Supplementary Material to: Recovering Accurate 3D Human Pose in The Wild Using IMUs and a Moving Camera

Timo von Marcard[1], Roberto Henschel[1], Michael J. Black[2], Bodo Rosenhahn[1], and Gerard Pons-Moll[3]

[1] Leibniz Universität Hannover, Germany
[2] MPI for Intelligent Systems, Germany
[3] MPI for Informatics, Saarland Informatics Campus, Germany
{marcard,henschel,rosenhahn}@tnt.uni-hannover.de, black@tue.mpg.de, gpons@mpi-inf.mpg.de

This document contains additional experiments to the paper "Recovering Accurate 3D Human Pose in The Wild Using IMUs and a Moving Camera"[1]. These experiments validate different aspects of VIP, our proposed method for combining IMU-based tracking with a single hand-held camera, and provide further details to the proposed 3DPW.

In Section 1, we validate that the explicit modeling of IMU heading errors is an important ingredient of the proposed method. In Section 2 we evaluate tracking accuracy of VIP for an additional IMU sensor setup. In order to demonstrate the challenges of our newly recorded dataset in comparison to existing datasets, we evaluate three monocular 3D pose estimation methods in Section 3.

1 Modeling Heading Drift

In order to evaluate the importance of modeling heading drift we run a modified version of VIP, where heading angles were excluded from optimization. For TotalCapture heading errors were negligible since they were recalibrated for each recording sequence. This is impractical for continuous operation and unrealistic if headings are not recalibrated all the time. During our recordings of 3DPW we observed heading errors exceeding 100°. Hence to setup a more realistic experiment we artificially distorted the IMU orientations by heading errors drawn from a normal distribution with standard deviations of $\sigma = 20°$ and $\sigma = 75°$. For $\sigma = 20°$ the MPJPE increases to 32.6mm, and for $\sigma = 75°$ the MPJPE jumps to 61.6mm. This underlines the importance of modeling heading under realistic conditions: VIP with heading optimized still achieves a MPJPE of 26mm.

2 Performance vs. Number of IMUs

In TotalCapture 13 IMUs are used for VIP and we evaluate 6 IMUs in VIP-IMU6. Another common setting is to use 9 IMUs attached to lower and upper legs, lower and upper arms and waist. For this sensor setup, VIP achieves a MPJPE of 28.3mm, which is 2.3mm worse compared to 13 IMUs, but might be sufficient if a reduced number of IMUs is desired.

3 State-of-the-art Monocular Pose Estimation Methods

In order to assess the challenges of 3DPW compared to existing indoor datasets such as x and y we ran the methods of [2–4] on 3DPW. All methods estimate 3D pose from either 2D landmarks [2], a single image [3] or both [4]. We ran these methods with the appropriate inputs such as ground truth 2D pose candidate or cropped images since these methods only work for single person. Since there is not existing method or technology that is able to compute ground-truth MoCap data in such unconstrained environments as in 3DPW, we use the results of VIP as a measure of ground-truth. After Procrustes alignment (rotation, scale and translation) the MPJPEs are 203.4mm for [2], 108.1mm for [3] and 108.2mm for [4]. These high errors indicate that the new dataset contains many poses that

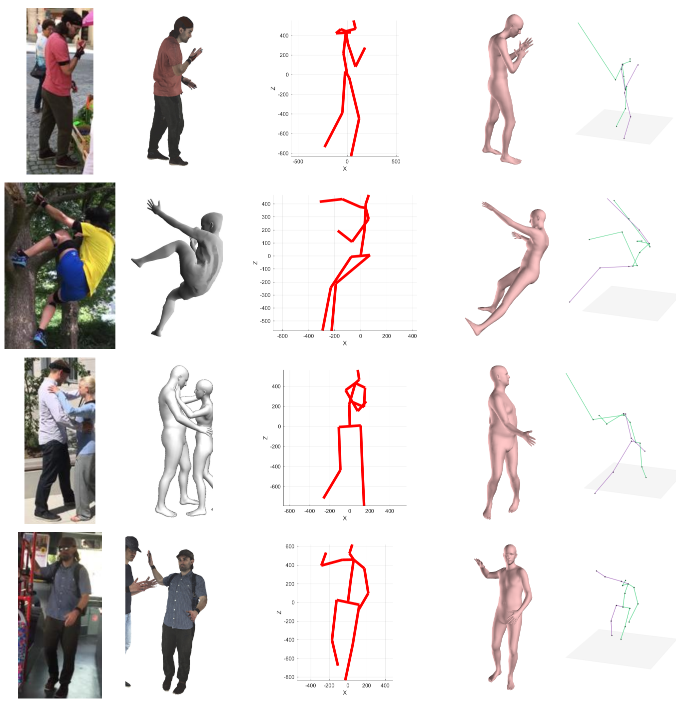

Fig. 1. Example results for images taken from Fig. 9 of the main document. For every example we show from left to right: a cropped image of the person to be estimated, our results, method [3], method [4] and method [2].

are challenging for state of the art methods. Please note, that these methods report an MPJPE of approx. 62-82mm on Human3.6m [5]. This gap in MPJPE indicates the difficulty of the dataset and demonstrates that there are many new challenges still to be addressed. Visual inspection corroborates this, with some representative examples depicted in Fig. 1.

References

1. von Marcard, T., Henschel, R., Black, M.J., Rosenhahn, B., Pons-Moll, G.: Recovering accurate 3d human pose in the wild using imus and a moving camera. European Conference on Computer Vision (ECCV) (2018)
2. Martinez, J., Hossain, R., Romero, J., Little, J.J.: A simple yet effective baseline for 3d human pose estimation. In: IEEE International Conference on Computer Vision (ICCV). (2017)
3. Popa, A.I., Zanfir, M., Sminchisescu, C.: Deep multitask architecture for integrated 2d and 3d human sensing. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)
4. Bogo, F., Kanazawa, A., Lassner, C., Gehler, P., Romero, J., Black, M.J.: Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image. In: European Conference on Computer Vision (ECCV), Springer International Publishing (2016)
5. Ionescu, C., Papava, D., Olaru, V., Sminchisescu, C.: Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) **36**(7) (2014) 1325–1339