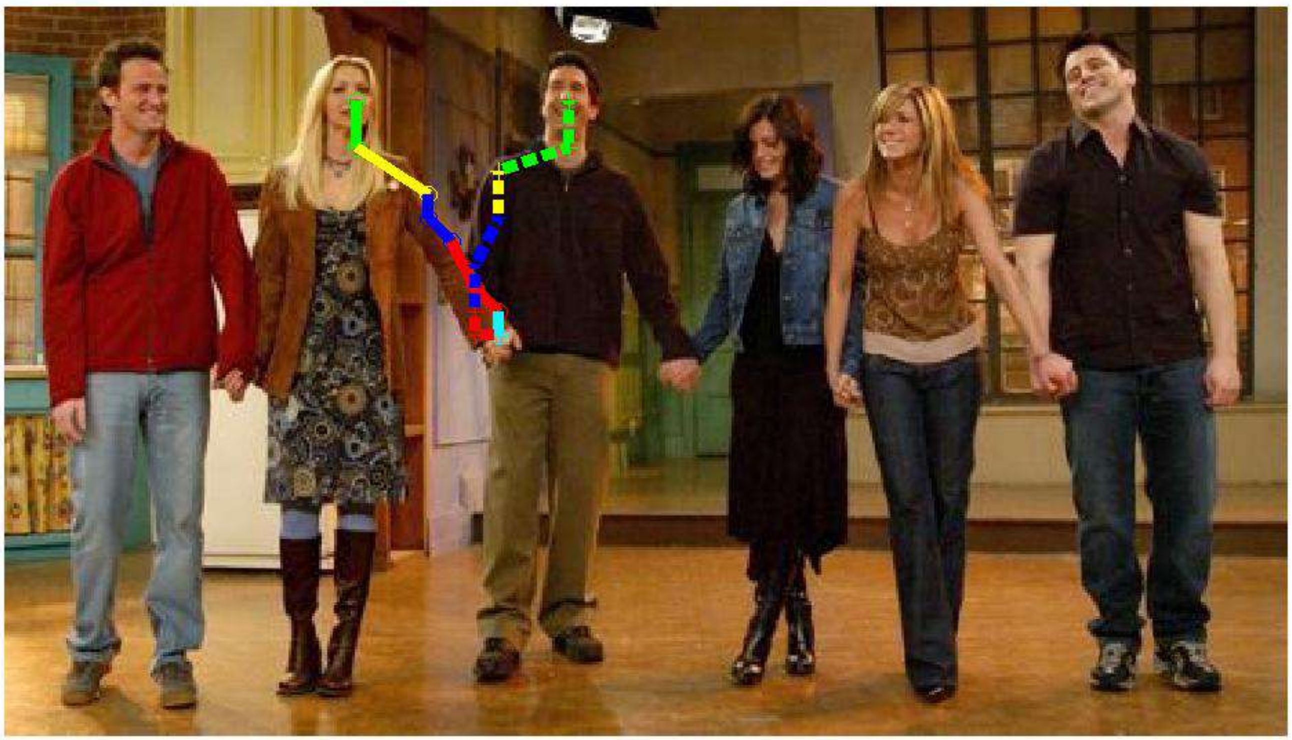


Acknowledgements



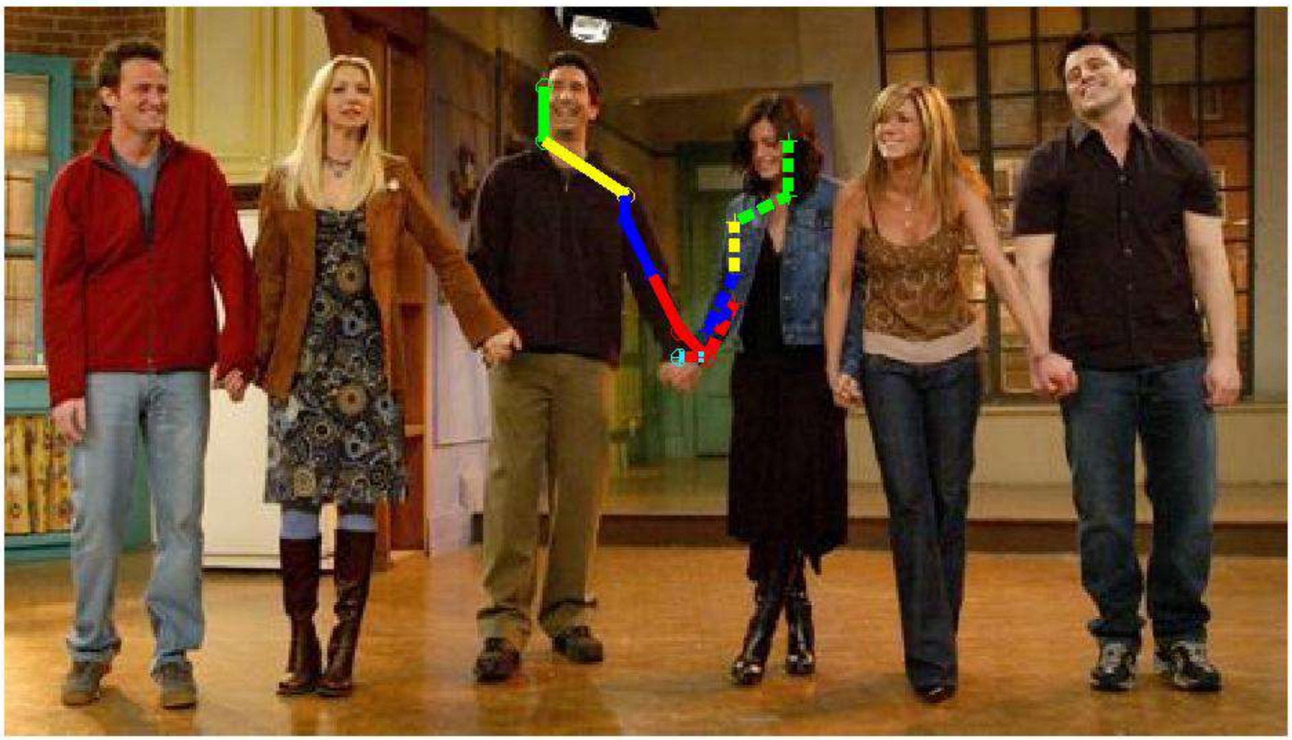
Thank Simon and MSR for internship

Acknowledgements



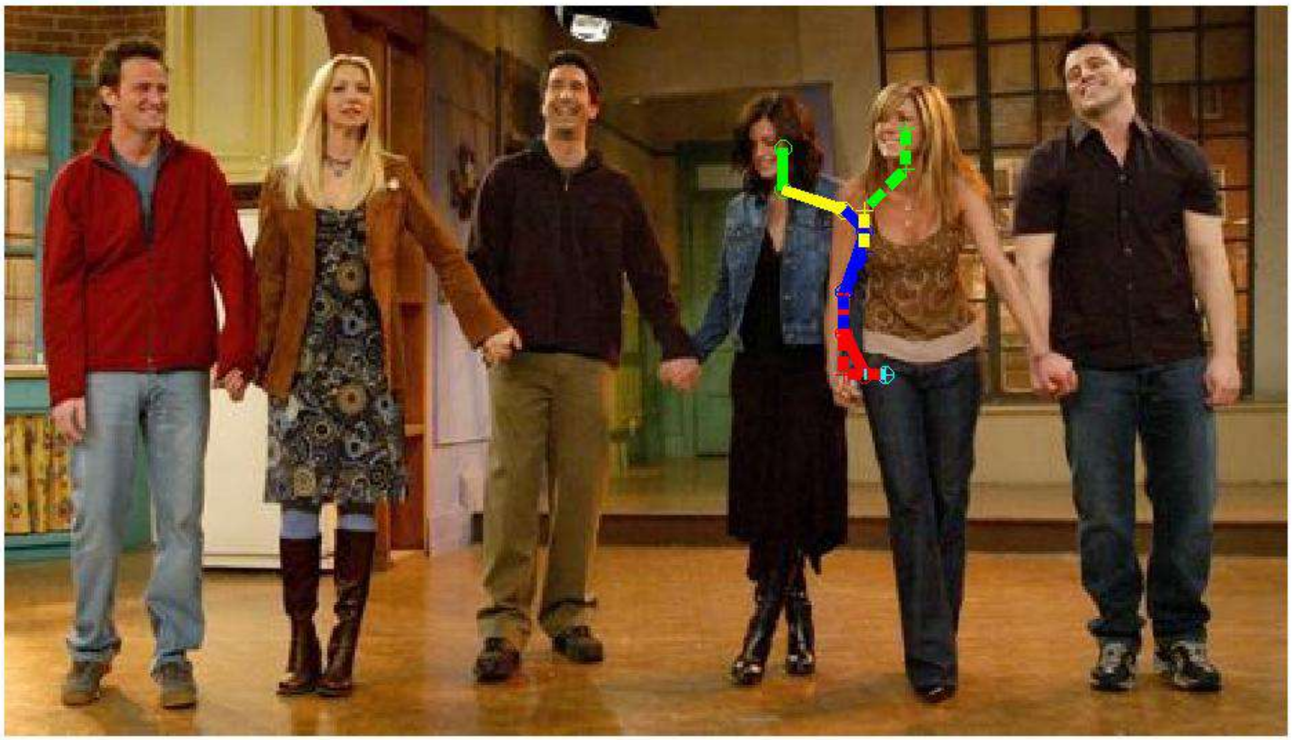
Thank Anitha for a lot of suggestions

Acknowledgements



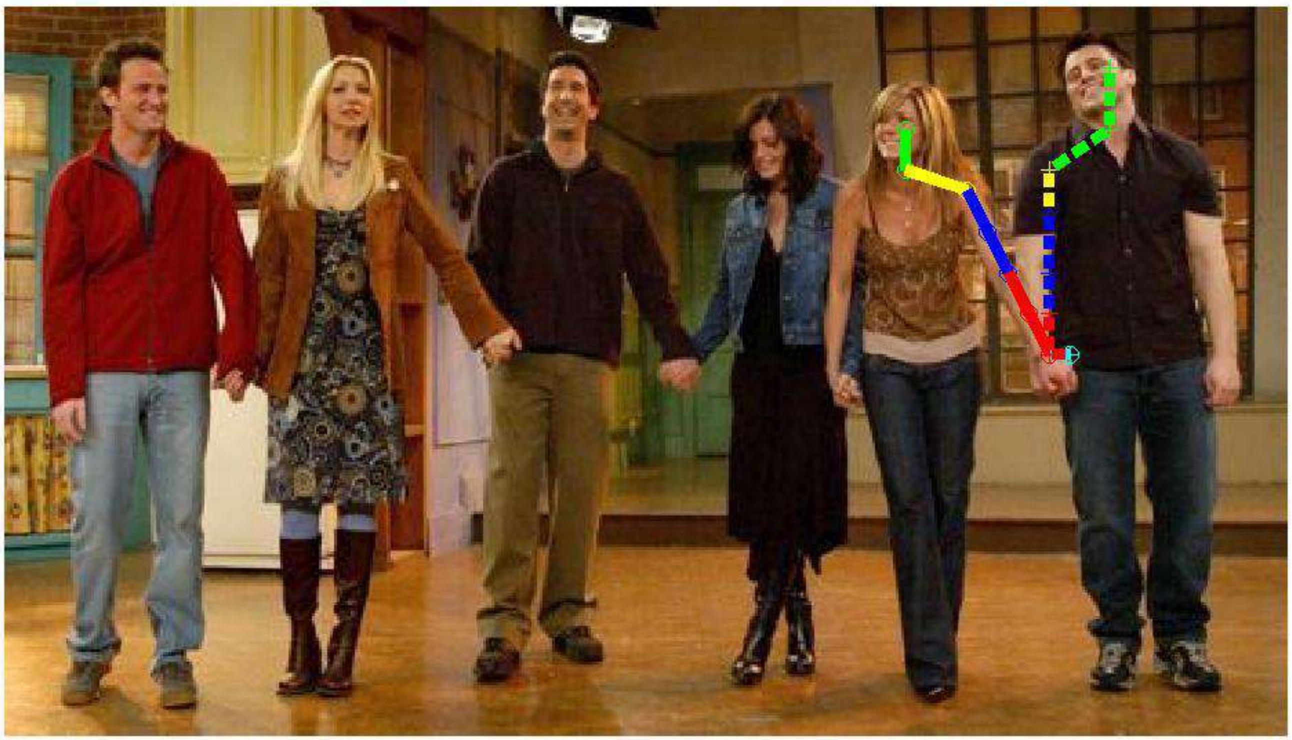
Thank Anarb for gettyimages

Acknowledgements



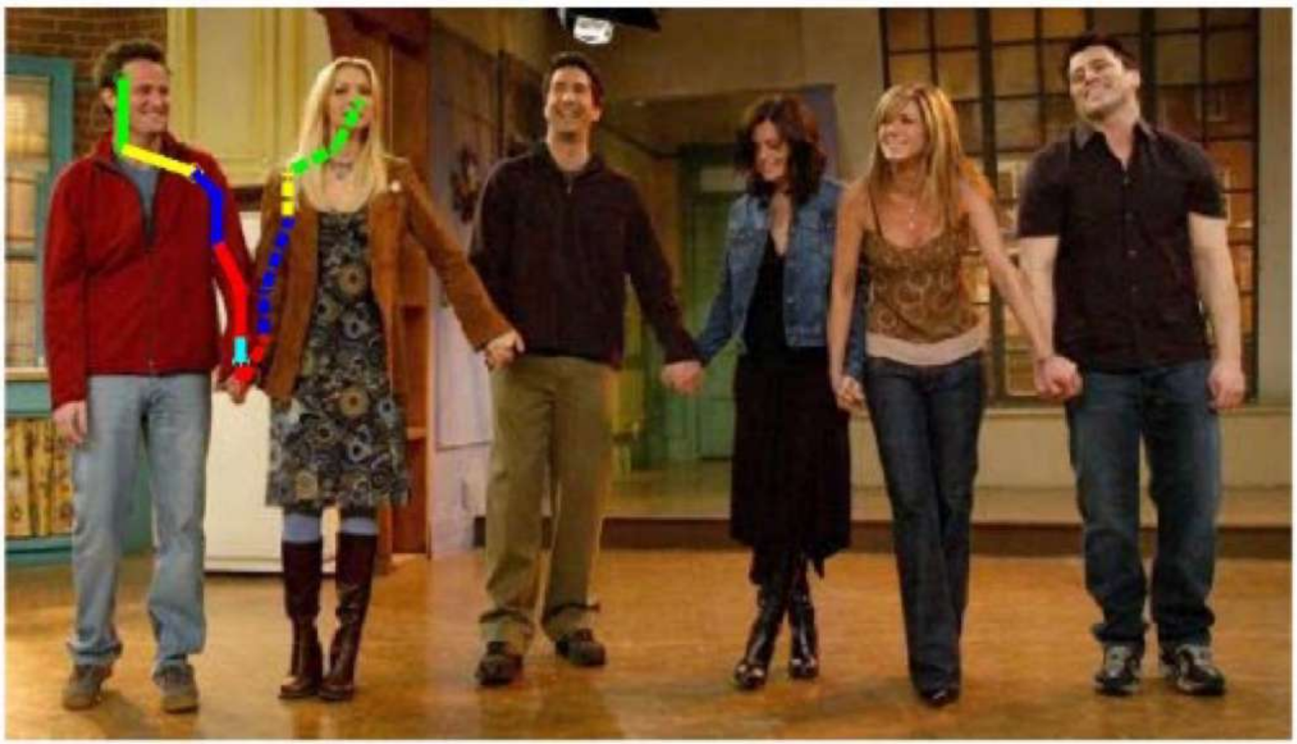
Thank Eletee for her beautiful smiling

Acknowledgements



Thank everybody for not falling asleep

Thank you



Articulated Pose Estimation

Percentage of Correctly Localized Parts on Proxemic Dataset

DataSet	Shoulders	Elbows	Wrists	Hands
Single Person	93.0	61.7	38.7	34.6
Interacting People	86.6	46.2	22.0	17.4
Difference	6.4	**15.5**	**16.7**	**17.2**

- Yi Yang & Deva Ramanan, CVPR 2011

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 $S(I,z) = w \cdot \phi(I,z)$