Code

Identity-Aware Hand Mesh Estimation and Personalization from RGB Images

Deying Kong[1], Linguang Zhang, Liangjian Chen, Haoyu Ma[1], Xiangyi Yan[1],Shanlin Sun[1], Xingwei Liu[1], Kun Han[1],Xiaohui Xie[1]

[1] University of California, Irvine [2] Meta (Facebook)

Paper

Problem Setting

Reconstruct 3D hand mesh from monocular RGB images.

Motivation

- Most of the SOTA methods are subject-***agnostic***.
  (a) The identity of the subject is often *ignored*.
  (b) However, this identity information is often practically *available* in VR/AR applications.
  (c) The consistency in hand shape (hand size, finger fatness etc.) is *not* strictly enforced among the images from same subject.

  We raise the *first question*:

  Can 3D hand reconstruction be further improved with the help of identity information?

- In practice, subjects ***unseen*** from the training set require hand model calibration. Existing methods use ***depth image*** to perform hand model personalization, which requires dedicated hardwares and complex procedure.

  We raise the *second question*:

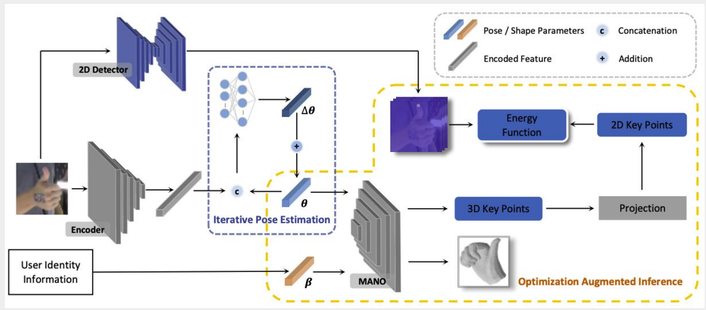
  Could we calibrate the hand model for unseen subject by using only RGB images?

Main Contributions

- Our work is the ***first*** to
  (a) systematically investigate the problem of hand mesh personalization from **only RGB** images, and
  (b) demonstrate its benefits to hand mesh and keypoints reconstruction via an ***Identity-aware*** hand mesh estimation model.
- A novel ***hand model personalization*** method is designed.
  For unknown subjects that are not seen in training, the proposed method is capable of calibrating the hand model using a few (<=20) ***unannotated RGB images*** of the same subject.
- Performance are evaluated on two large-scale public datasets, HUMBI and DexYCB.

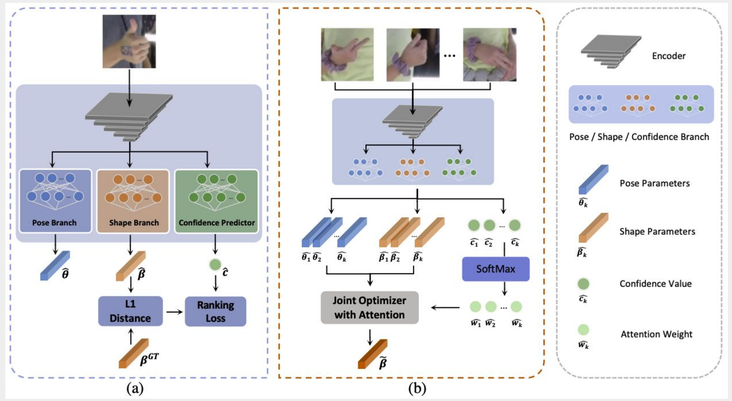
Proposed Method

Identity-aware hand mesh estimation model.



Key proposal: Instead of *estimating* MANO shape parameters from the input image, we feed the *groundtruth/calibrated* shape parameters β directly into the network, explicitly forcing the **shape consistency** among images from the same subject.

Hand calibration pipeline from ***only*** RGB images.



(a) A ***confidence predictor*** is trained on top of the baseline model via a ranking loss. The baseline model differs from our model in that it also predicts the MANO shape parameters β.

(b) During calibration phase, several images from the same subject are fed into the baseline model. The final calibrated shape is obtained by solving the following optimization problem, where $\mathcal{M}(\cdot)$ denotes the MANO model.

$$\min_{\tilde{\beta}} \sum_{k=1}^{K} w_k \cdot \|\mathcal{M}(\tilde{\beta}, \hat{\theta}_k) - \mathcal{M}(\hat{\beta}_k, \hat{\theta}_k)\|_F$$

Quantitative & Qualitative Results

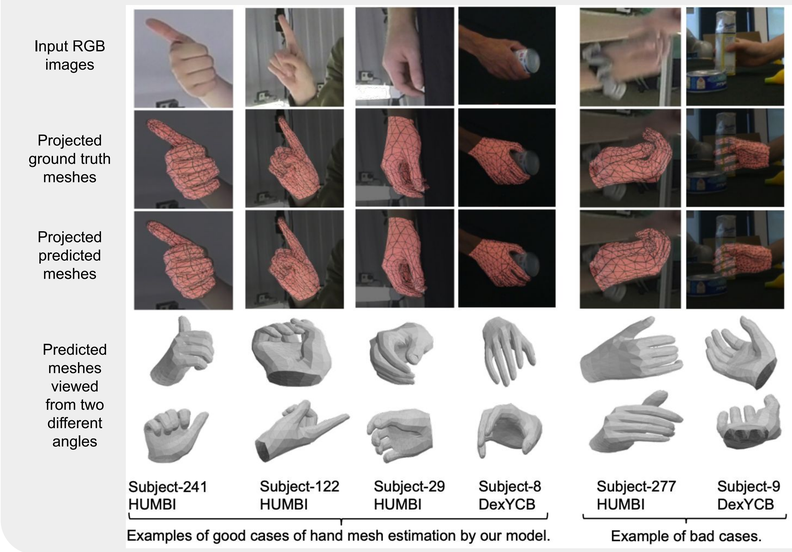
Results on mesh/keypoints reconstruction

Table 1: Numerical results on DexYCB and HUMBI datasets.

Method	DexYCB		HUMBI	
	MPJPE ↓	MPVPE ↓	MPJPE ↓	MPVPE ↓
CMR-PG [9]	20.34	19.88	11.64	11.37
Without Optimization at Inference Time				
Baseline	21.58	20.95	12.13	11.82
Ours, GT shape	18.83	18.27	11.41	11.11
Ours, Calibrated	18.97	18.42	11.51	11.21
With Optimization at Inference Time				
Baseline	18.03	17.92	10.75	10.60
Ours, GT shape	16.60	16.29	10.17	9.94
Ours, Calibrated	**16.81**	**16.55**	**10.31**	**10.28**

Table 2: Comparison with existing methods on Dex-YCB.

Methods	MPJPE↓	MPVPE ↓
Boukhayma et al. [4]	27.94	27.28
Spurr *et al* [42] + ResNet50	22.71	-
Spurr *et al* [42] + HRNet32	22.26	-
Boukhayma et al. [4] †	21.20	21.56
CMR-PG [9]	20.34	19.88
Metro [24]	19.05	17.71
Ours, Calibrated	**16.81**	**16.55**



Results on hand shape calibration/personalization.

Table 3: Performance of hand model calibration.

Metrics	HUMBI	DexYCB
MSE_{mano}	0.07	0.04
W-error (mm)	0.88	1.02
L-error (mm)	1.71	1.20

W-error: hand width error.
L-error: hand length error.