Real-time Body Tracking with One Depth Camera and Inertial Sensors
Supplemental Material

Thomas Helten* Meinard Müller† Hans-Peter Seidel* Christian Theobalt*
*Saarland University and MPI Informatik †International Audio Laboratories Erlangen
{thelten,theobalt}@mpi-inf.mpg.de meinard.mueller@audiolabs-erlangen.de

The supplemental material for ICCV the submission "Real-time Body Tracking with One Depth Camera and Inertial Sensors" consists of this document and the accompanying video.

Content

In the video, we show qualitative tracking results especially for sequences D_5 and D_6. We also describe graphically the key components of our proposed tracker. In this document, we include some additional information about the sequences included in our evaluation dataset[1] and present some extended quantitative results that did not fit spatially into our paper.

Table 1 gives an overview over the six sequences of our evaluation dataset. While sequence D_1 contains comparably simple motions such as arm and leg rotations, the other five sequences are challenging for depth based trackers each in its own way. Sequence D_2 introduces considerable faster motions compared to Sequence D_1, including punching and kicking motions. Sequence D_3 contains full body motions including jumping jacks, skiing motions, and squats. Especially the latter ones are interesting, because the induce inertial features that are almost equal over all phases of the motion, which is challenging for the inertial-based retrieval. This stems from the fact that the arms and the lower legs do not change their orientation with respect to the trunk. In Sequence D_4, the arms touch the body at different locations which challenges the geodesic extrema-based database lookup as used by Baak *et al.* [1]. In particular, this prevents correct geodesic extrema detection by introducing loops to the shape of the person. Sequence D_5 first introduces non-frontal poses including walking sidewards and skiing motions performed lateral w.r.t. the depth camera. Finally, Sequence D_6 completes our evaluation dataset by introducing walking in circles, rotating on the spot and occlusions of the arms.

In Fig. 1, we show the error bars for all sequences of the Stanford dataset introduced in [2]. Finally, the full set of joint errors, as described in Sect. 8.2. of the paper are depicted in Fig. 2.

*This work was funded by the ERC Starting Grant "CapReal".
[1] http://resources.mpi-inf.mpg.de/InertialDepthTracker

Table 1. Description of the six sequences from the evaluation dataset.

Scene	Description	#Frames
D_1	Arm rotations, leg rotations, bending of upper body, and grabbing	1366
D_2	Punching, kicking, fast arm motion, and jumping	445
D_3	Jumping jacks, skiing, and squats	527
D_4	Arms at the hips, arms crossed, and hands behind head	930
D_5	Straight walking sidewards, and skiing sidewards	930
D_6	Circular walking, rotation on the spot, and moving arms behind the body	885

References

[1] A. Baak, M. Müller, G. Bharaj, H.-P. Seidel, and C. Theobalt. A data-driven approach for real-time full body pose reconstruction from a depth camera. In *ICCV*, 2011.

[2] V. Ganapathi, C. Plagemann, S. Thrun, and D. Koller. Real time motion capture using a single time-of-flight camera. In *CVPR*, 2010.

[3] J. Taylor, J. Shotton, T. Sharp, and A. W. Fitzgibbon. The Vitruvian manifold: Inferring dense correspondences for one-shot human pose estimation. In *CVPR*, 2012.

[4] M. Ye, X. Wang, R. Yang, L. Ren, and M. Pollefeys. Accurate 3d pose estimation from a single depth image. In *ICCV*, pages 731–738, 2011.

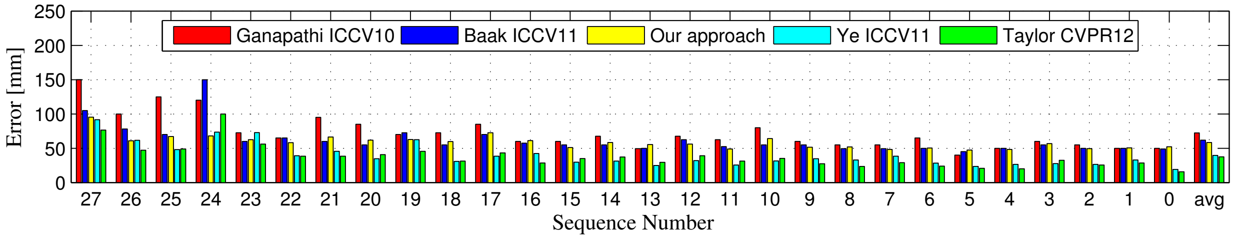

Figure 1. Evaluation on the Stanford dataset presented in [2]. **(red)** Ganapathi *et al.* [2] **(blue)** Baak *et al.* [1] **(yellow)** Our tracker. **(cyan)** Ye *et al.* [4] (not real-time). **(green)** Taylor *et al.* [3].

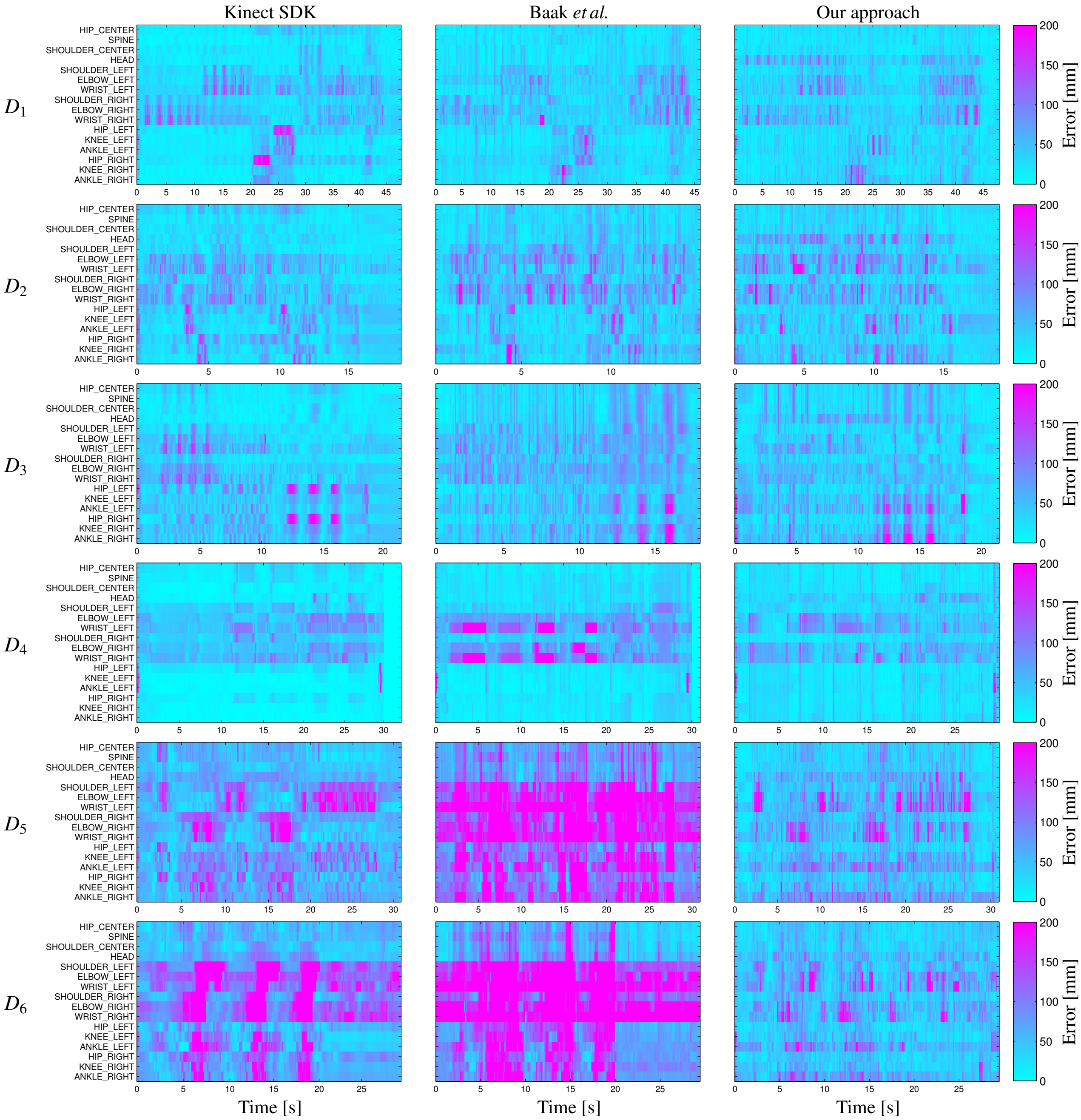

Figure 2. Joint Errors for all joints and all sequences $D_1, \ldots, D_6$ from the evaluation dataset, according to Sect. 8.2. in the paper. The three columns represent the three different trackers: **(left)** The Kinect SDK's joint tracker, **(middle)** the approach presented by Baak *et al.*[1], and **(right)** our approach.