Supplementary File of Paper "HybrIK: A Hybrid Analytical-Neural Inverse Kinematics Solution for 3D Human Pose and Shape Estimation"

A. Rigid Registration of Global Rotation

In the SMPL model [3], the pose parameters θ control the rotations of the rigid body parts. The three joints named `spine`, `left hip` and `right hip` form a rigid body part, which is controlled by the global root rotation. Therefore, the global rotation can be determined by registering the rest pose template of `spine`, `left hip` and `right hip` to the predicted locations of these three joints. Let t_1, t_2 and t_3 denote their locations in the rest pose template, and p_1, p_2 and p_3 denote the predicted locations. Our goal is to find a rigid rotation that optimally aligns the two sets of joints. Here, we assume the root joint of the predicted pose and the rest pose are aligned. Hence, the problem is formulated as:

$$R_0 = \arg\min_{R\in\mathbb{SO}^3} \sum_{i=1}^{3} \|p_i - Rt_i\|_2^2. \tag{1}$$

This formula can be written in matrix form:

$$R_0 = \arg\min_{R\in\mathbb{SO}^3} \|P_0 - RT_0\|_F^2, \tag{2}$$

where $\|\cdot\|_F$ denotes the Frobenius norm,P_0 denotes $[p_0\ p_1\ p_2]$, and T_0 denotes $[t_0\ t_1\ t_2]$. Let us simplify the expression in Eq. 2 as:

$$\begin{aligned} &\min_{R\in\mathbb{SO}^3} \|P_0 - RT_0\|_F^2 \\ \Leftrightarrow &\min_{R\in\mathbb{SO}^3} \text{trace}((P_0 - RT_0)^{\mathrm{T}}(P_0 - RT_0)) \\ \Leftrightarrow &\min_{R\in\mathbb{SO}^3} \text{trace}(P_0^{\mathrm{T}}P_0 + T_0^{\mathrm{T}}T_0 - 2P_0^{\mathrm{T}}RT_0). \end{aligned} \tag{3}$$

Note that $P_0^{\mathrm{T}}P_0$ and $T_0^{\mathrm{T}}T_0$ are independent of R. Thus the original problem is equivalent to:

$$\begin{aligned} &\arg\min_{R\in\mathbb{SO}^3} \|P_0 - RT_0\|_F^2 \\ \Leftrightarrow &\arg\max_{R\in\mathbb{SO}^3} \text{trace}(P_0^{\mathrm{T}}RT_0). \end{aligned} \tag{4}$$

Further, we can leverage the property of the matrix trace,

$$\text{trace}(P_0^{\mathrm{T}}RT_0) = \text{trace}(RT_0P_0^{\mathrm{T}}). \tag{5}$$

Then, we apply Singular Value Decomposition (SVD) to the joint locations:

$$T_0P_0^{\mathrm{T}} = U\Lambda V^{\mathrm{T}}. \tag{6}$$

The problem is equivalent to:

$$\begin{aligned} &\arg\max_{R\in\mathbb{SO}^3} \text{trace}(RT_0P_0^{\mathrm{T}}) \\ \Leftrightarrow &\arg\max_{R\in\mathbb{SO}^3} \text{trace}(RU\Lambda V^{\mathrm{T}}) \\ \Leftrightarrow &\arg\max_{R\in\mathbb{SO}^3} \text{trace}(\Lambda V^{\mathrm{T}}RU). \end{aligned} \tag{7}$$

Note that U, V and R are orthogonal matrices, so $M = V^{\mathrm{T}}RU$ is also an orthogonal matrix. Then, for all $1 \leq j \leq$ we have:

$$\begin{aligned} m_j^{\mathrm{T}}m_j = 1 = \sum_{i=1}^{3} m_{ij}^2 \\ \Rightarrow\ m_{ij}^2 \leq 1 \Rightarrow |m_{ij}| \leq 1. \end{aligned} \tag{8}$$

Besides, Λ is a diagonal matrix with non-negative values, *i.e.* λ_1, λ_2, $\lambda_3 \geq 0$. Therefore:

$$\begin{aligned} \text{trace}(\Lambda V^{\mathrm{T}}RU) &= \text{trace}(\Lambda M) \\ &= \sum_{i=1}^{3} \lambda_i m_{ii} \leq \sum_{i=1}^{3} \lambda_i. \end{aligned} \tag{9}$$

The trace is maximized if $m_{ii} = 1, \forall 1 \leq i \leq 3$. That means $M = \mathcal{I}$, where $\mathcal{I}$ is the identity matrix. Finally, the optimal rotation R_0 is:

$$\begin{aligned} V^{\mathrm{T}}R_0U = \mathcal{I} \\ \Rightarrow R_0 = VU^{\mathrm{T}}. \end{aligned} \tag{10}$$

B. More Ablation Experiments

	GT β		Estimated β		Zero β	
	MPJPE	PVE	MPJPE	PVE	MPJPE	PVE
Error	72.7	87.4	80.0	94.5	81.1	95.4

Table 1. **Reconstruction error with different shape parameters β.**

Effect of β In this experiment, we analyze the effect of the shape parameters β in Tab. 1. Using the ground-truth β brings 5 mm improvement of MPJPE and PVE on 3DPW

	Human3.6M		3DPW	
	Predicted Pose	HybrIK	Predicted Pose	HybrIK
MPJPE (24 *jts*) ↓	51.3	48.1	88.2	79.2

Table 2. **Error correction capability of HybrIK** on 3DPW and Human3.6M.

dataset. Using zero β brings 1 mm error. It shows that there are lots of room for improvement by estimating more accurate β.

Comparison with Baseline Models In this experiment, we compare HybrIK with two baselines to validate its effectiveness. Firstly, we want th compare with the model that directly predicts SMPL parameters without any auxiliary loss. This model is a degraded version of HMR [2]. We find it is hard to train and the model learns limited information. The model achieves over 100 mm error on Human3.6M [1]. Secondly, we add 3D keypoint prediction to help the network to extract features. The model still learns to predict SMPL parameters directly. However, still over 100 mm error achieves on Human3.6M [1] dataset.

Error correction capability of HybrIK In this experiment, we examine the error correction capability of HybrIK on 3DPW [4] and Human3.6M [1] datasets. Quantitative results are reported in Tab. 2.

C. Qualitative Results

Fig. 1 provides qualitative results of our approach from the different datasets involved in our experiments (LSP, MPI-INF-3DHP, Human3.6M, 3DPW). Fig. 2 includes typical failure cases that are attributed to erroneous bone length estimation (shape parameters β) and 3D keypoint estimation, which lead to misalignment and unnatural joint bending, respectively.

References

[1] 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. 2

[2] Angjoo Kanazawa, Michael J Black, David W Jacobs, and Jitendra Malik. End-to-end recovery of human shape and pose. In *CVPR*, 2018. 2

[3] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J Black. Smpl: A skinned multi-person linear model. *TOG*. 1

[4] Timo von Marcard, Roberto Henschel, Michael Black, Bodo Rosenhahn, and Gerard Pons-Moll. Recovering accurate 3d human pose in the wild using imus and a moving camera. In *ECCV*, 2018. 2

Figure 1. Qualitative results from various datasets, LSP (rows 1-3), MPI-INF-3DHP (row 4), 3DPW (rows 5-6), H36M (rows 7-8).

Figure 2. Erroneous reconstructions of our method. Typical failure cases can be attributed to inaccurate bone length estimation (shape parameters β) and 3D keypoint estimation.