

# Leveraging Anthropometric Measurements to Improve Human Mesh Estimation and Ensure Consistent Body Shapes

## Supplementary Material

### 7. Anthropometric Measurements

The selection of the anthropometric measurements is mainly adopted from AnthroNet [28]. In total, 36 measurements are selected, which can be divided into 23 lengths and 13 circumferences. All measurements are taken based on the standard SMPL-X T-pose. The reference landmarks are chosen by matching the vertices on the default mesh with the landmarks defined by the anthropometric survey of the U.S. army personnel [14]. A visualization of the landmarks can be found in Figure 4 and 5. The lengths are calculated by computing the Euclidean distance between two landmarks or the difference along the coordinate axis pointing upwards for certain heights. The lengths are visualized in Figure 6 and 7. Table 9 lists the enclosing landmarks for each length. To measure the circumferences, we adopt the code from [2]. For each measurement, a plane is created, the intersection between the mesh and the plane are extracted and the convex hull of the result is calculated. During this process, the mesh is restricted to the body part to be measured. A visualization of the circumferences can be found in Figure 8 and a list of the landmarks and the normal vectors spanning the plane in Table 6.

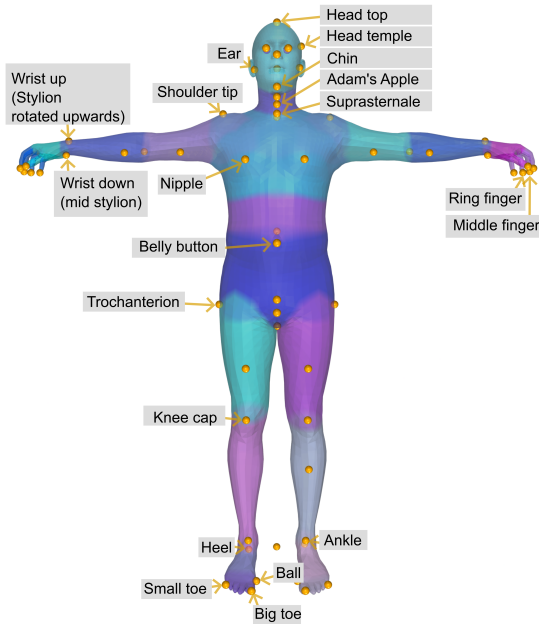


Figure 4. Visualization of the used landmarks with a standard T-pose SMPL-X mesh in front view.

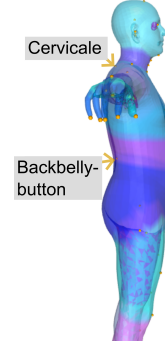


Figure 5. Visualization of a subset of the used landmarks in side view.

Idx	Circumference	Normal Vector	Position
1	Waist	Up	Belly button
2	Chest	Up	Nipple
3	Hip	Up	Pubic bone
4	Head	Up	Head temple
5	Neck	Spine to head	Adam's apple
6/7	Upper Arm	Shoulder to elbow	Center of the bicep
8/9	Forearm	Elbow to wrist	Widest point of the forearm
10/11	Thigh	Up	Center of the thigh
12/13	Calf	Up	Widest point of the calf

Table 6. Definitions of circumferences by landmarks and the normal vector spanning the plane.

### 8. 3D Human Shape Ground Truth Analysis

We further analyze the GT shape consistency for the common datasets Human3.6M [16] and MPI-INF-3DHP [22]. We find that for Human3.6M, the bone lengths derived from the 3D annotations are fixed, but not for MPI-INF-3DHP.

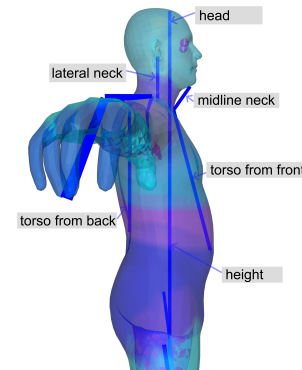


Figure 6. Visualization of used lengths with a standard T-pose SMPL-X mesh in side view.

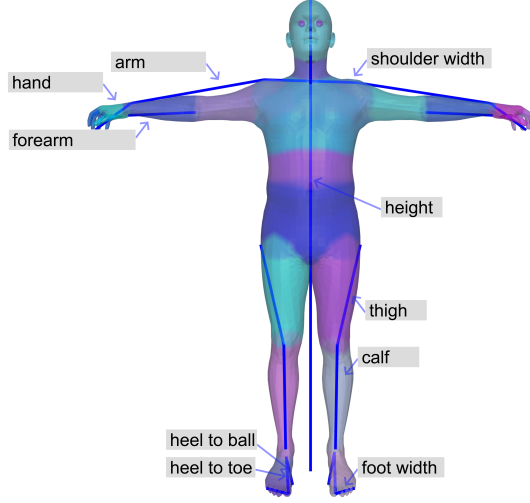


Figure 7. Visualization of used lengths with a standard T-pose SMPL-X mesh in front view.

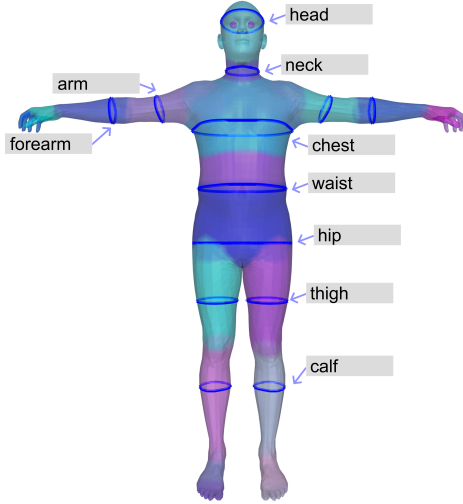


Figure 8. Visualization of used circumferences with a standard T-pose SMPL-X mesh in front view.

Therefore, we do not report the deviations of 3D joint annotations for Human3.6M, since there are none. We further evaluate the SMPL-X annotations for both datasets provided by NeuralAnnot [24] which are used by HME models as GT for training. See Tables 7, 8 for details.

## 9. Evaluating A2B Models

We measure two types of errors to evaluate the performance of our A2B models. The first type ( $\beta$  error) shows the error if we take the GT  $\beta$  parameters, derive anthropometric measurements (B2A), input them into the A2B models and evaluate the MSE of the predicted  $\beta$  parameters. The second type ( $A$  error) calculates B2A from the predicted  $\beta$

3D joint annotations				SMPL-X annotations			
Measure	$\sigma$	r. $\sigma$	r. range	Measure	$\sigma$	r. $\sigma$	r. range
head	0.19	1.03%	2.08%	head	0.21	0.75%	4.87%
hip width	0.22	0.89%	1.80%	hip circ.	1.16	1.16%	9.13 %
forearm	0.21	0.87%	1.77%	forearm	0.45	1.80%	9.75%
upper arm	0.29	0.90%	1.82%	arm	0.83	1.59%	8.19%
lower leg	0.60	1.49%	3.06%	lower leg	1.05	2.56%	11.54%
thigh	3.83	7.91 %	41.90%	thigh	0.77	2.02%	9.47%
				height	2.76	1.56%	8.24%
				$\beta$ param.	0.18		

Table 7. GT data analysis for MPI-INF-3DHP [22]. Bone length analysis based on the 3D joint locations (left) and on SMPL-X annotations by NeuralAnnot (right). Standard deviation  $\sigma$ , relative standard deviation  $\frac{\sigma}{avg}$  and relative range  $\frac{\max - \min}{avg}$  of anthropometric measurements are reported. Standard deviations are given in cm, despite for the  $\beta$  parameters. The values are averaged between left and right body parts and between all persons in of each dataset. The  $\beta$  parameter standard deviation is averaged over all  $\beta$  parameters.

SMPL-X annotations			
Measure	$\sigma$	r. $\sigma$	r. range
head	0.41	1.51%	10.28%
hip circ.	1.24	1.19%	8.90%
forearm	0.83	3.30%	27.93%
arm	0.77	2.58%	22.88%
lower leg	0.43	1.18%	12.20%
thigh	0.66	1.27%	9.43%
height	3.40	2.06%	15.66%
$\beta$ param.	0.20		

Table 8. GT data analysis for Human3.6M [16]: Analysis of SMPL-X annotations by NeuralAnnot. Standard deviation  $\sigma$ , relative standard deviation  $\frac{\sigma}{avg}$  and relative range  $\frac{\max - \min}{avg}$  of anthropometric measurements are reported. Standard deviations are given in cm, despite for the  $\beta$  parameters. The values are averaged between left and right body parts and between all persons in of each dataset. The  $\beta$  parameter standard deviation is averaged over all  $\beta$  parameters.

parameters and evaluates the mean difference between the GT and predicted anthropometric measurements (all 36) in mm. These evaluations are a kind of cycle consistency evaluation for A2B and B2A. Figure 9 provides a visualization of the evaluation scheme. The part that is also included in the training is highlighted with thicker arrows. The anthropometric error is only used during evaluation.

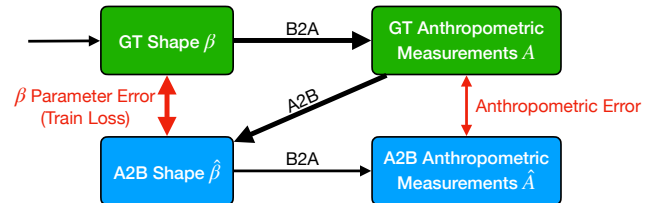


Figure 9. Visualization of the A2B evaluation and training procedures. The training part is highlighted with thicker arrows. During training, the  $\beta$  parameter error is used. For evaluations, the  $\beta$  parameter error and the anthropometric error are calculated.

Idx	Length Name	From	To
1	Shoulder width	Left shoulder tip (left acromion)	Right shoulder tip
2	Back torso height	Cervicale	Back belly button
3	Front torso height	Suprasternale (top of the breastbone)	Belly button
4	Head	Head top	Cervicale
5	Midline neck	Chin	Suprasternale
6	Lateral neck	Center between the ears	Cervicale
7	Height	Head top	Center between heels
8/9	Hand right/left	Center between middle and ring finger	Stylion rotated downwards
10/11	Arm right/left	Acromion	Wrist
12/13	Forearm right/left	Elbow	Stylion rotated downwards
14/15	Thigh right/left	Outer point at the femur (Trochanterion)	Knee cap
16/17	Calf right/left	Knee cap	Ankle
18/19	Foot width right/left	Small toe	Big toe
20/21	Heel to ball right/left	Heel	Ball
22/23	Heel to toe right/left	Heel	Big toe

Table 9. Definitions of lengths by their two enclosing landmarks.

In the main paper, we test our A2B models on the AGORA [26] dataset and randomly sampled body shapes. Since AGORA is a synthetic dataset, it might not reflect the real world. The same holds for randomly sampled body shapes. Therefore, we additionally test our best A2B models on the real-world SSP-3D dataset [33] which consists of diverse body shapes. We display the results in Table 10.

	$\varnothing$ error of $\beta$ [ $10^{-2}$ ]			$\varnothing$ error of $A$ [mm]		
	m	f	n	m	f	n
NN	1.73	0.97	2.74	0.634	0.803	0.968
SVR	0.13	0.0039	0.066	0.167	0.114	0.182

Table 10. Results of our A2B models on the SSP-3D dataset using n(eutral), m(ale) and f(emale) meshes.

All A2B models accurately estimate the diverse real-world body shapes with low error.

## 10. Keypoint Selection for fit3D

We use the fit3D [12] dataset for our evaluations, since this is the only sports dataset with public SMPL-X annotations. We evaluate on the SMPL-X joints, since these are trivial to obtain from SMPL-X meshes and there is no regressor available for the fit3D annotated 3D joints. SMPL-X has 144 defined joints. Since our focus is mainly on the body and not on the hands and face, we remove most of these joints. In the end, we select a subset of 37 SMPL-X joints: pelvis, left hip, right hip, spine1, left knee, right knee, spine2, left ankle, right ankle, spine3, left foot, right foot, neck, left collar, right collar, head, left shoulder, right shoulder, left elbow, right elbow, left wrist, right wrist, left index, left thumb, right index, right thumb, left big toe, left small toe, left heel, right big toe, right small toe, right heel, right eye, left eye, right ear, left ear, nose.

## 11. Generation of Pseudo GT Anthropometric Measurements

As we do not have access to the athletes of ASPset and fit3d to obtain real anthropometric measurements, we need an alternative to simulate this process. For ASPset, as a first step, we run IK on the GT 3D joint locations. From the generated meshes, we obtain the necessary anthropometric parameters with B2A. Then, we use the median values of these measurements as the GT anthropometric values. We call these parameters *pseudo GT* throughout this paper, since this is not directly the GT, but obtained from IK executed on the GT 3D joint locations and the B2A computation from the created meshes. These parameters are used in this paper to generate the pseudo GT  $\beta$  parameters by A2B prediction.

We do not have access to the athletes of the fit3D dataset either. Therefore, we need some kind of GT data to mimic measurements. Obviously, there is no GT available for the official test set evaluation on the evaluation server. We therefore split the official training dataset into a training, validation, and test set for our evaluations. We perform a leave-one-out cross validation, therefore all eight athletes from the official training dataset are used in our evaluation. With this selection, we have real GT shape parameters available. We do not use these directly, since this would skip the measuring process that is needed in real applications. Further, the GT data is not consistent (see Section 3 in the main paper). Therefore, we apply B2A and use the median measurements over time in order to simulate the measuring process and obtain a single set of anthropometric measurements per person. In real applications, this step is omitted because the anthropometric parameters can be measured directly from the athletes before starting the recording.

We consider this strategy as a valid method for evaluations, since our main goal is to improve the HME performance as much as possible with only marginal overhead.

pose	orig.	measure	NN m	SVR m	NN n	SVR n	median
SMPLer-X	86.02	SMPLer-X	85.89	85.69	86.02	85.99	86.04
SMPLer-X FT	79.09	SMPLer-X FT	78.92	78.88	79.59	79.37	79.44
SMPLer-X FT	-	GT	65.63	65.84	64.77	<b>64.76</b>	-
SMPLer-X FT	-	SMPLer-X	73.41	73.29	73.79	73.63	73.66
IK-UU	67.54	IK-UU	66.92	66.60	67.28	67.12	67.16
IK-UU	-	SMPLer-X	63.80	<u>63.64</u>	63.92	63.78	63.82
IK-UU	-	SMPLer-X FT	69.46	69.27	69.83	69.63	69.69
IK-UU	-	GT	56.44	56.56	<b>55.18</b>	55.19	-

Table 11. MPJPE results in mm for the test split of ASPset. Results are given for different methods and replaced *beta* parameters with A2B results (columns NN/SVR) or the median of the original  $\beta$  parameters from the model noted in the *measure* column. SMPLer-X FT stands for the best fine-tuned variant of SMPLer-X (fine-tuned with the meshes obtained from IK executed on the GT 3D joints). The *orig* column contains the results without replaced  $\beta$  parameters. We highlight the best result for each model and the best option for the combination of IK-UU pose and SMPLer-X  $\beta$  parameters, since this combination outperforms the original IK-UU result, too.

Our main focus is sports, which contains extreme poses that let existing HME models fail, sometimes even to detect a human at all. Examples can be found in the supplementary videos. As professional athletes are measured anyway, the additional effort for the measurements is negligible in this context.

## 12. Inverse Kinematics

We use the inverse kinematics approach with a VPoser extension, as proposed in the code by [27], to fit SMPLX meshes to given 3D keypoints. VPoser is a learned prior for human poses, since the raw SMPL-X model definition allows impossible poses for humans. VPoser learned plausible poses from the large AMASS [21] dataset and helps IK to generate only plausible poses. IK learns the best SMPL-X parameters ( $\beta$  and  $\theta$ ) that fit the mesh to the given 3D joint locations by minimizing the error between the given joint locations and the regressed joint locations from the mesh. IK is an iterative algorithm and adjusts the pose and the shape parameters with a gradient descent minimization approach in each step. Besides the already described joint error, IK further penalizes abnormal poses with a VPoser error and extreme body shapes with a  $\beta$  parameter error. Therefore, the total loss for IK can be described as:

$$\mathcal{L}_{IK} = \lambda_1 \mathcal{L}_{joint} + \lambda_2 \mathcal{L}_{VPoser} + \lambda_3 \mathcal{L}_{\beta}, \quad (1)$$

whereby  $\mathcal{L}_{joint}$  is the summarized Squared Error of the estimated keypoints,  $\mathcal{L}_{VPoser}$  and  $\mathcal{L}_{\beta}$  are the sums of the squared values of the VPoser and  $\beta$  parameters, respectively. This makes sense since the VPoser and  $\beta$  parameter distributions are centered around zero. We set the weighting factors  $\lambda_1 = 10$ ,  $\lambda_2 = 0.0007$ , and  $\lambda_3 = 0.01$  in our experiments. We use relatively low values for  $\lambda_2$  and  $\lambda_3$ , since sports datasets incorporate extreme poses and our main interest is to achieve the most perfect pose.

We execute IK per frame, which results in a slight jitter in between the frames, but leads to more accurate joint positions. Since IK needs multiple iterations to adjust the standard T-pose parameters to achieve a pose that is roughly close to the desired UU pose, we speed up the process by initializing the pose and shape parameters with the result from the previous frame if available. This also enhances the final result slightly. We acknowledge that IK is relatively slow regarding the runtime, but our main focus is the precision. For sport analysis, which is our focus, the runtime is not critical, but a very precise result is crucial.

## 13. Fine-tuning HME Models with Pseudo GT Meshes

Fine-tuning existing HME models on pure 3D joints datasets is not possible, since they need mesh annotations for training. However, with IK, we can generate pseudo GT meshes. We exemplarily test a fine-tuning of SMPLer-X on ASPset with this approach. Experiments show that using their fine-tuning script with 1.6M iterations leads to worse results than the results without fine-tuning. Therefore, we reduce the number of iterations with early stopping and achieve better results with fine-tuning only for 32K iterations.

The results shown in Table 11 prove that fine-tuning on IK generated meshes can lead to a significant improvement of the scores. Replacing the  $\beta$  parameters of the fine-tuned results with the A2B  $\beta$  parameters boosts the performance even more. These are the best results achieved with any existing HME model throughout this study.

Moreover, we experiment with using the SMPLer-X body shape parameters combined with the poses estimated by IK applied to the UU results (see last two rows of Table 11). Using the  $\beta$  parameters from SMPLer-X leads to a slightly better result than the original 3D joint based result



(without IK). This evaluation shows that 3D HPE models are better in precisely locating the joints of humans than HME models, but HME models are better in estimating the shape of humans. We also try to use the  $\beta$  parameters of the fine-tuned variant together with the UU IK poses like before. However, this resulted in a performance drop compared to the body shape parameters from the original SMPLer-X without fine-tuning. These experiments show that fine-tuning HME models on pseudo ground truth leads to a better performance regarding the keypoints, but the estimated  $\beta$  parameters have worse quality. This can further be proven by replacing the  $\beta$  parameters from the fine-tuned SMPLer-X variant with the  $\beta$  parameters from the not fine-tuned model, which results in a performance gain of over 5 mm compared to the original results from the fine-tuned version (rows 2 and 4 in Tab. 11). However, our method using the UU IK poses and the A2B body shape parameters with GT anthropometric measurements achieves the overall best results.

We provide a comprehensive summary and visualization of all results on the ASPset dataset in Section 14. This includes results of existing HME models, results of our approach, and the fine-tuning results.

## 14. Summary of the Results

We execute a multitude of experiments with different combinations of pose and shape parameters. Figure 10 summarizes the results with their pose and shape origins for ASPset. In general, the poses estimated by IK based on the UU results (red branch in Fig. 10) are more precise than the poses estimated by SMPLer-X (light blue branch in Fig. 10). Further, the body shape parameters from our A2B models with GT anthropometric measurements (green boxes in Fig. 10) achieve the best results for all poses. We provide more qualitative examples comparing SMPLer-X with this approach in the supplementary video. Without access to the GT, all models benefit slightly from A2B model results with the median anthropometric measurements from B2A of the estimated meshes by the respective model (boxes with same color for all three branches in Fig. 10). Moreover, SMPLer-X A2B body shape parameters perform best when analyzing body shapes without GT access (light blue boxes in Fig. 10). Fine-tuning SMPLer-X with IK created meshes (dark blue branch in Fig. 10) improves the performance of SMPLer-X, although the quality of the body shape deteriorates. This can be seen as by comparing the shapes from SMPLer-X and fine-tuned SMPLer-X (dark blue and light blue boxes in Fig. 10) with fine-tuned and IK poses.

Since fit3D is a larger dataset, fine-tuning UU works better, which further leads to better IK meshes with an MPJPE of 37.02 mm. Enforcing consistent meshes with GT or IK A2B shape parameters decreases the performance

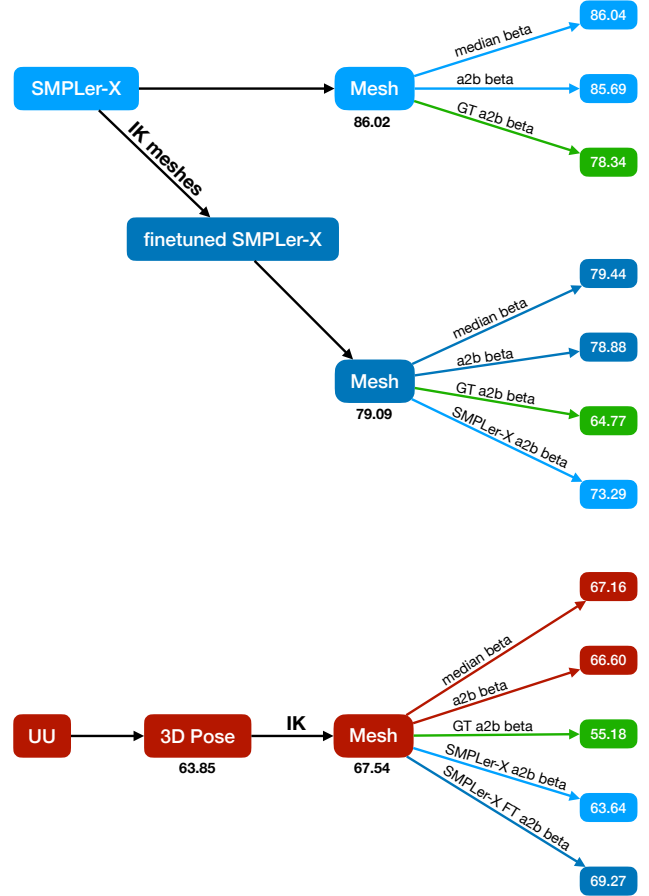


Figure 10. Overview of the main results for the ASPset dataset. All results are MPJPE results in mm. Results below *mesh* boxes show the result with the original  $\beta$  parameters. All results after arrows to the right are results with replaced  $\beta$  parameters. The type of the  $\beta$  parameters is noted on the arrow and is color-coded.

slightly in this case. However, A2B shape parameters achieve slightly better scores than median values. This also holds for OSX and Multi-HMR. Overall, the approach with UU, IK, and A2B body shape parameters achieves an over 33 mm lower MPJPE than any HME model. The same also holds for the MVE, which can be improved by over 30 mm with our approach. The scores can be found in the main paper.

We provide two videos in the supplementary material that show qualitative results for ASPset and fit3D. Figure 11 shows one example visualization for both datasets. We include the GT and predicted meshes in the fit3D visualization and display the GT and estimated body shapes in T-pose right next to each other. For ASPset, we visualize the estimated meshes and the GT and estimated joints, since we do not have GT meshes here.

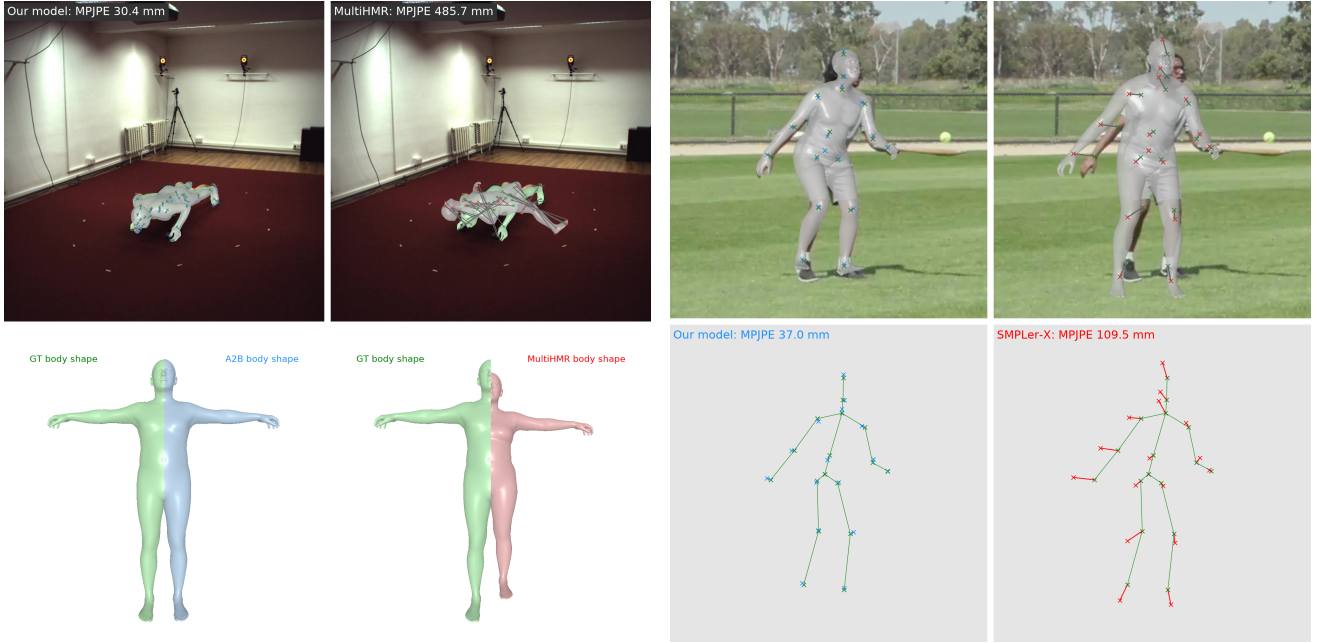


Figure 11. Example frames from our supplementary videos. It shows qualitative results of our approach compared to MultiHMR for fit3d (left) and qualitative results of SMPLer-X and our approach for example frames from ASPset (right). In fit3d visualizations, we display the **GT meshes in green** and the estimated meshes in gray. The GT joints are also displayed in green while the estimated joints from our model are visualized in **blue**. The **MultiHMR joints are shown in red**. Corresponding joints are connected. We display the exact MPJPE values in the top left of each frame. Recall that the visualization is in 2D, but the evaluation is in 3D. Therefore, sometimes the MPJPE values may seem odd. In the lower part, we show the estimated body shapes in T-pose. The **GT body shape is shown in green** and the estimated body shape from **our model in blue**. The **MultiHMR body shape is shown in red**. For ASPset visualizations, we display the estimated meshes and the GT and estimated joints. **GT joints are shown in green**, estimated joints from **our model in blue**, and the **SMPLer-X joints in red**. Corresponding joints are connected. In the lower part, we show the GT and estimated joints in the same way, but without the mesh and image to reduce distraction. We further display the MPJPE values.