Multi-View Multi-Person 3D Pose Estimation with Uncalibrated Camera Networks

Carnegie Mellon University

Yan Xu, Kris Kitani

BMVC 2022

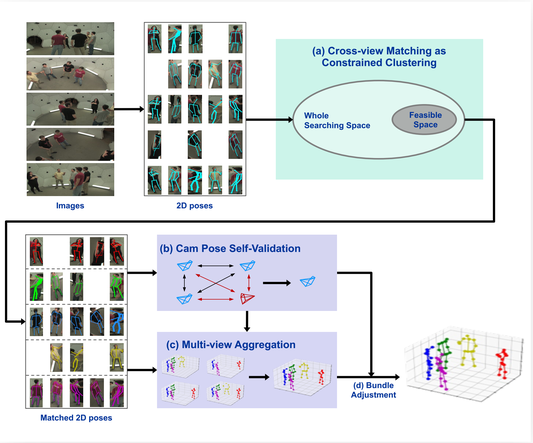
Background

1. Multi-view multi-person 3D human pose estimation mostly done in controlled env
2. In many real-life scenarios, camera poses are not likely to be readily available
3. We targets uncalibrated cam networks

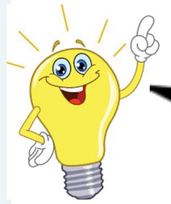
Method

Our method includes three steps:

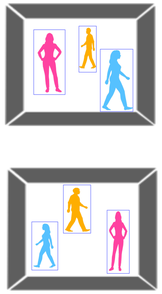
1. Match human boxes through clustering
2. Associate body joints for point corresponds
3. Solve cam and human pose



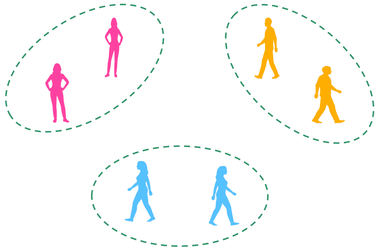
You are welcome to refer to our paper or reach out to the author (QR) for details.
Email: xuyan@cmu.edu.



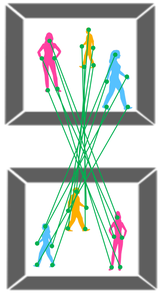
Clustering can be used to match human bounding boxes across views and obtain body joints 2D-2D correspondences!



Multi-view images and human detection



Match human across camera views through clustering



Associating body joints to obtain point correspondences

Cross-view matching constraints

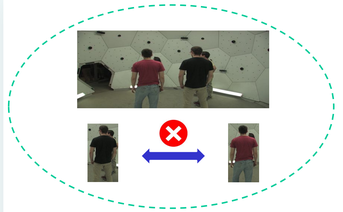
• **Constraint 1:** Cluster size larger than 2

At least two cameras for solving depth ambiguity

• **Constraint 2:** Cluster size smaller than number of cams

At most observed by N cameras

• **Constraint 3:** Ppl from same camera not in the same cluster



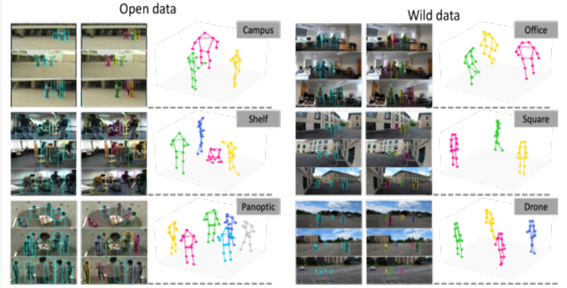
Observations from the same cam must be different ppl

Result

• Reaches SOTA performance without requiring camera poses, 3D data, or network training

Campus	CamPose	Training	Actor 1	Actor 2	Actor 3	Average
Huang *et al.* [22]	✓	✓	98.0	94.8	97.4	96.7
Tu *et al.* [55]	✓	✓	97.6	93.8	98.8	96.7
Zhang *et al.* [63]	✓	✓	98.2	94.1	97.4	96.6
Reddy *et al.* [48]	✓	✓	97.9	**95.2**	99.1	97.4
Belagiannis *et al.* [4]	✓	-	93.5	75.7	84.4	84.5
Ershadi *et al.* [15]	✓	-	94.2	92.9	84.6	90.6
Dong *et al.* [13]	✓	-	97.6	93.3	98.0	96.3
Perez-Yus *et al.* [46]	✓	-	98.4	93.4	98.3	96.7
Ours	-	-	**99.0**	94.7	**99.6**	**97.8**
Shelf	**CamPose**	**Training**	**Actor 1**	**Actor 2**	**Actor 3**	**Average**
Huang *et al.* [22]	✓	✓	98.8	96.2	97.2	97.4
Tu *et al.* [55]	✓	✓	99.3	94.1	97.6	97.0
Zhang *et al.* [63]	✓	✓	99.3	95.1	97.8	97.4
Reddy *et al.* [48]	✓	✓	99.1	96.3	98.3	98.2
Wu *et al.* [58]	✓	✓	99.3	96.5	97.3	97.7
Belagiannis *et al.* [4]	✓	-	75.3	69.7	87.6	77.5
Ershadi *et al.* [15]	✓	-	93.3	75.9	94.8	88.0
Dong *et al.* [13]	✓	-	98.8	94.1	97.8	96.9
Perez-Yus *et al.* [46]	✓	-	98.9	92.3	97.8	96.5
Ours	-	-	**99.6**	**95.2**	**98.5**	**97.8**

• Generalizes well to the in-the-wild data



• Can estimate cam poses across various settings