

# 3D Pictorial Structures for Multiple Human Pose Estimation

Vasileios Belagiannis<sup>1</sup>, Sikandar Amin<sup>2,3</sup>, Mykhaylo Andriluka<sup>3,4</sup>,  
Bernt Schiele<sup>3</sup>, Nassir Navab<sup>1</sup>, Slobodan Ilic<sup>1</sup>

<sup>1</sup>Computer Aided Medical Procedures (CAMP), Technische Universität München, Germany

<sup>2</sup>Intelligent Autonomous Systems, Technische Universität München, Germany

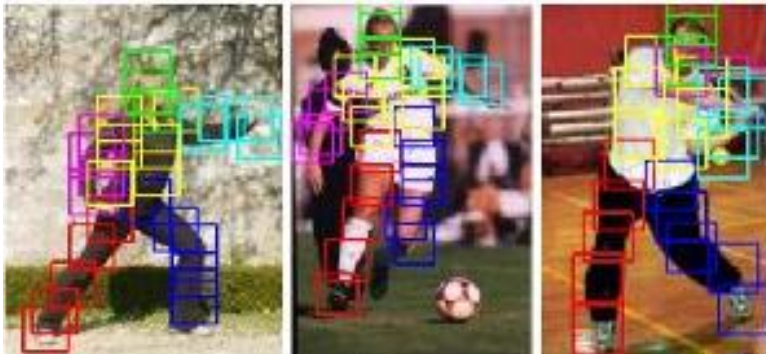
<sup>3</sup>Computer Vision and Multimodal Computing, Max Planck Institute for Informatics Saarbrücken, Germany

<sup>4</sup>Stanford University, USA

**Paper ID: O-2C-6**

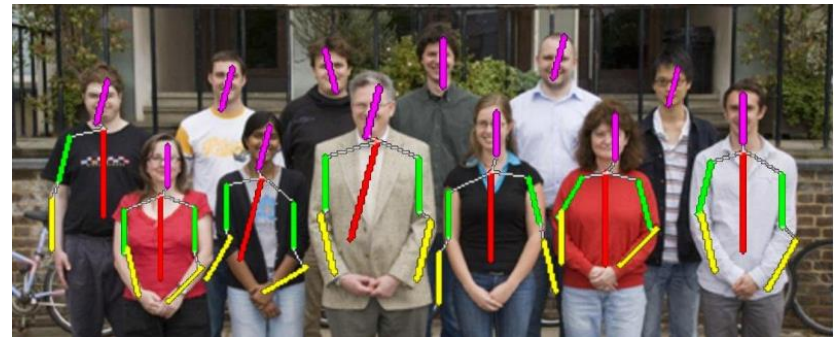
# 2D Human Pose Estimation

## Single-View – Single-Human



Y. Yang and D. Ramanan. Articulated pose estimation with flexible mixtures-of-parts. In CVPR, 2011.

## Single-View – Multiple-Human



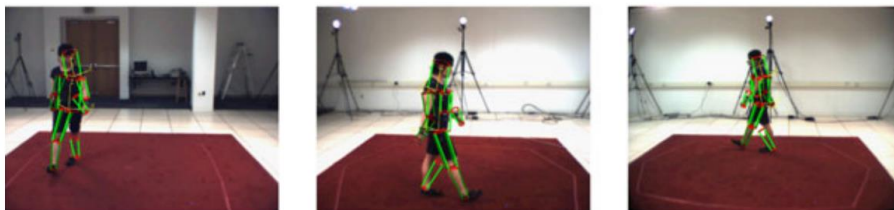
Marcin Eichner and Vittorio Ferrari. We are family: Joint pose estimation of multiple persons. In ECCV, 2010.

# 3D Human Pose Estimation

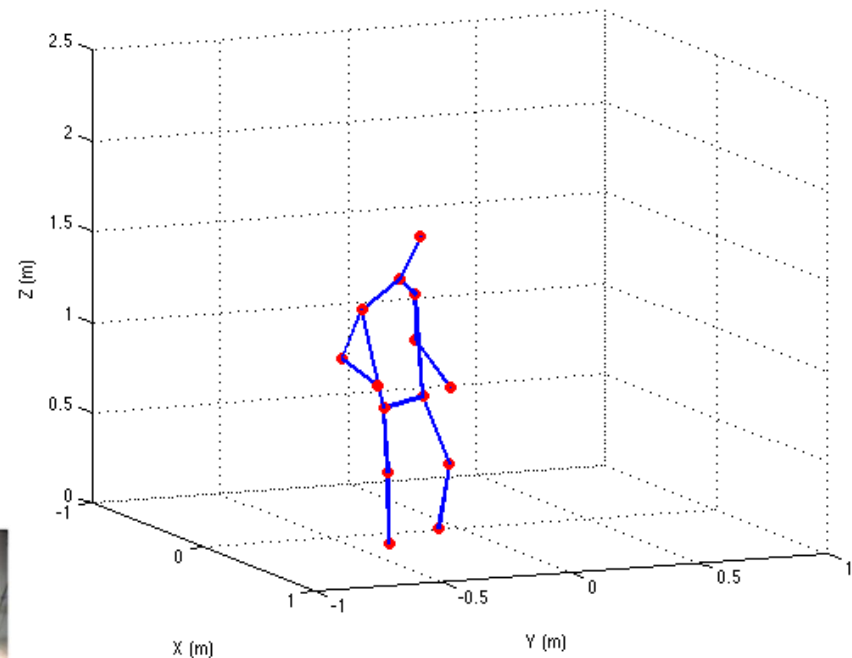
## Multi-View – Single-Human



M. Burenius, et al. 3D pictorial structures for multiple view articulated pose estimation. In CVPR, 2013.



Sigal, Leonid, et al. "Loose-limbed people: Estimating 3d human pose and motion using non-parametric belief propagation.", In IJCV 2012.



Introduction

3DPS Model

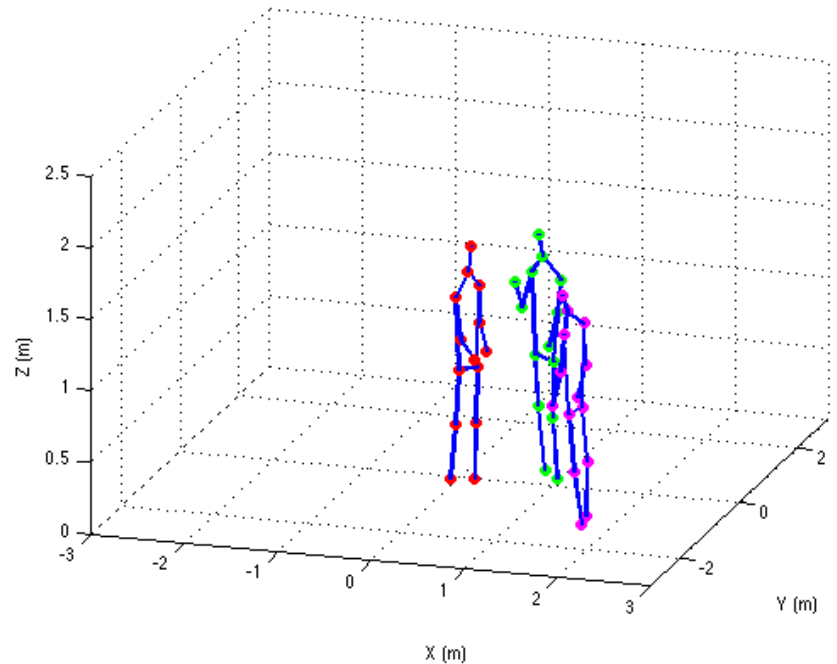
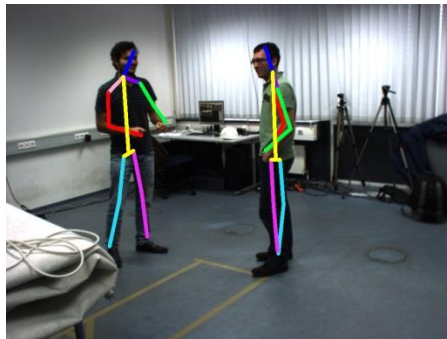
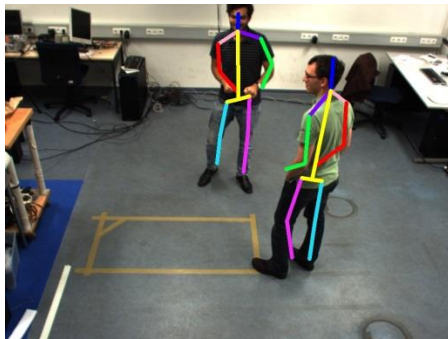
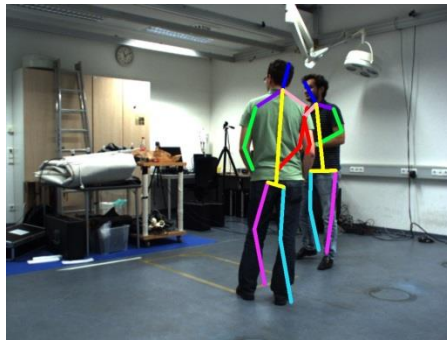
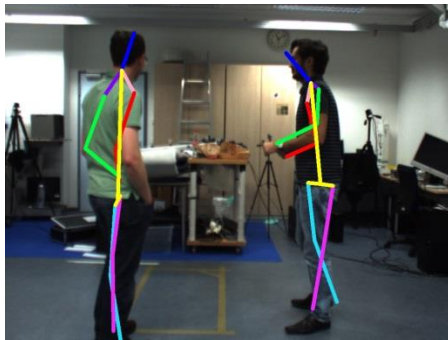
Pose Inference

Experiments

Conclusion

# 3D Human Pose Estimation

## Multi-View – Multiple Human

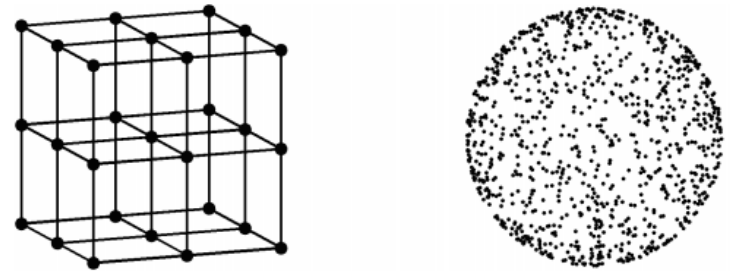
[Introduction](#)[3DPS Model](#)[Pose Inference](#)[Experiments](#)[Conclusion](#)

# 3D Human Pose Estimation

- Challenges

- Large state space (6DoF)

## Multi-View – Multiple-Human



M. Burenius, et al. 3D pictorial structures for multiple view articulated pose estimation. In CVPR, 2013.

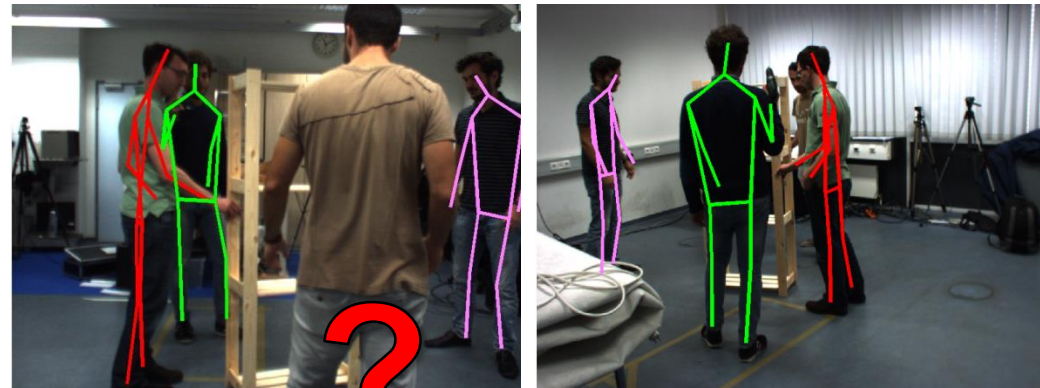
$$|\Omega_T| = 32^3 \times |\Omega_R| = 8^3 \times N$$

# 3D Human Pose Estimation

- Challenges

- Large state space (6DoF)
- Unknown identity

## Multi-View – Multiple-Human



Introduction

3DPS Model

Pose Inference

Experiments

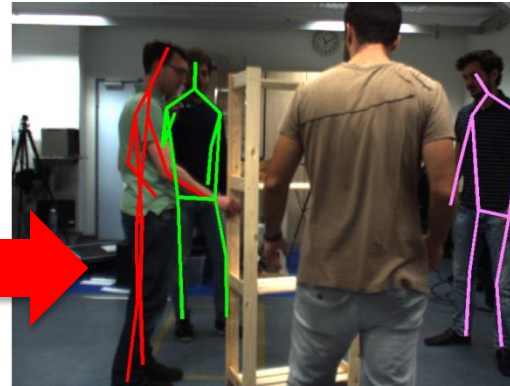
Conclusion

# 3D Human Pose Estimation

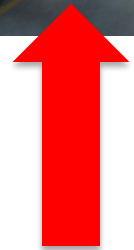
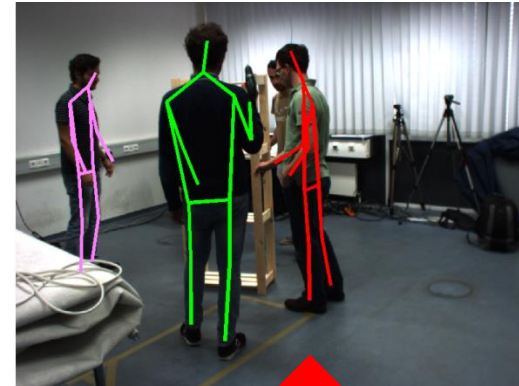
- Challenges

- Large state space (6DoF)
- Unknown identity
- Occlusion

**Occlusion  
between  
individuals**



## Multi-View – Multiple-Human



**Self-occlusion**

Introduction

3DPS Model

Pose Inference

Experiments

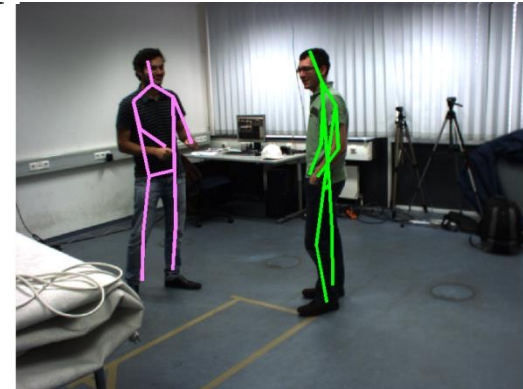
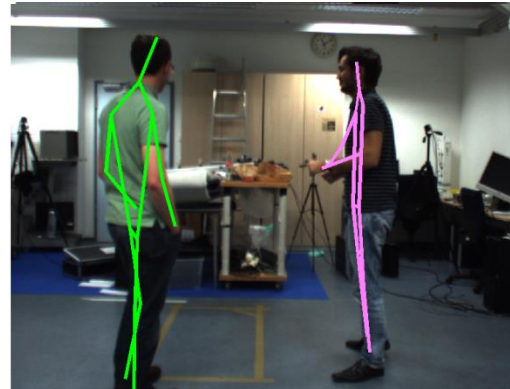
Conclusion

# 3D Human Pose Estimation

- **Challenges**

- Large state space (6DoF)
- Unknown identity
- Occlusion
- Dynamic environment
- Unconstrained motion

## Multi-View – Multiple-Human



# Related Work

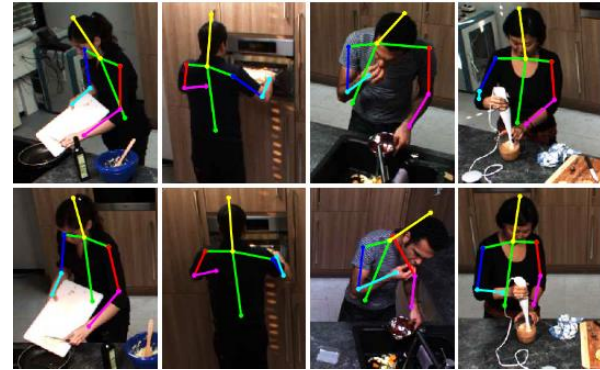
- **Pictorial structures model**
  - Fischler, Martin A., and Robert A. Elschlager, IEEE Transactions 1973.
  - Felzenszwalb, Pedro F., and Daniel P. Huttenlocher, IJCV 2005.



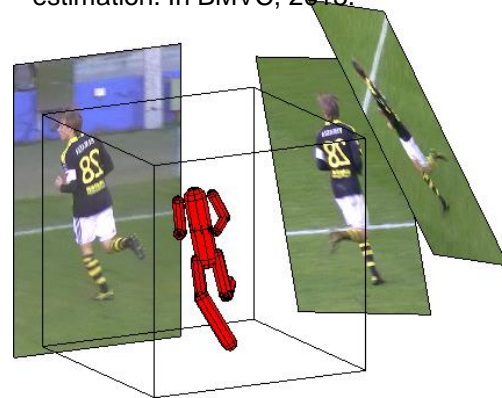
M. Andriluka, S. Roth, and B. Schiele. Pictorial structures revisited: People detection and articulated pose estimation. In CVPR, 2009.

## Related Work

- **Multi-view pictorial structures**
  - S. Amin et al., BMVC 2013
    - Skeleton inference in 2D
    - Triangulation (Single 3D skeleton)
  - M. Burenius et al., CVPR 2013
    - 3D volume discretization
    - Single 3D skeleton inference
- **Loose-limbed people**
  - L. Sigal et al., IJCV 2011
    - Continuous state space



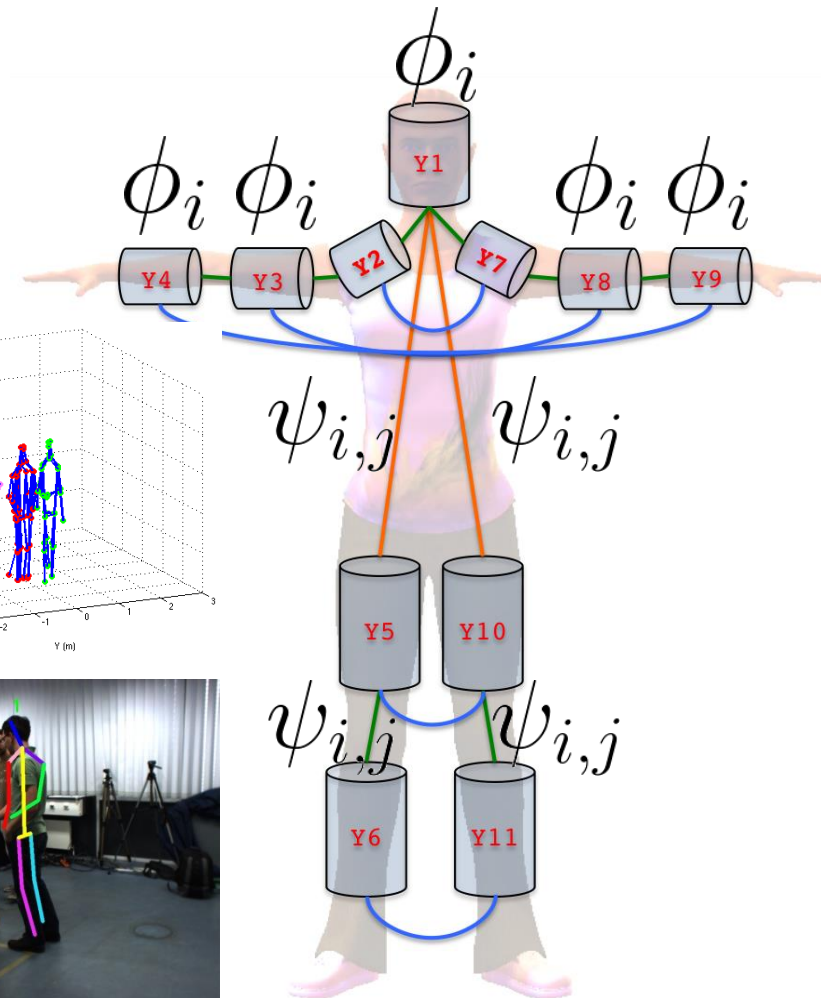
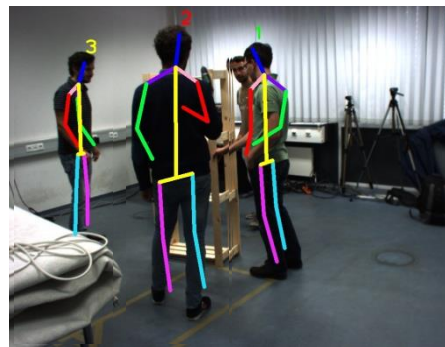
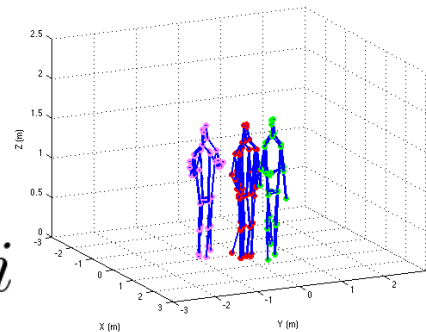
S.Amin, M.Andriluka, M.Rohrbach, and B.Schiele. Multi- view pictorial structures for 3d human pose estimation. In BMVC, 2013.



M. Burenius, J. Sullivan, and S. Carlsson. 3d pictorial structures for multiple view articulated pose estimation. In CVPR, 2013.

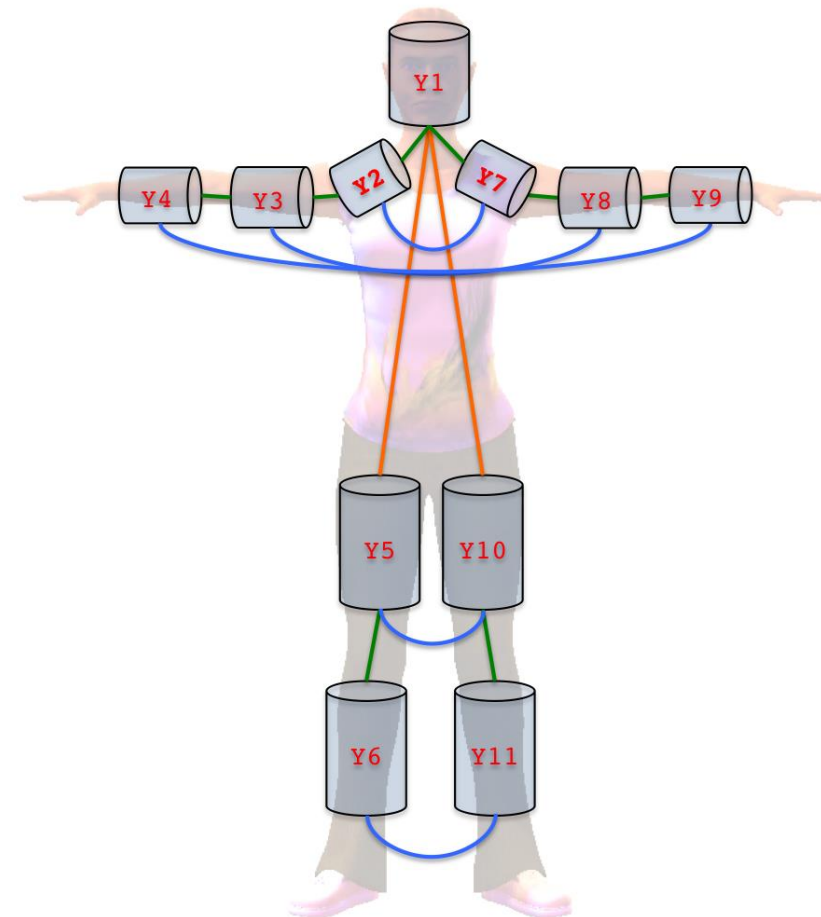
# Our Contributions

- 3D pictorial structures (3DPS) model
  - Single & multiple human pose estimation
- State space generation
  - Reduced search space
- Potential functions
  - Two- and multi-view unary  $\phi_i$
  - Body prior as pairwise  $\psi_{i,j}$
- Multiple human pose inference
  - Progressive skeleton parsing



# 3DPS Model

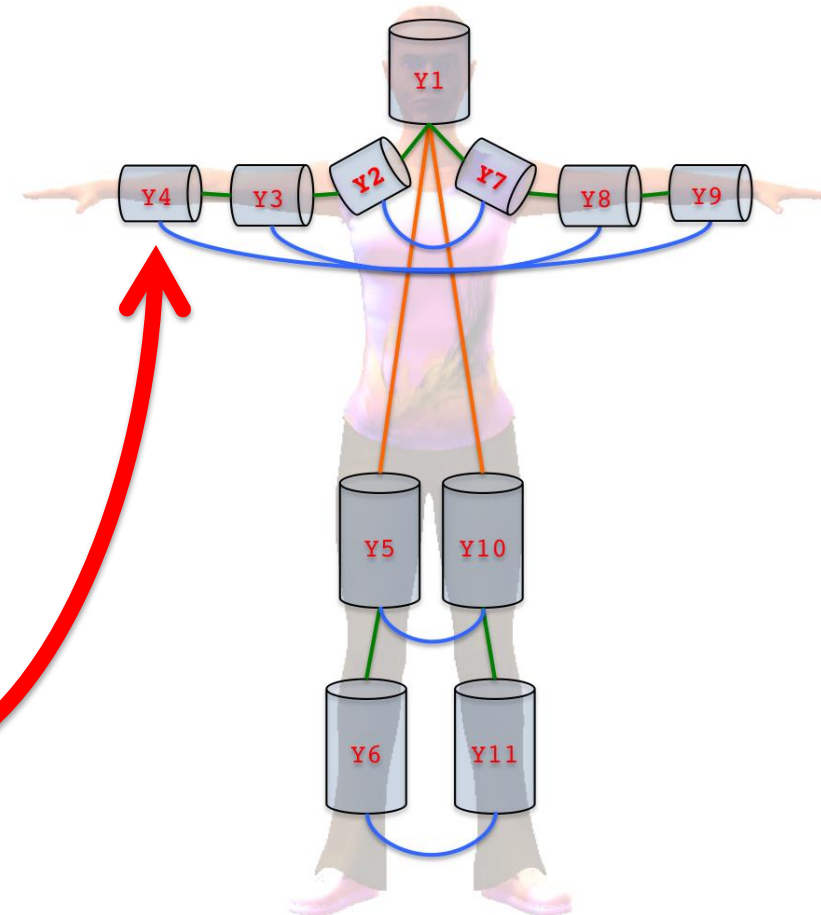
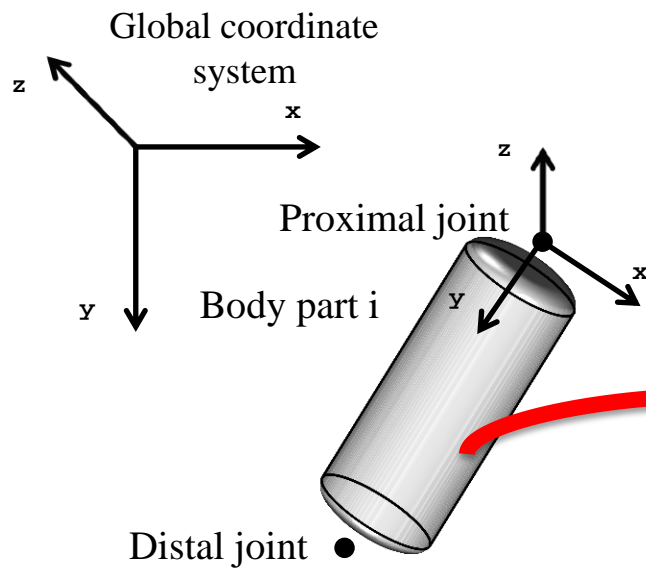
- Human body representation
  - Undirected graphical model
- Conditional Random Field (CRF)
  - Graph node – **body part** (random variable)
  - Graph edge – **body part** constraints
    - **Collision**
    - **Rotation**
    - **Translation**



## 3DPS Model

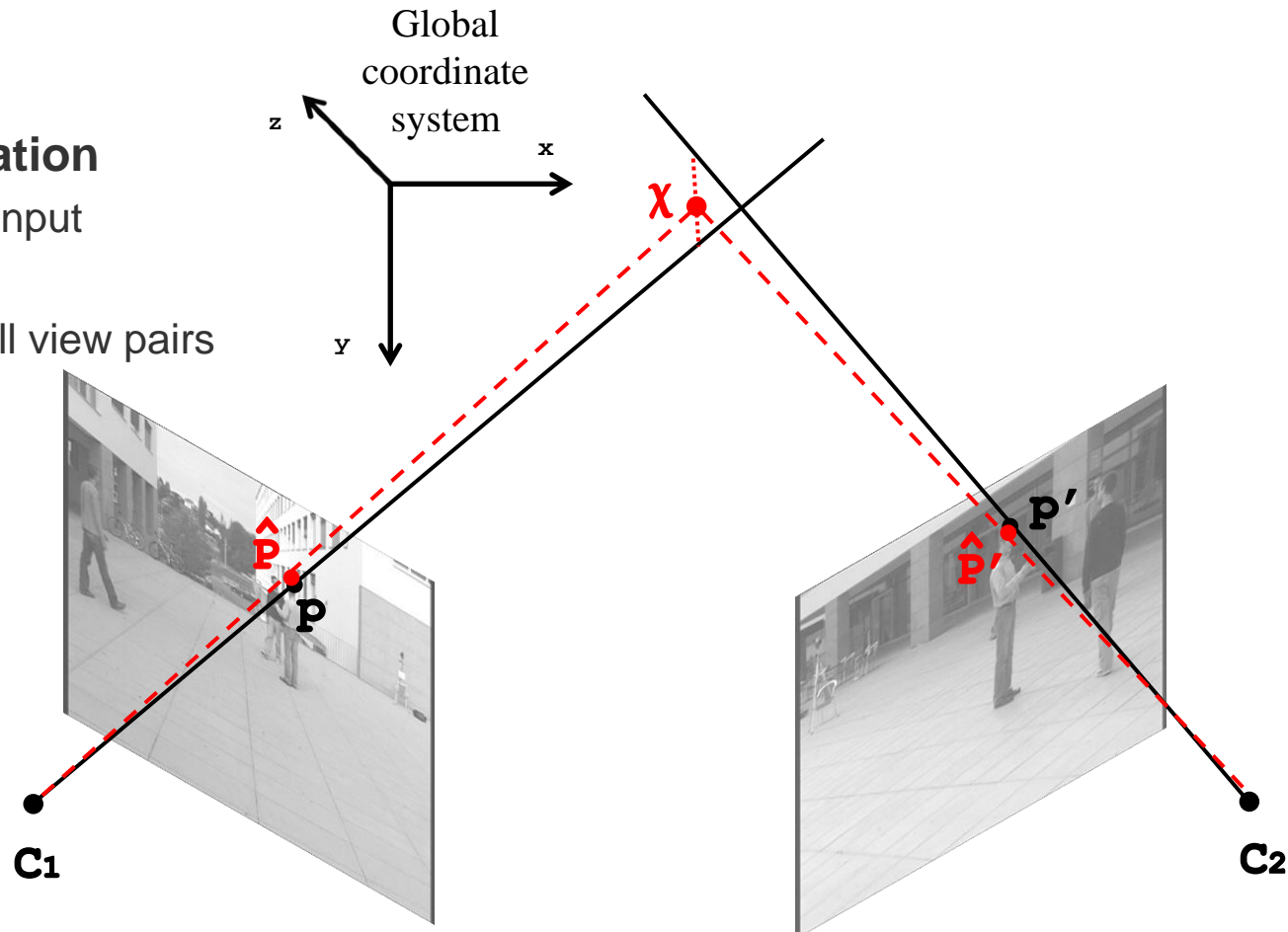
- **Body part configuration**

- Proximal and distal joint
- Orientation in 3-space



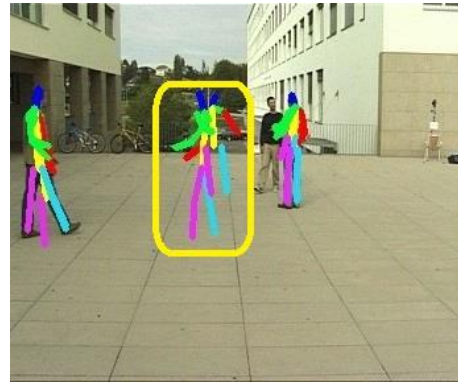
# State Space

- **Hypotheses generation**
  - 2D part detection input
  - Triangulation
  - Combinations of all view pairs

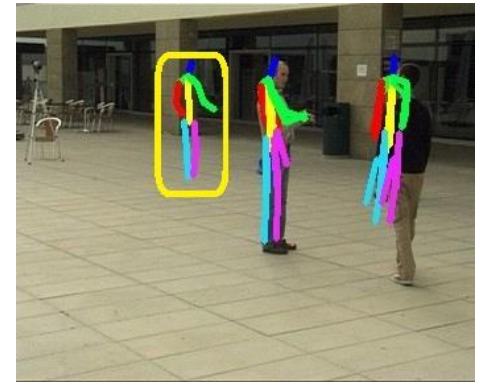


# State Space

- **Hypotheses generation**
  - 2D part detection
  - Triangulation
  - Combinations of all view pairs
- **Incorrect hypotheses**
  - False positive 2D detections
  - Triangulation of individuals with unknown identity



Camera A

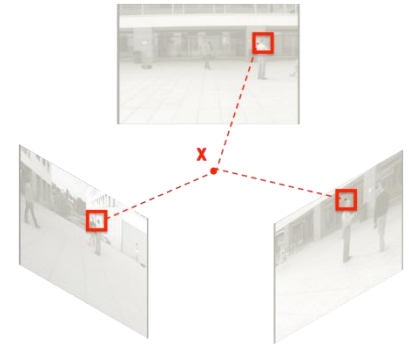
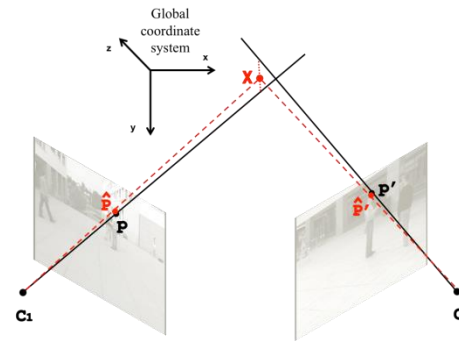


Camera B

# Posterior Estimation

## Potential Functions

- Unary
- Pairwise



Detection Confidence

Reprojection Error

Part Visibility

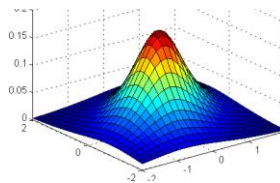
$$p(\mathbf{y} \mid \mathbf{x}) = \frac{1}{Z(\mathbf{x})} \prod_i^n \phi_i^{conf}(y_i, \mathbf{x}) \prod_i^n \phi_i^{repr}(y_i, \mathbf{x}) \prod_i^n \phi_i^{vis}(y_i, \mathbf{x}) \cdot$$

$$\prod_i^n \phi_i^{len}(y_i, \mathbf{x}) \cdot \prod_{(i,j) \in E_{kin}} \psi_{i,j}^{tran}(y_i, y_j) \cdot \prod_{(i,j) \in E_{kin}} \psi_{i,j}^{rot}(y_i, y_j) \cdot \prod_{(i,j) \in E_{col}} \psi_{i,j}^{col}(y_i, y_j)$$

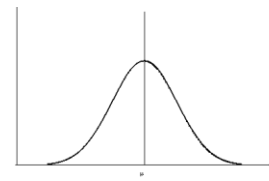
Part length



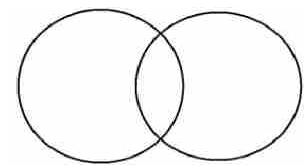
Translation



Rotation

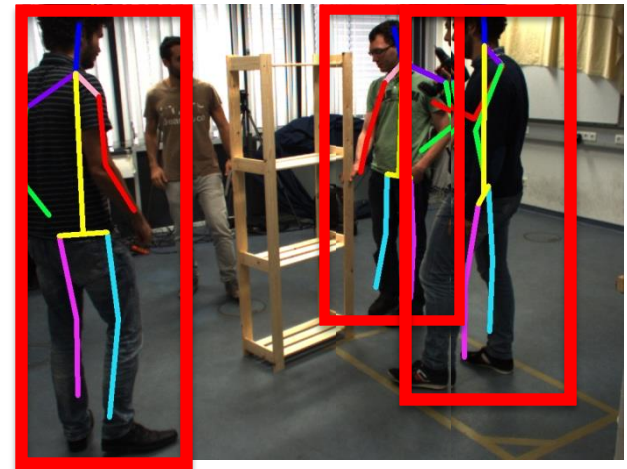
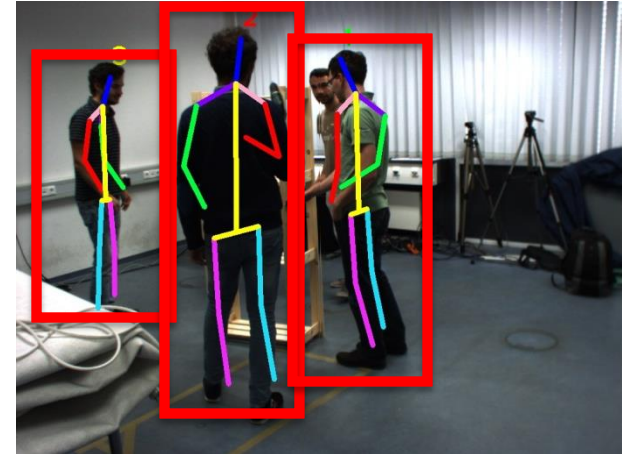


Collision



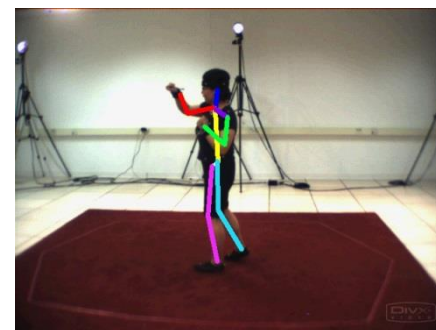
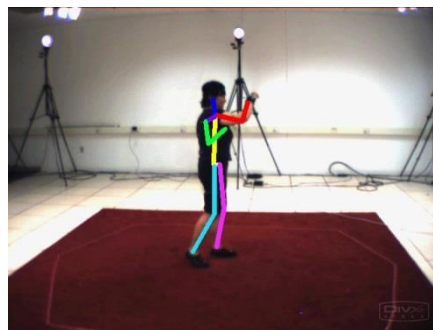
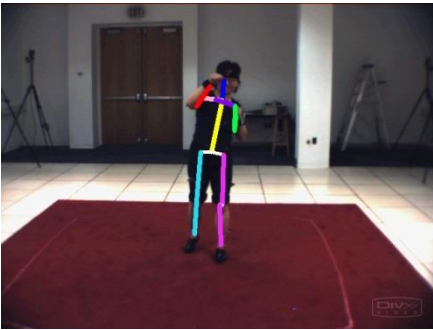
# Multiple Human Pose Inference

- Posterior estimation
  - Loopy belief-propagation
- Computation of individual number and location with a human detector
- Progressively parse skeletons
  - Sampling from the posterior
  - Projecting each sample across all views for verification

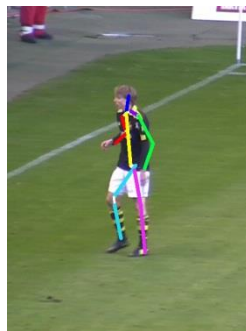


## Experiments (single-human)

- HumanEva-I dataset (Sigal et al., IJCV 2010)



- KTH Multiview Football Dataset II (Burenus et al. CVPR 2013)



Introduction

3DPS Model

Pose Inference

Experiments

Conclusion

## Experiments (single-human)

- HumanEva-I dataset (3D joint error in millimeters)



Sequence	Walk	Box
Amin et al. [2]	54.5	47.7
Sigal et al. [24]	89.7	-
Our method	68.3	62.7

- KTH Multiview Football Dataset II



	Bur. [8]	Our	Bur. [8]	Our
Body Parts	CAM2	CAM 2	CAM3	CAM 3
Arm	40.5	57.0	47.5	62.0
Legs	85.0	70.5	95.0	74.0
Average	62.7	63.8	71.2	68.0

Introduction

3DPS Model

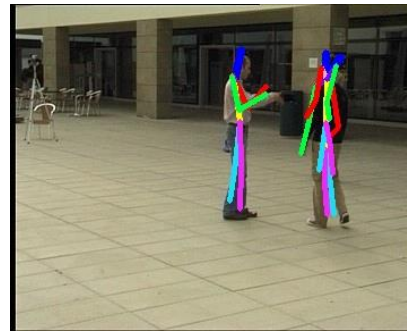
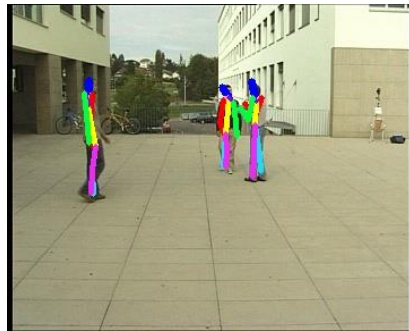
Pose Inference

Experiments

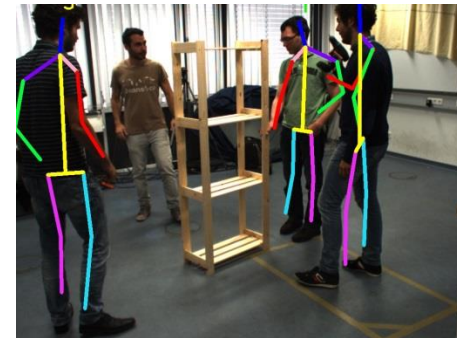
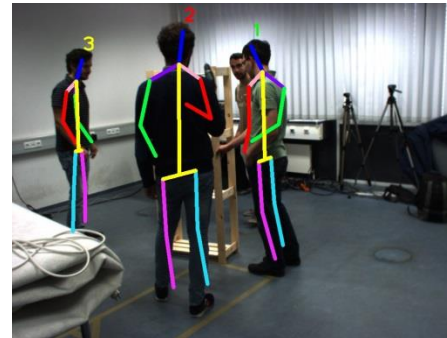
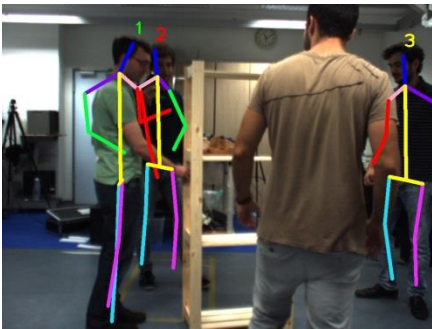
Conclusion

# Experiments (multiple-human)

- Campus dataset (proposed)



- Shelf dataset (proposed)



Introduction

3DPS Model

Pose Inference

Experiments

Conclusion

## Experiments (multiple-human)

- Campus dataset (PCP score)

Inference	Single Human		Multiple Human
	Amin et al. [2]	Our	Our
Actor 1	81	82	82
Actor 2	74	73	72
Actor 3	71	73	73
Average	75.3	<b>76.0</b>	<b>75.6</b>

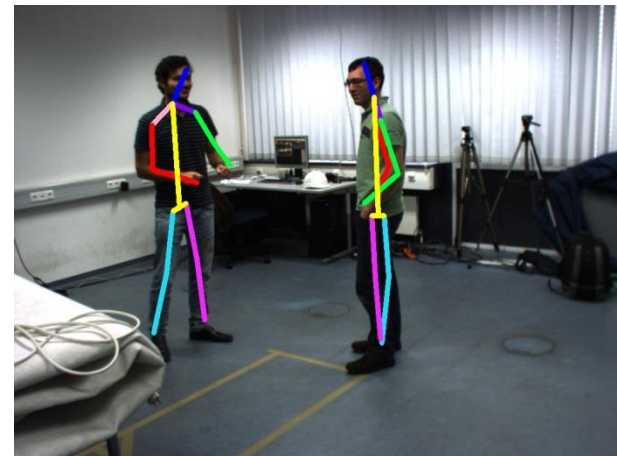
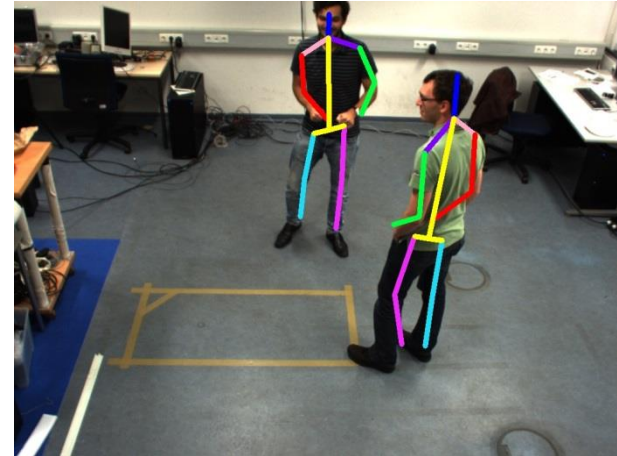
## Experiments (multiple-human)

- Shelf dataset (PCP score)

Inference	Single Human		Multiple Human
	Amin et al. [2]	Our	Our
Actor 1	65	66	66
Actor 2	62	65	65
Actor 3	81	83	83
Average	69.3	<b>71.3</b>	<b>71.3</b>

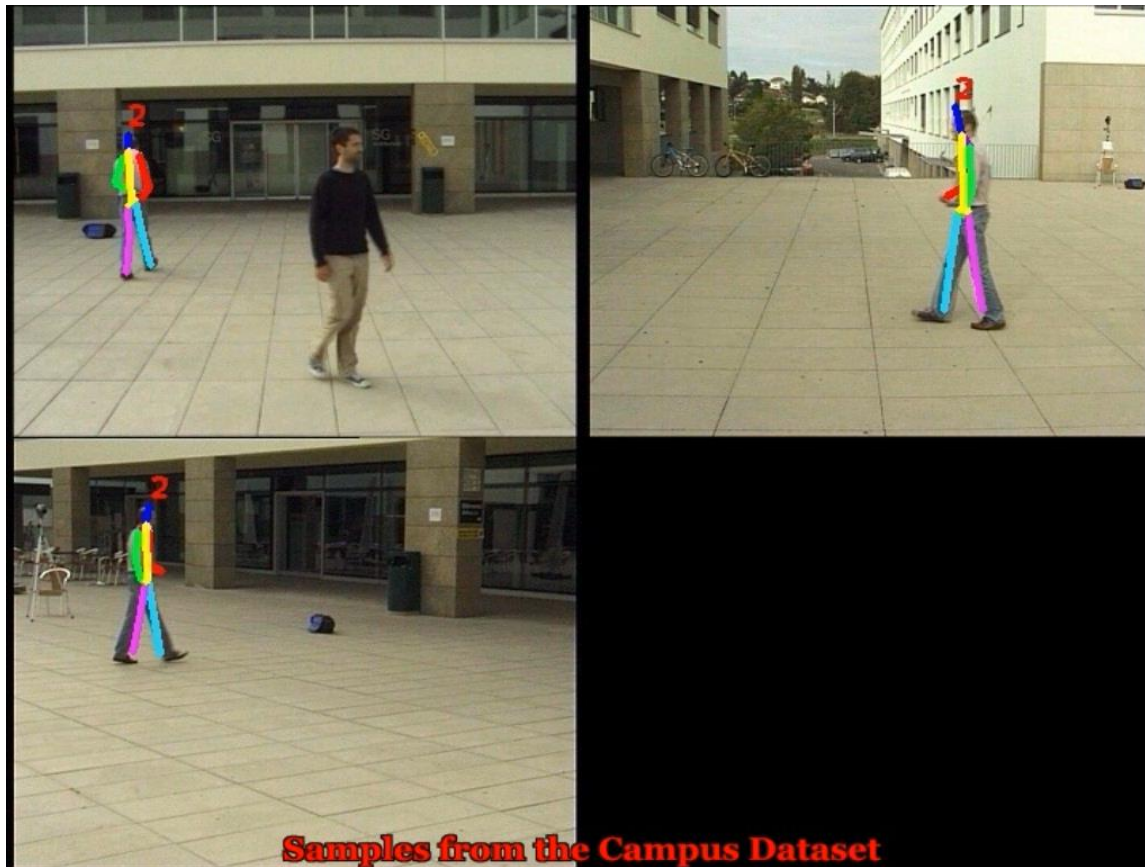
# Conclusion

- 3DPS model for recovering 3D human body poses
- Common state space between all individuals
- Multi-view potential functions
- Applicable to single or multiple human pose estimation



## Future Work

- Temporal consistency
  - Robustness against incorrect inferred poses and smoother solution
- Identity recover
  - Separate and smaller state space for each individual



# Thank you!

- Campus and Shelf datasets available at:
  - <http://campar.in.tum.de/Chair/MultiHumanPose>
- Poster ID: O-2C-6