Three-dimensional Reconstruction of Human Interactions

Supplementary Material

Mihai Fieraru[1] Mihai Zanfir[1] Elisabeta Oneata[1]
Alin-Ionut Popa[1] Vlad Olaru[1] Cristian Sminchisescu[2,1]

[1]**Institute of Mathematics of the Romanian Academy,** [2]**Lund University**

[1]{firstname.lastname}@imar.ro, [2]cristian.sminchisescu@math.lth.se

In this supplementary material, we include detail on design choices, methodology, and implementation. We also present additional quantitative insight in order to better asses the impact of our proposed geometric alignment model. For qualitative, visual results, illustrating our 3d reconstructions and contact segmentation predictions, please see our *Supplementary Video* online at our project page http://vision.imar.ro/ci3d.

1. Region Splitting

We use different versions of splitting the body into regions, ranging form coarser ones (75 regions) to finer ones (9 regions). The data collection and predictions methods are done using the initial 75 regions. Yet, since all coarser region splits are obtained by merging the initial fine regions (see fig. 2), we can automatically transfer all annotations and predictions to a smaller number of regions.

2. Contact-Based Tasks - Implementation

We implement the learning methods using the open-source PyTorch [5] framework. We use the architecture of the Graph CNN described in [4], with the following modifications: the input features have size $N_{reg} \times 30$, the branch predicting the camera parameters is removed and the size of the output $F_p^{'}$ is $N_{reg} \times 20$. Each Θ_{S_p} and Θ_{C_p} is a fully connected layer, with $F_p^{''}$ of size $N_{reg} \times 10$.

To standardize the input, we rescale it such that the bounding box of the two people (obtained from the input 2d skeletons) has a reference height of 220px.

Our training procedure employs a data augmentation step. Following [1], we perform random rescaling $50\% - 110\%$, cropping of size 368px $\times$ 368px randomly within 40px from the center of the two body poses, and random horizontal flipping. In addition, we randomly switch the order of the two input poses.

We initialize the networks using the weights of the ResNet50 [3] model trained on ImageNet [2]. Optimization is performed using mini-batch stochastic gradient descent, with 15 images per batch. We start with a learning rate 0.001, with 0.9 momentum and reduce it by a variable factor when the validation loss plateaus. We train the contact classification network for 35 epochs and the contact segmentation and signature network for 100 epochs. For both, we select the model that minimizes the validation loss.

3. Reconstruction Results

To assess the quality of the estimated reconstructions using our alignment term, we plot the histogram of 3D contact distances between the regions in the contact signature. As seen in fig. 1, most of the errors between the regions annotated to be in contact are smaller than 10mm.

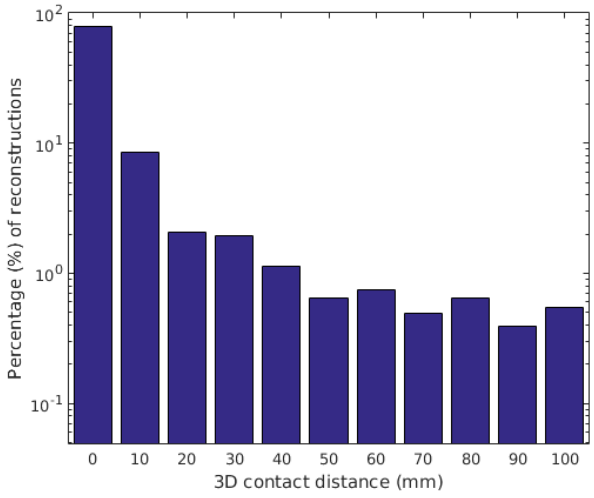

Figure 1: Histogram of the 3d contact distances on the CHI3D dataset (note the logscale on frequency counts). We observe that an overwhelming majority of errors (i.e. 80%) is concentrated between $0 - 10$ mm.

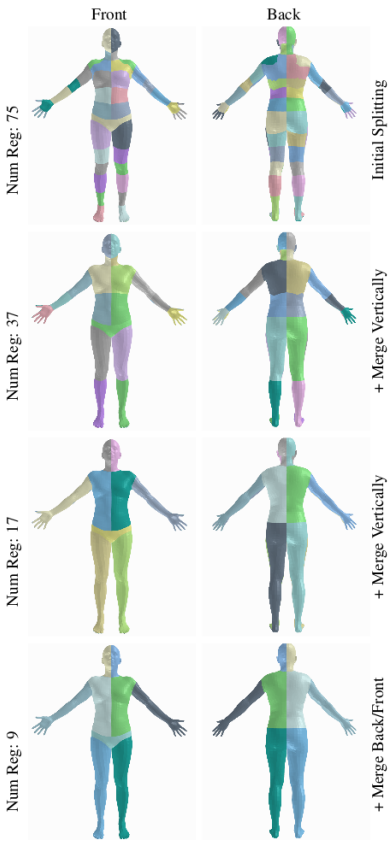

Figure 2: Granularity of region splitting on mesh surfaces. Each row shows a different split of the body surface into regions. We initially split the body into 75 anatomically semantic regions (row 1). Each new splitting is generated by grouping the previous finer regions into coarser ones, either by merging them vertically (rows 2, 3) or by merging corresponding frontal and posterior regions (row 4). See our supplementary video for annotation detail.

References

[1] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *CVPR*, 2017. 1

[2] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR*, 2009. 1

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016. 1

[4] Nikos Kolotouros, Georgios Pavlakos, and Kostas Daniilidis. Convolutional mesh regression for single-image human shape reconstruction. In *CVPR*, 2019. 1

[5] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017. 1