Dense Depth Alignment for Human Pose and Shape Estimation - Supplementary Material

December 19, 2023

1 Optimal Alignment

We are interested in finding the scale $s \in \mathrm{R}$ and translation $\boldsymbol{t} \in \mathrm{R}^3$ that optimally align the point cloud $\boldsymbol{P}_1 \in \mathrm{R}^{N\times 3}$ to another point cloud $\boldsymbol{P}_2 \in \mathrm{R}^{N\times 3}$:

$$s_{opt}, \boldsymbol{t}_{opt} = \underset{s,\boldsymbol{t}}{\operatorname{argmin}} ||s\boldsymbol{P}_1 + \boldsymbol{t} - \boldsymbol{P}_2||^2. \tag{1}$$

We can rewrite the Equation 1 as a summation over $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ (note that points in $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ are paired):

$$s_{opt}, \boldsymbol{t}_{opt} = \underset{s,\boldsymbol{t}}{\operatorname{argmin}} \sum_{i=1}^{N} ||sP_{1i} + \boldsymbol{t} - P_{2i}||^2. \tag{2}$$

We first take the derivative with respect to $\boldsymbol{t}$ first to find the optimal t in terms of $s, \boldsymbol{P}_1, \boldsymbol{P}_2$.

$$\frac{\partial \sum_{i=1}^{N} ||sP_{1i} + \boldsymbol{t} - P_{2i}||^2}{\partial \boldsymbol{t}} = 2\sum_{i=1}^{N}(\boldsymbol{t} + sP_{1i} - P_{2i}) = 0$$

and find

$$\boldsymbol{t}_{opt} = \bar{\boldsymbol{P}}_2 - s\bar{\boldsymbol{P}}_1 \tag{3}$$

where $\bar{\boldsymbol{P}}_1$ and $\bar{\boldsymbol{P}}_2$ are the means over the points in the respective point cloud. Substituting this into Equation 1 and taking the derivative with respect to s this time, we get

$$\frac{\partial ||s(\boldsymbol{P}_1 - \bar{\boldsymbol{P}}_1) - (\boldsymbol{P}_2 - \bar{\boldsymbol{P}}_2)||^2}{\partial s} = 2s(\boldsymbol{P}_1 - \bar{\boldsymbol{P}}_1)\odot(\boldsymbol{P}_1 - \bar{\boldsymbol{P}}_1) + 2(\boldsymbol{P}_1 - \bar{\boldsymbol{P}}_1)\odot(\boldsymbol{P}_2 - \bar{\boldsymbol{P}}_2) = 0 \tag{4}$$

where $\odot$ is the Hadamard product. Resolving Equation 4 gives

$$s_{opt} = \frac{(\boldsymbol{P}_1 - \bar{\boldsymbol{P}}_1)\odot(\boldsymbol{P}_2 - \bar{\boldsymbol{P}}_2)}{(\boldsymbol{P}_1 - \bar{\boldsymbol{P}}_1)\odot(\boldsymbol{P}_1 - \bar{\boldsymbol{P}}_1)} \tag{5}$$

and finally

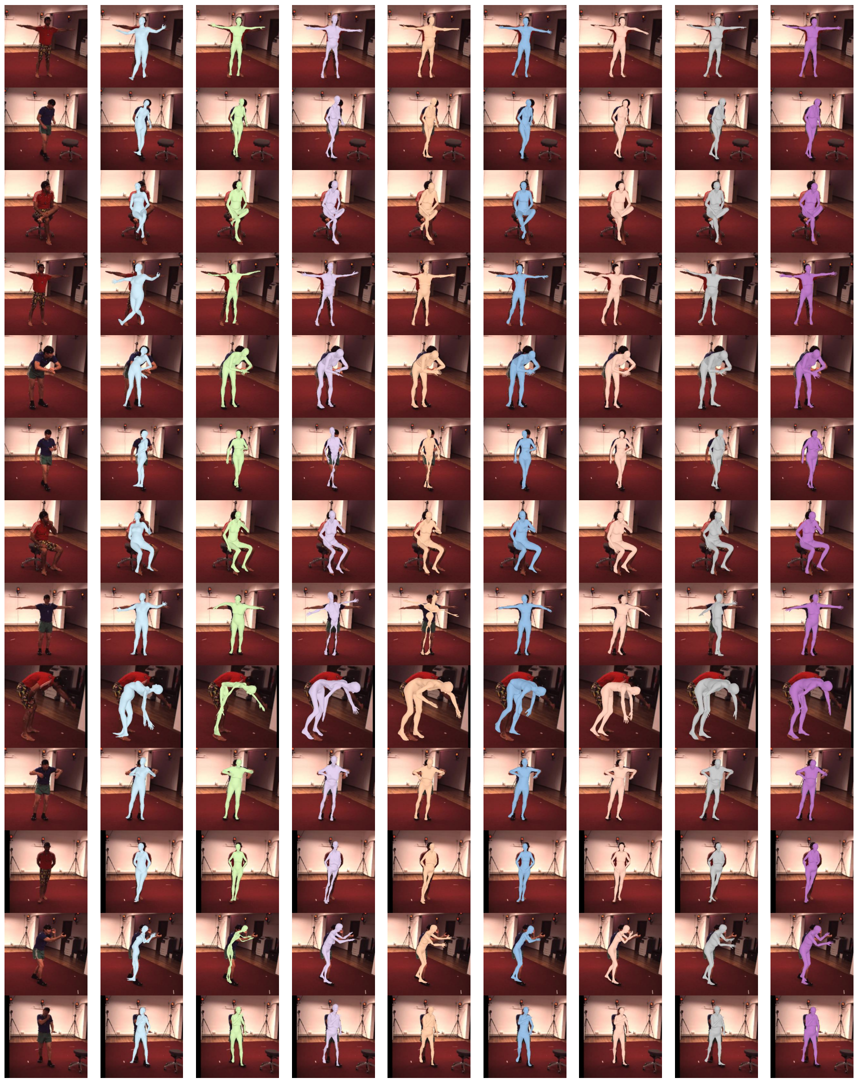
$$\boldsymbol{t}_{opt} = \bar{\boldsymbol{P}}_1 - \frac{(\boldsymbol{P}_1 - \bar{\boldsymbol{P}}_1) \odot (\boldsymbol{P}_2 - \bar{\boldsymbol{P}}_2)}{(\boldsymbol{P}_1 - \bar{\boldsymbol{P}}_1) \odot (\boldsymbol{P}_1 - \bar{\boldsymbol{P}}_1)} \bar{\boldsymbol{P}}_2 \tag{6}$$

In practice we simplify the procedure by using zero mean versions $\boldsymbol{P}'_1 = \boldsymbol{P}_1 - \bar{\boldsymbol{P}}_1$ and $\boldsymbol{P}'_2 = \boldsymbol{P}_2 - \bar{\boldsymbol{P}}_2$ in place of $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ respectively and calculate the optimal alignment as

$$\begin{aligned} s_{opt} &= \frac{\boldsymbol{P}'_1 \odot \boldsymbol{P}'_2}{\boldsymbol{P}'_1 \odot \boldsymbol{P}'_1} \\ \boldsymbol{t}_{opt} &= \boldsymbol{0} \end{aligned} \tag{7}$$

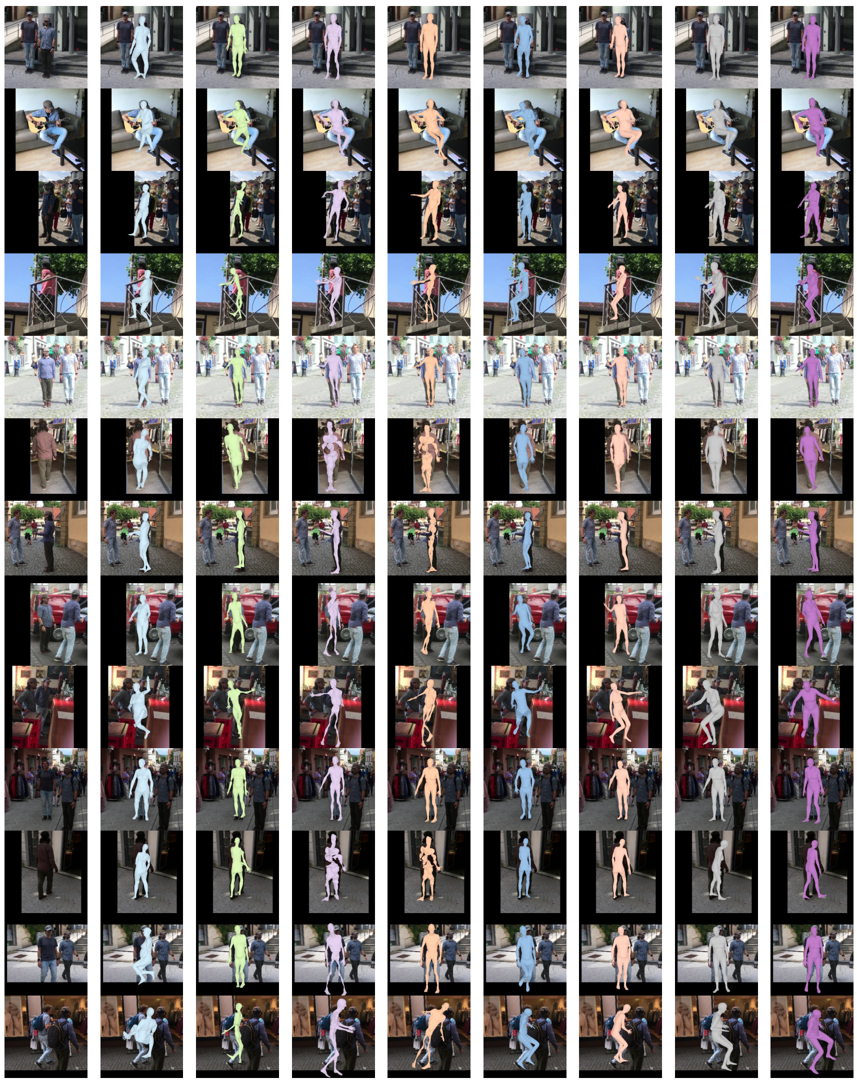
2 More Qualitative Results on H36M and 3DPW

We present the predictions of all models in Table 1 in our paper on random examples from H36M and 3DPW test sets in Figure 1 and Figure 2.



(a) Input	(b) Baseline HMR	(c) $\mathcal{L}_{\text{Depth}}$	(d) $\mathcal{L}_{3D}$	(e) $\mathcal{L}_{3D}$, $\mathcal{L}_{\text{Depth}}$	(f) Baseline HMR*	(g) $\mathcal{L}_{\text{Depth}}$*	(h) $\mathcal{L}_{3D}$*	(i) $\mathcal{L}_{3D}$, $\mathcal{L}_{\text{Depth}}$*

Figure 1: **Random Results from H36M Test Dataset.** All models are trained with $\mathcal{L}_{2D}$ in addition to the stated losses. Models with * are initialized with Camera Pretraining.



(a) Input | (b) Baseline HMR | (c) $\mathcal{L}_{\text{Depth}}$ | (d) $\mathcal{L}_{3D}$ | (e) $\mathcal{L}_{3D}$, $\mathcal{L}_{\text{Depth}}$ | (f) Baseline HMR* | (g) $\mathcal{L}_{\text{Depth}}$* | (h) $\mathcal{L}_{3D}$* | (i) $\mathcal{L}_{3D}$, $\mathcal{L}_{\text{Depth}}$*

Figure 2: **Random Results from 3DPW Test Dataset.** All models are trained with $\mathcal{L}_{2D}$ in addition to the stated losses. Models with * are initialized with Camera Pretraining.