Supplementary Material for Identity-Aware Hand Mesh Estimation and Personalization from RGB Images

Anonymous ECCV submission

Paper ID 1076

1 Introduction

In this supplementary document, we mainly cover the following topics.

- Cross-validation experiments on the DexYCB dataset.
- Additional ablation study.
- Additional qualitative results of our proposed personalization method from only RGB images.

2 Cross-Validation on DexYCB

Table 1: Cross-validation experiments on DexYCB. Three-fold cross-validation is performed.

Method	Split-A { ⋆ }/{7}/{8, 9}		Split-B { ⋆ }/{3}/{0, 2}		Split-C { ⋆ }/{10}/{5, 6}		Average	
	MPJPE	MPVPE	MPJPE	MPVPE	MPJPE	MPVPE	MPJPE ↓	MPVPE ↓
Chen *et al* [2], CMR-PG	20.34	19.88	17.77	17.30	13.50	13.09	**17.20**	**16.76**
Without Optimization at Inference Time								
Baseline	21.58	20.95	17.84	17.33	14.01	13.59	17.81	17.29
Ours, GT Shape	18.83	18.27	16.34	15.86	13.11	12.62	16.09	15.58
Ours, Calibrated	18.97	18.42	16.98	16.51	13.11	12.66	16.35	15.86
With Optimization at Inference Time								
Baseline	18.03	17.92	15.93	15.72	11.75	11.88	15.24	15.17
Ours, GT Shape	16.60	16.29	14.63	14.34	11.18	10.90	14.14	13.84
Ours, Calibrated	16.81	16.55	15.48	15.34	11.18	11.13	**14.49**	**14.34**

Since the DexYCB dataset only contains 10 subjects, we further conduct cross-validation experiments to verify that our model can consistently improve the baseline and outperform the current state-of-the-art method CMR-PG [2]. Experiments are performed on three splits, and results are reported in Table. 1. Split-A refers to the original split across subjects in [1]. In split-A, subject-7 is

in validation set, subject-8 and subject-9 are in test set, while all other seven subjects are allocated to the training set. We denote split-A as { ⋆ }/{7}/{8, 9}. The right column of Table. 1 reports the average performance of these cross-validation experiments.

3 Additional Ablation Studies

Table 2: Ablation study of the proposed method on DexYCB dataset. "GT" is short for ground truth. With GT hand shape, our identity-aware model takes in the GT hand shape parameters. Without GT hand shape, our model takes in the calibrated hand shape parameters which are obtained through our proposed personalization pipeline.

Methods				MJPJE↓
Without Optimization at Inference Time				
Baseline				21.58
	GT hand shape			
Ours	✓			18.83
	✗			18.97
With Optimization at Inference Time				
Baseline				18.03
	GT hand shape	GT root position	GT 2D Predictions	
Ours	✓	✓	✓	2.66
	✓	✗	✓	7.88
	✓	✓	✗	13.85
	✓	✗	✗	16.60
	✗	✗	✗	**16.81**

4 Qualitative Results of our personalization pipeline.

We further provide the comparison of the calibrated hand mesh versus the ground truth hand mesh for all the subjects in the test set. Note that all these test subjects are *not* seen in the training set. For each subject, please refer to the PNG image named as `$dataset_$subjectID_gt_and_calibrated.png` under the folder `calibrated_versus_ground_truth_hand_meshes`.

References

1. Chao, Y.W., Yang, W., Xiang, Y., Molchanov, P., Handa, A., Tremblay, J., Narang, Y.S., Van Wyk, K., Iqbal, U., Birchfield, S., et al.: Dexycb: A benchmark for cap-

turing hand grasping of objects. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9044–9053 (2021)
2. Chen, X., Liu, Y., Ma, C., Chang, J., Wang, H., Chen, T., Guo, X., Wan, P., Zheng, W.: Camera-space hand mesh recovery via semantic aggregation and adaptive 2d-1d registration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 13274–13283 (2021)