TesseTrack: End-to-End Learnable Multi-Person Articulated 3D Pose Tracking

N. Dinesh Reddy[1] Laurent Guigues[2*] Leonid Pishchulin[2*] Jayan Eladeth[2] Srinivasa G. Narasimhan[1]

[1]Carnegie Mellon University [2]Amazon [*]Equal Contribution

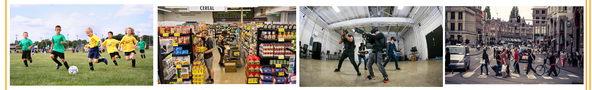
Project: http://www.cs.cmu.edu/~ILIM/projects/IM/TesseTrack/

SCAN ME

CVPR VIRTUAL JUNE 19-25

Introduction

Goal: Multi-Person 3D Pose Tracking using End-to-End Learning

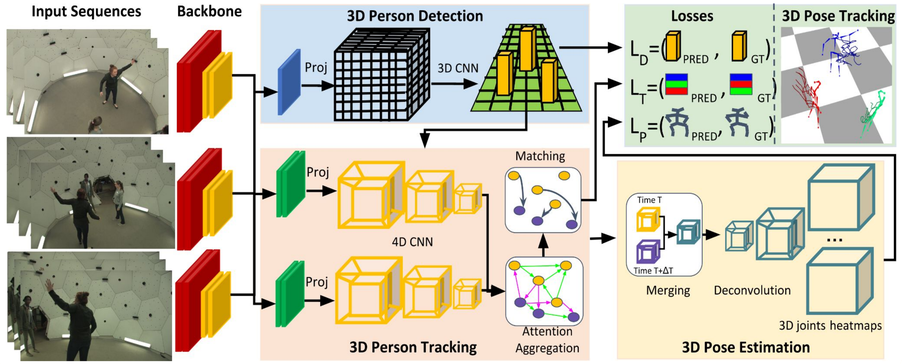


Sports Shopping AR/VR Automation

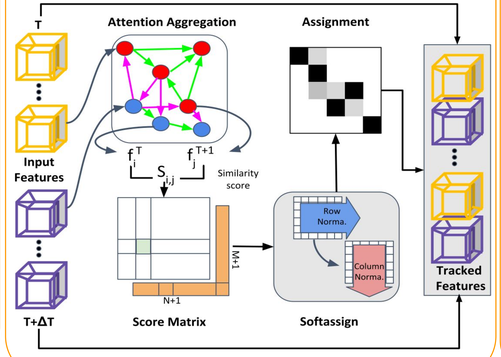
Contributions:

- A novel spatio-temporal formulation that allows simultaneous 3D body joint reconstruction and tracking of multiple individuals.
- A novel learnable tracking formulation that allows extending person-specific spatio-temporal representation learning.
- A novel evaluation framework for multi-person 3D pose tracking.
- An in-depth ablation study of the proposed approach and thorough comparisons to current methods on several standard benchmarks.

System Pipeline



Learnable Tracking Framework



SOTA Comparisions

Method	Actor-1	Actor-2	Actor-3	Total
Belagiannis	93.5	75.7	84.4	84.5
Ershadi	94.2	92.9	84.6	90.6
Dong	97.2	93.3	98.0	96.3
Tu Et al.	97.6	93.8	98.8	96.7
TesseTrack	**97.9**	**95.2**	**99.1**	**97.4**

Campus Dataset

Method	Actor-1	Actor-2	Actor-3	Total
Belagiannis	93.5	75.7	84.4	84.5
Ershadi	94.2	92.9	84.6	90.6
Dong	97.2	93.3	98.0	96.3
Tu Et al.	97.6	93.8	98.8	96.7
TesseTrack	**97.9**	**95.2**	**99.1**	**97.4**

Shelf Dataset

	Monocular Method(MPJPE,mm)					Multi-View Method(MPJPE,mm)				
Method	Martinez	Iskakov	Pavollo	Cheng	Tessetrack	Martinez	Pavlakos	Padoy	Iskakov	TesseTrack
All	62.9	49.9	46.8	40.1	**44.6**	57.0	56.7	49.1	20.8	**18.7**

Human3.6M Dataset

	Multi-View(5 Views)		Monocular	
Method	Tu et al.	TesseTrack	Tu Et al.	TesseTrack
MPJPE(mm)	17.7	**7.3**	51.1	**18.9**

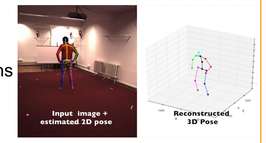
Panoptic Dataset

State-of-the Art

Single person 3D Pose Estimation:

✚ Good Reconstruction accuracy

▬ Fails with Multi-person or occlusions

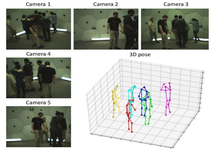


Francesc Moreno-Noguer. **3d human pose estimation from a single image via distance matrix regression.**. In CVPR 2017.
Julieta Martinez, Rayat Hossain, et al.. **A simple yet effective baseline for 3d hu-man pose estimation**. In ICCV, 2017.

Multi-Person 3D Pose Estimation:

✚ Good 3D learning Frameworks

▬ Fails Accurate Tracking

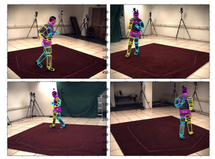


Hanyue Tu, et al. **Voxelpose:Towards multi-camera 3d human pose estimation in wild environment**. ECCV, 2020.
Je Sun, Bin Xiao, et al.. **Deephigh-resolution representation learning for human pose esti-mation**. In CVPR, 2019

Multi-Person 3D Pose Tracking:

✚ Working tracking Frameworks

▬ PieceWise Tracking modules



Mykhaylo Andriluka, Et al..**Monocular 3d pose estimation and tracking by detection.** In CVPR 2010
Andrei Zanfir et al.. **Monocular 3d pose and shape estimation of multiple people in natural scenes**.in cvpr 2018

Losses

Detection Loss:

$$L_D^t = \sum_{w=1}^{W}\sum_{h=1}^{H}\sum_{d=1}^{D} ||V_{Pred}^{w,h,d} - V_{GT}^{w,h,d}||$$

Tracking Loss:

$$L_T^t = -\sum_{(i,j)\in G} \log P_{i,j}$$

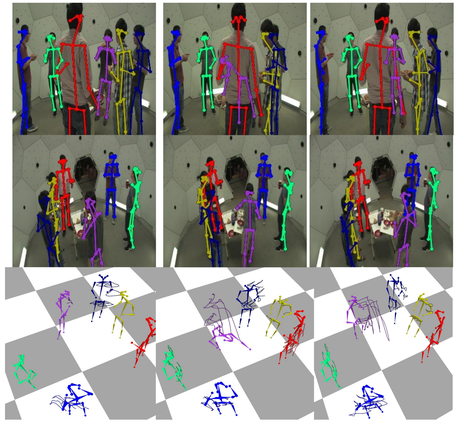
3D Pose Estimation Loss:

$$L_P^{t,d} = \sum_{q=1}^{Q} [||k_{Pred}^q - k_{GT}^q||_1 - \beta.\log(T_{Pred}^q(k_{GT}^q))]$$

Total Loss:

$$L = \sum_{t\in D}\left[L_D^t + \alpha L_T^t + \gamma \sum_{p\in TP(t)} L_P^{t,p}\right]$$

Qualitative Results



Panoptic Dataset

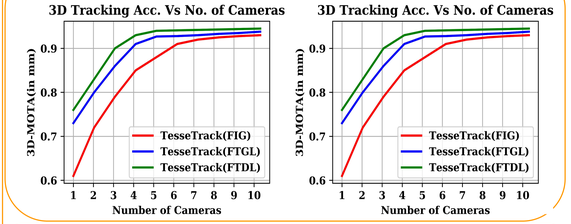
Wild Dataset

3D Pose Tracking Accuracy(3D-MOTA)

Method	Neck	head	Shou.	Elbow	Wrist	Hip	Knee	Ankle	Avg
FIG	89.7	87.4	90.8	88.0	82.2	92.7	89.1	92.4	87.6
FTGL	93.9	91.7	93.0	92.1	87.4	94.4	**93.9**	94.6	92.1
FTDL	**94.6**	**93.6**	**93.4**	**92.7**	**88.2**	**94.7**	93.8	**95.0**	**94.1**

Panoptic Dataset

Effect of No. of Cameras on Accuracy



Acknowledgement

This paper was supported in parts by NSF Grants IIS1900821 and CNS-2038612, DOT RITA Mobility-21 Grant 69A3551747111, and a PhD fellowship from Amazon.