Supplementary Material for "Accurate 3D Hand Pose Estimation for Whole-Body 3D Human Mesh Estimation"

Gyeongsik Moon[1] Hongsuk Choi[1] Kyoung Mu Lee[1,2]

[1]Dept. of ECE & ASRI, [2]IPAI, Seoul National University, Korea
{mks0601,redarknight,kyoungmu}@snu.ac.kr

In this supplementary material, we present more experimental results that could not be included in the main manuscript due to the lack of space.

A. Qualitative comparisons

A.1. MSCOCO

Figure A and B show that our Hand4Whole produces more accurate results on in-the-wild images of MSCOCO. In particular, ours produce much better 3D hands results.

A.2. 3DPW

The video in this link[1] shows that our Hand4Whole produces more accurate and plausible expressive 3D human mesh than ExPose [2] and FrankMocap [10] on videos of 3DPW [11]. In particular, ours achieves much better and stable hands results when hands are invisible by using body and hand MCP joint features. On the other hand, ExPose and FrankMocap do not use body features, which results in implausible 3D hands. Hand4Whole, ExPose, and FrankMocap are run on every single frame without leveraging temporal information. We did not apply any post-processing, such as average filtering, on the outputs. The results of ExPose and FrankMocap are obtained by their officially released codes.

License of the Used Assets

- MSCOCO dataset [7] belongs to the COCO Consortium and are licensed under a Creative Commons Attribution 4.0 License.
- Human3.6M dataset [5]'s licenses are limited to academic use only.
- MPII dataset [1] is released for academic research only and it is free to researchers from educational or research institutes for non-commercial purposes.
- 3DPW dataset [11] is released for academic research only and it is free to researchers from educational or research institutes for non-commercial purposes.
- FreiHAND dataset [12] is released for academic research only and it is free to researchers from educational or research institutes for non-commercial purposes.
- FFHQ dataset [6]'s individual images were published in Flickr by their respective authors under either Creative Commons BY 2.0, Creative Commons BY-NC 2.0, Public Domain Mark 1.0, Public Domain CC0 1.0, or U.S. Government Works license. The dataset itself (including JSON metadata, download script, and documentation) is made available under Creative Commons BY-NC-SA 4.0 license by NVIDIA Corporation.
- Stirling dataset [4] is released for academic research only and it is free to researchers from educational or research institutes for non-commercial purposes.
- EHF dataset [9] is released for academic research only and it is free to researchers from educational or research institutes for non-commercial purposes.
- AGORA dataset [8] is released for academic research only and it is free to researchers from educational or research institutes for non-commercial purposes.
- ExPose [2] codes are released for academic research only and it is free to researchers from educational or research institutes for non-commercial purposes.
- FrankMocap [10] codes are CC-BY-NC 4.0 licensed.
- PIXIE [3] codes are released for academic research only and it is free to researchers from educational or research institutes for non-commercial purposes.

[1] https://www.youtube.com/watch?v=Ym_CH8yxBso

Figure A. Qualitative comparison of the proposed Hand4Whole, ExPose [2], FrankMocap [10], and PIXIE [3] on MSCOCO.

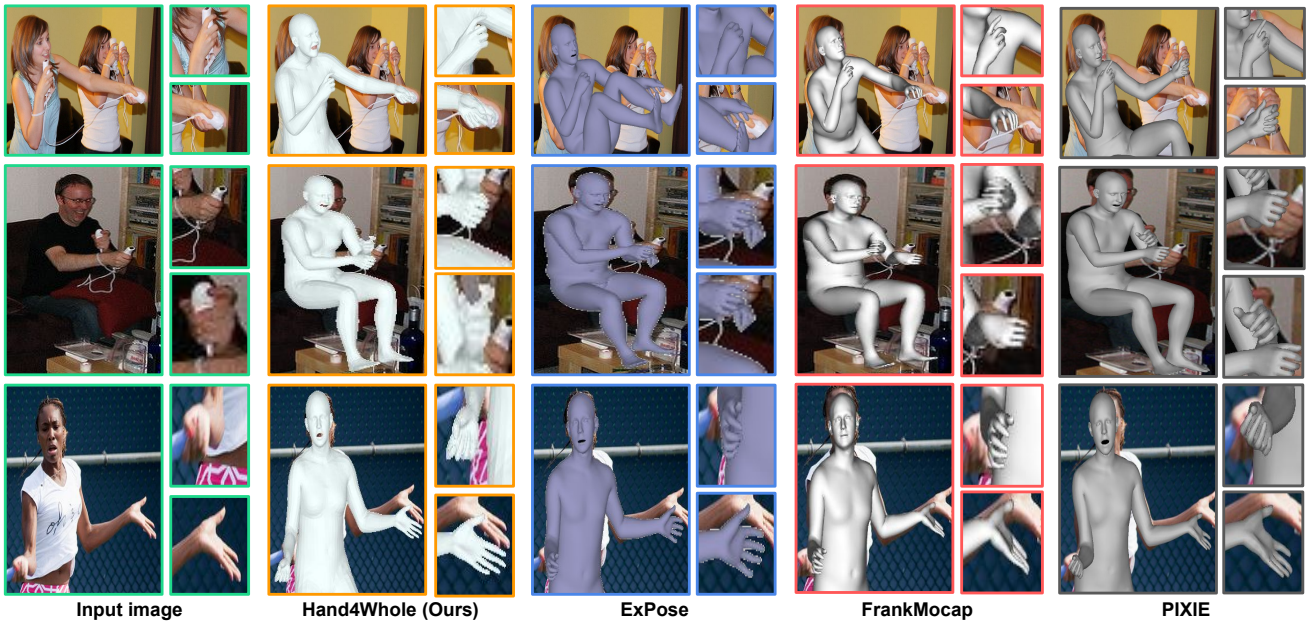

Figure B. Qualitative comparison of the proposed Hand4Whole, ExPose [2], FrankMocap [10], and PIXIE [3] on MSCOCO.

References

[1] Mykhaylo Andriluka, Leonid Pishchulin, Peter Gehler, and Bernt Schiele. 2D human pose estimation: New benchmark and state of the art analysis. In *CVPR*, 2014.

[2] Vasileios Choutas, Georgios Pavlakos, Timo Bolkart, Dimitrios Tzionas, and Michael J Black. Monocular expressive body regression through body-driven attention. In *ECCV*, 2020.

[3] Yao Feng, Vasileios Choutas, Timo Bolkart, Dimitrios Tzionas, and Michael J. Black. Collaborative regression of expressive bodies using moderation. In *3DV*, 2021.

[4] Zhen-Hua Feng, Patrik Huber, Josef Kittler, Peter Hancock, Xiao-Jun Wu, Qijun Zhao, Paul Koppen, and Matthias Rätsch. Evaluation of dense 3D reconstruction from 2D face images in the wild. *FG*, 2018.

[5] Catalin Ionescu, Dragos Papava, Vlad Olaru, and Cristian Sminchisescu. Human3.6M: Large scale datasets and predictive methods for 3D human sensing in natural environments. *TPAMI*, 2014.

[6] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *CVPR*, 2019.

[7] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft COCO: Common objects in context. In *ECCV*, 2014.

[8] Priyanka Patel, Chun-Hao P. Huang, Joachim Tesch, David T. Hoffmann, Shashank Tripathi, and Michael J. Black. AGORA: Avatars in geography optimized for regression analysis. In *CVPR*, 2021.

[9] Georgios Pavlakos, Vasileios Choutas, Nima Ghorbani, Timo Bolkart, Ahmed AA Osman, Dimitrios Tzionas, and Michael J Black. Expressive body capture: 3D hands, face, and body from a single image. In *CVPR*, 2019.

[10] Yu Rong, Takaaki Shiratori, and Hanbyul Joo. FrankMocap: A monocular 3d whole-body pose estimation system via regression and integration. In *ICCV Workshop*, 2021.

[11] Timo von Marcard, Roberto Henschel, Michael J Black, Bodo Rosenhahn, and Gerard Pons-Moll. Recovering accurate 3D human pose in the wild using IMUs and a moving camera. In *ECCV*, 2018.

[12] Christian Zimmermann, Duygu Ceylan, Jimei Yang, Bryan Russell, Max Argus, and Thomas Brox. FreiHAND: A dataset for markerless capture of hand pose and shape from single RGB images. In *ICCV*, 2019.