Part-aware Measurement for Robust Multi-View Multi-Human 3D Pose Estimation and Tracking Supplementary Material

Hau Chu[1,3] Jia-Hong Lee[1,3] Yao-Chih Lee[2,5] Ching-Hsien Hsu[2,6]
Jia-Da Li[1,3] Chu-Song Chen[1,4]
[1]National Taiwan University [2]Academia Sinica, Taiwan
[3]{hauchu8733,jiahonghenrylee,jiadali}@ntu.edu.tw, [4]chusong@csie.ntu.edu.tw
[5]yclee1231@iis.sinica.edu.tw, [6]applychristw@gmail.com

1. Supplementary Material

1.1. Implementation Details

Paremeters Selection. In this work we have several parameters: α_{2D}, α_{epi}, τ, ε are thresholds of 2D velocity, epipolar distance, time interval and number of positive affinities respectively, λ_α is a penalty rate of time interval. Here we show our empirical selections of parameters for each dataset in Table 1. α_{2D}, α_{epi}, τ, ε are based on the frame size and the distance between people and camera. For Campus [1], the image size is 360×288 and the actors are far from cameras. Therefore, α_{2D} and α_{epi} is set to be smaller number. Yet ε adjust with a strict value since actors are mostly captured completely, and we expect that 2D poses can all be accurately estimated. On the other hand, Shelf [1] and Panoptic [2] have larger image size and humans are captured in a small area which is close to cameras. Thus, occlusion and out-of-view often occur. We then define the parameter with a more flexible value. For other two parameters, τ and λ_α basically depends on the fps of video, *e.g*. the three datates are all captured at 25 fps.

Dataset	Campus	Shelf	Panoptic
α_{2D}	30	70	60
α_{epi}	15	60	30
τ	3	3	3
ε	14	10	10
λ_α	3	3	3

Table 1: Parameters selection for different datasets.

Initialization Procedure. To introduce our 3D pose initialization procedure more clearly, it is detailed in Algorithm 1.

1.2. Qualitative Results

Here, we demonstrate more qualitative results of our approach on three datasets in Figure 1, Figure 2 and Figure 3.

Algorithm 1: Initialization Procedure

Input:
Unmatched 2D poses $\mathbb{U}_c | c \in \mathbf{C}$
Previously tracked 3D skeletons $\mathbf{X}_{t'} \in \mathbf{P}$ at time t'
Output:
New tracked skeletons $\{\mathbf{X}_t\}$ at time t

Initialization: $\mathbb{U} \leftarrow \mathit{unmatched\ 2D\ poses\ of\ Camera\ 1}$
foreach $c \in \{\mathbf{C} - c_1\}$ **do**
  $\mathbb{U}_c \leftarrow \mathit{unmatched\ 2D\ poses\ of\ Camera\ c}$
  $\mathbf{E} \in \mathbb{R}^{|\mathbb{U}| \times |\mathbb{U}_c|}$
  foreach $\boldsymbol{x}_t \in \mathbb{U}$ **do**
    foreach $\boldsymbol{x}_{t,c} \in \mathbb{U}_c$ **do**
      $\mathbf{E} \leftarrow \mathrm{EpipolarConstraint}(\mathbb{U}, \mathbb{U}_c)$
      $\mathit{Match}(\boldsymbol{x}_t, \boldsymbol{x}_{t,c}), \mathbb{U}'_c \leftarrow \mathrm{HungarianAlgorithm}(\mathbf{E})$
      $\mathbb{U} \leftarrow \mathbb{U} \cup \mathbb{U}'_c$
    end
  end
end
foreach $\boldsymbol{x}_{cluster} \in \mathbb{U}$ **do**
  if $\mathrm{Length}(\boldsymbol{x}_{cluster}) \geq 2$ **then**
    $\hat{\mathbb{P}} \leftarrow \mathrm{JointsFilter}(\boldsymbol{x}_{cluster})$
    $\mathbf{X}_t \leftarrow \mathrm{3DReconstruction}(\hat{\mathbb{P}})$
    $\mathbf{P} \leftarrow \mathbf{P} \cup \{\mathbf{X}_t\}$
  end
end

References

[1] Vasileios Belagiannis, Sikandar Amin, Mykhaylo Andriluka, Bernt Schiele, Nassir Navab, and Slobodan Ilic. 3d pictorial structures revisited: Multiple human pose estimation. *IEEE transactions on pattern analysis and machine intelligence*, 38(10):1929–1942, 2015.

[2] Hanbyul Joo, Tomas Simon, Xulong Li, Hao Liu, Lei Tan, Lin Gui, Sean Banerjee, Timothy Godisart, Bart Nabbe, Iain Matthews, et al. Panoptic studio: A massively multiview system for social interaction capture. *IEEE transactions on pattern analysis and machine intelligence*, 41(1):190–204, 2017.

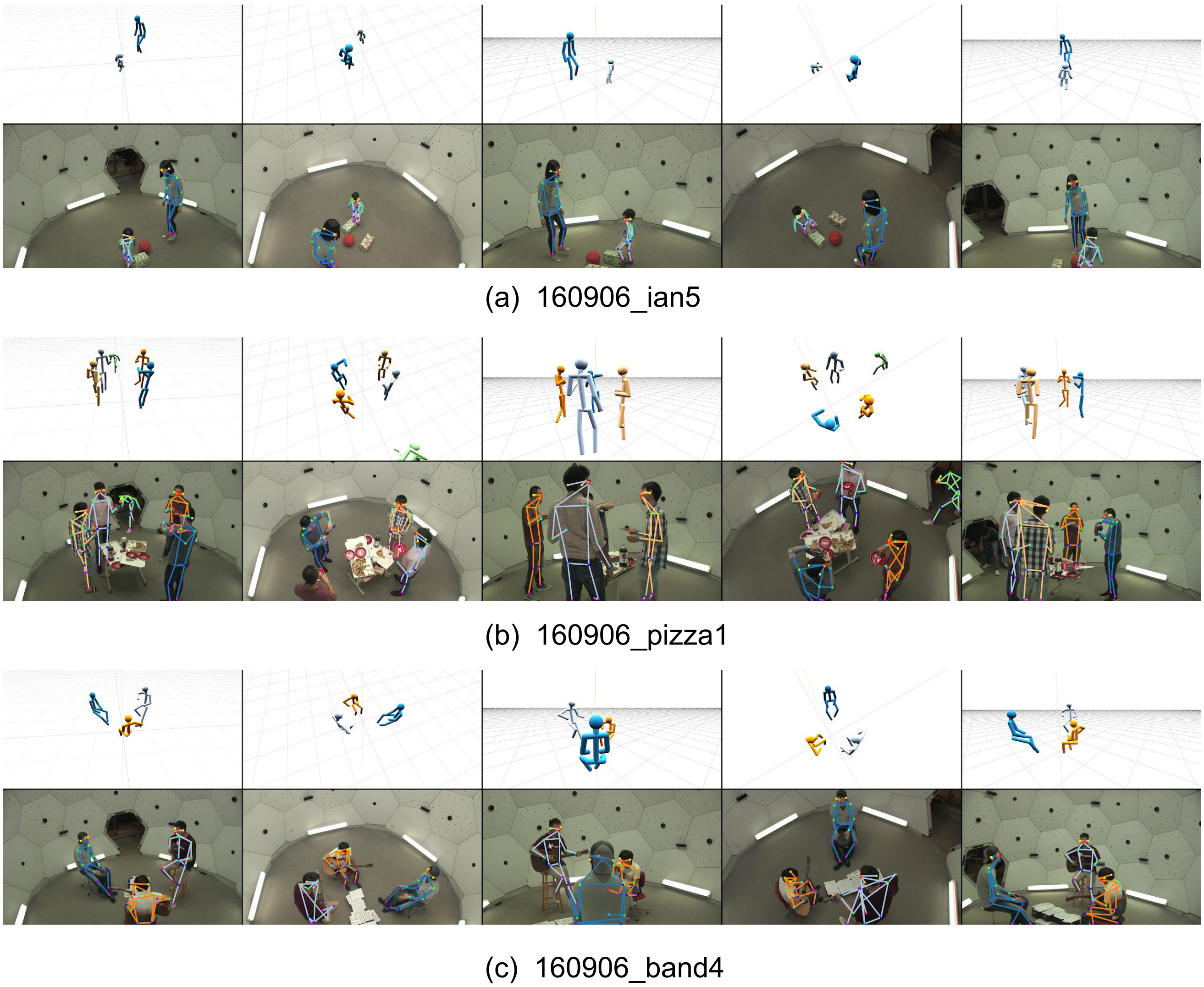

Figure 1: Qualitative results of Panoptic, three sub-datasets is demonstrated.

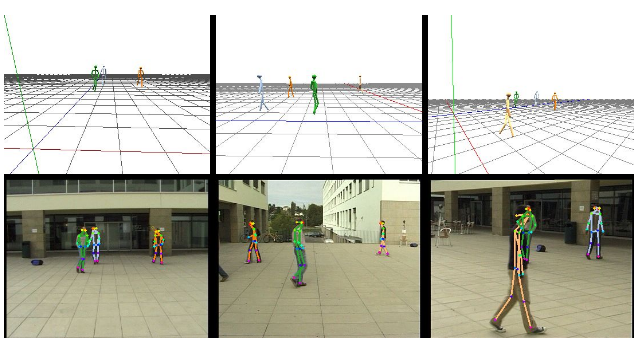

Figure 2: Qualitative result of Campus

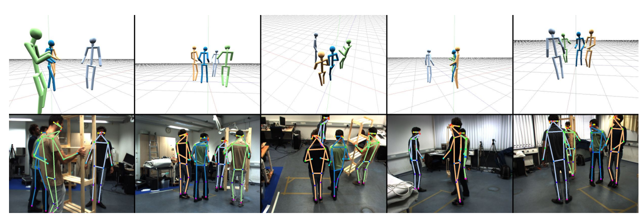

Figure 3: Qualitative result of Shelf