Learning to Fit Morphable Models

Vasileios Choutas[1, 2, *], Federica Bogo[2, *], Jingjing Shen [2], Julien Valentin[2]

[1]MPI for Intelligent Systems, Tübingen, DE, [2]Microsoft,

* Work performed at Microsoft

Microsoft

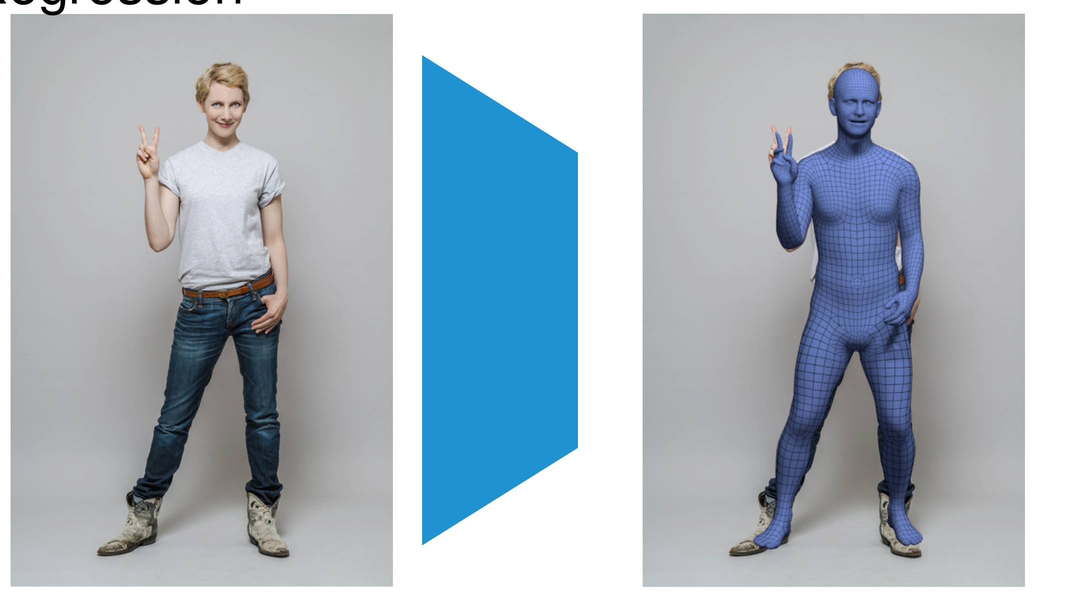
MAX PLANCK INSTITUTE FOR INTELLIGENT SYSTEMS

Goal

Fit parametric models of human bodies, hands or faces to sparse input signals in an accurate, robust, and fast manner.

Related work

Regression



- Accurate
- Robust
- Misalignment with input data

ExPose [5]

Optimization

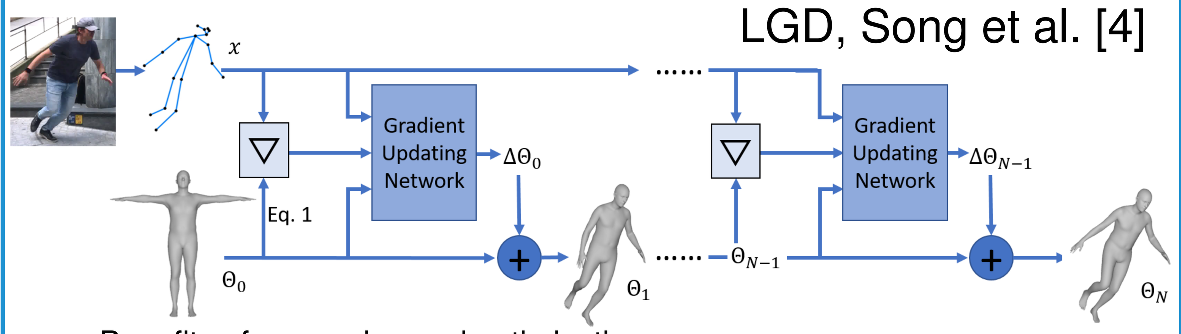
$$E(\Phi; L) = \underbrace{E_{\text{landmarks}}}_{\text{Data term}} + \underbrace{E_{\text{identity}} + E_{\text{expression}} + E_{\text{joints}} + E_{\text{temporal}} + E_{\text{intersect}}}_{\text{Regularizers}}$$

Wood et al. [3]

- Tight alignment to observations
- Hard to come-up and balance energy terms
- Significant time investment for real-time performance

Learned optimization

LGD, Song et al. [4]



- Benefits of regression and optimization
- Learn priors from data
- Fast inference thanks to performant Machine Learning backends.

Neural Fitter

Levenberg–Marquardt algorithm:

$$(J^T J + \lambda \text{diag}(J^T J))\Delta\mathbf{\Theta} = J^T \mathcal{R}$$

Learned optimization (ours):

$$u(\Delta\mathbf{\Theta}_n, \boldsymbol{g}_{n-1}, \mathbf{\Theta}_{n-1}) = \boldsymbol{\lambda}\Delta\mathbf{\Theta}_n + (-\boldsymbol{\gamma}\boldsymbol{g}_{n-1})$$

$$\boldsymbol{\lambda}, \boldsymbol{\gamma} = f_{\boldsymbol{\lambda},\boldsymbol{\gamma}}(\mathcal{R}(\mathbf{\Theta}_{n-1}), \mathcal{R}(\mathbf{\Theta}_{n-1} + \Delta\mathbf{\Theta}_n)), \boldsymbol{\lambda}, \boldsymbol{\gamma} \in \mathbb{R}^{|\mathbf{\Theta}|}$$

Algorithm Neural fitting

Require: Input data D

$\mathbf{\Theta}_0 = \Phi(D)$

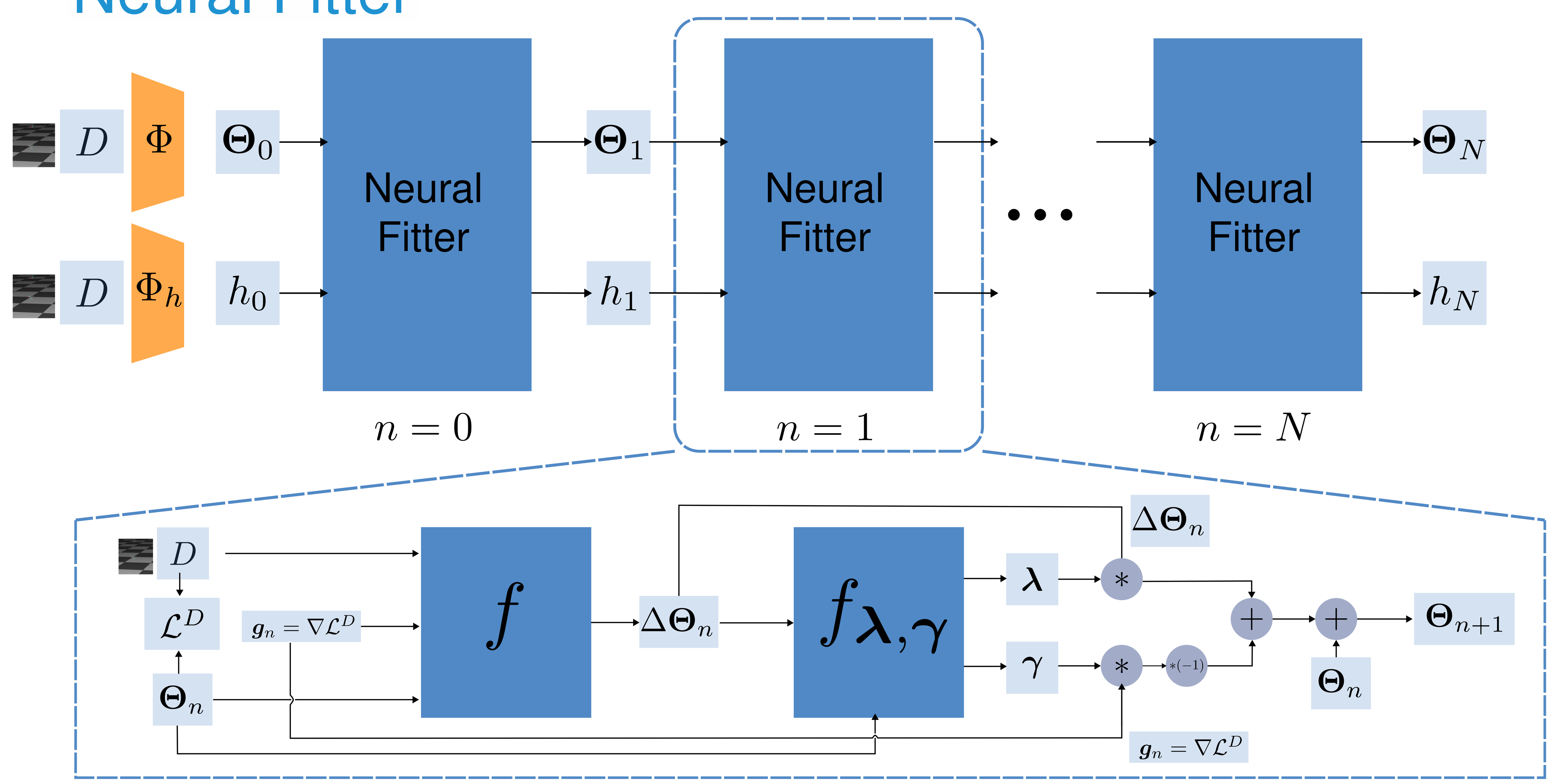
$h_0 = \Phi_h(D)$

while not converged **do**

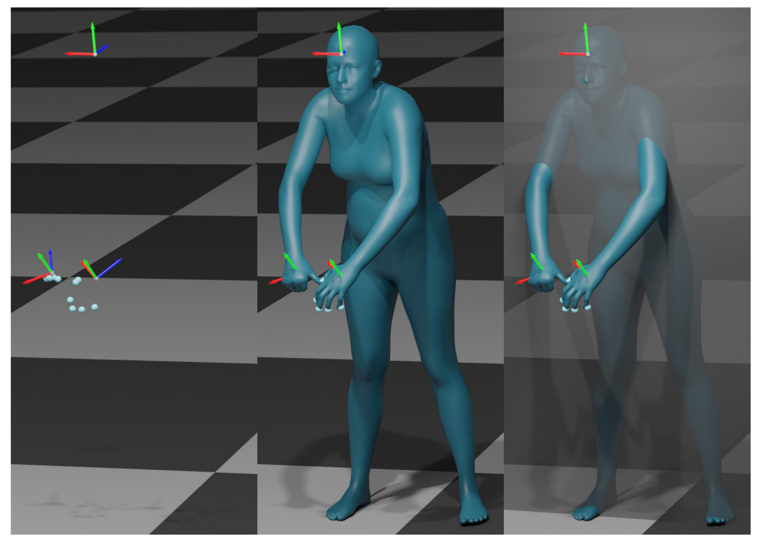
$\Delta\mathbf{\Theta}_n, h_n \leftarrow f([\boldsymbol{g}_{n-1}, \mathbf{\Theta}_{n-1}], D, h_{n-1})$

$\mathbf{\Theta}_n \leftarrow \mathbf{\Theta}_{n-1} + u(\Delta\mathbf{\Theta}_n, \boldsymbol{g}_{n-1}, \mathbf{\Theta}_{n-1})$

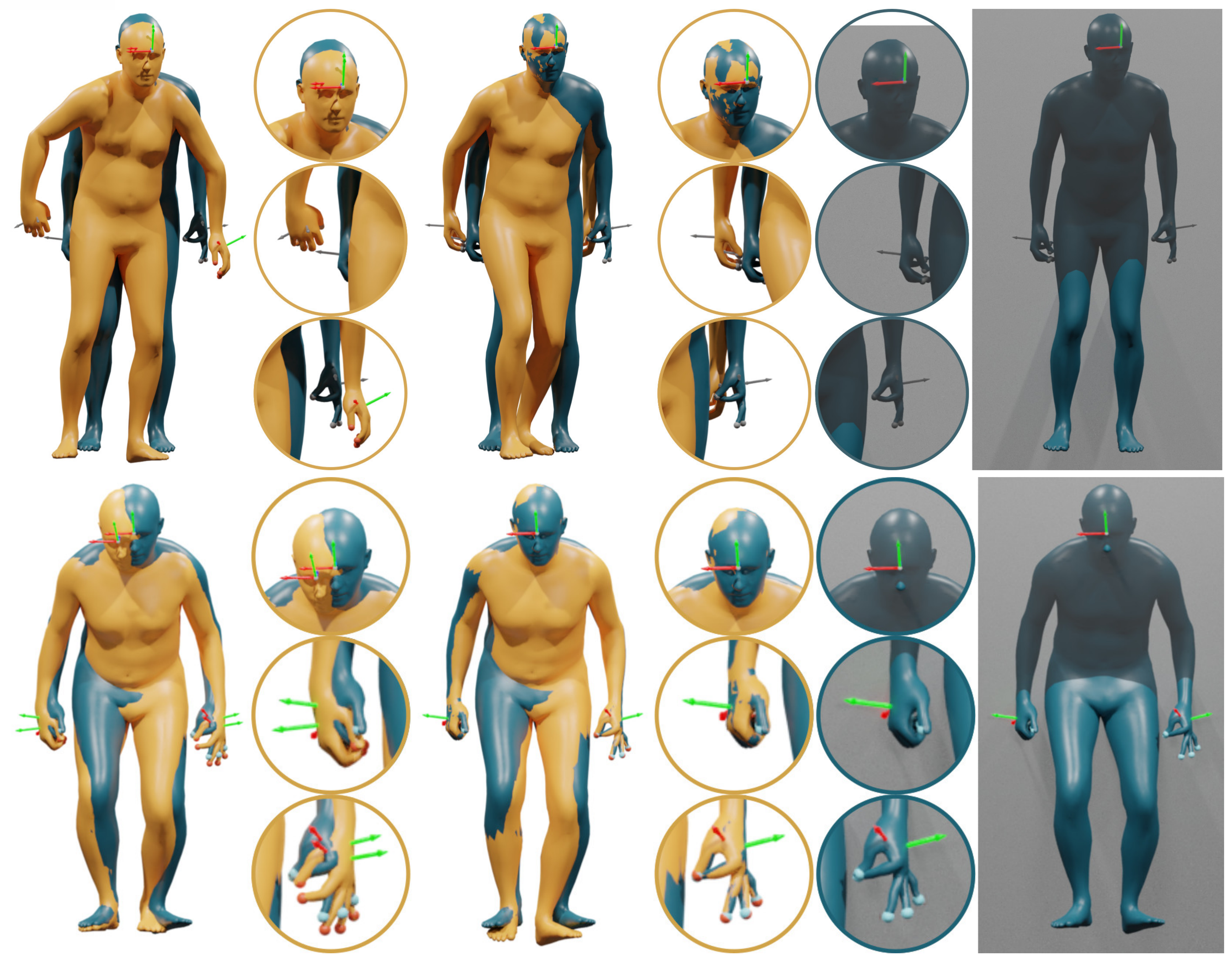
end while



Body fitting to HMD data



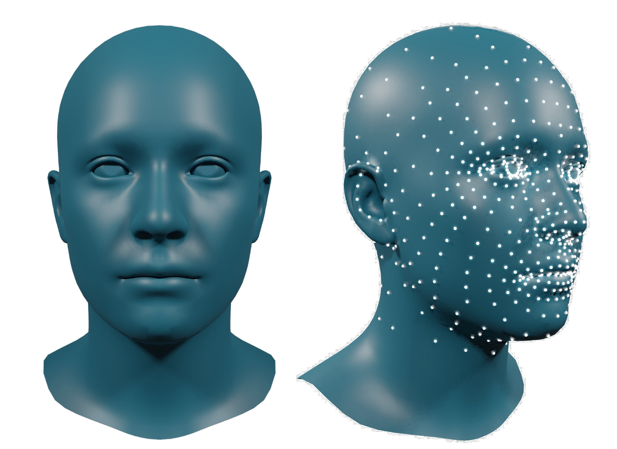
Method	Vertex-to-Vertex (mm) ↓ Full body F	Full body H	Head F	Head H	L / R hand F	L / R hand H	MPJPE (mm) ↓ F	MPJPE (mm) ↓ H	GrPe. (mm) ↓ F	GrPe. (mm) ↓ H
L-BFGS, GMM	73.1	116.2	2.9	3.4	3.2 / 3.0	5.6 / 5.3	49.7	137.26	70.8	74.0
L-BFGS, GMM, Tempo.	72.6	113.3	2.9	3.4	3.3 / 3.1	6.8 / 6.5	49.4	132.1	70.7	73.5
L-BFGS, VAE Enc.	76.1	119.3	3.9	4.1	5.3 / 4.7	8.7 / 7.6	52.6	140.5	63.6	66.7
Dittadi et al. [2]	N/A		N/A		N/A		43.3	N/A	N/A	
Ours $\Phi, (N = 0)$	44.2	69.7	19.1	22.7	27.8 / 25.9	32.1 / 29.9	38.9	84.9	16.1	20.1
Ours $(N = 5)$	**26.1**	**49.9**	**2.2**	**3.2**	**3.0 / 3.3**	**3.1 / 3.7**	**18.1**	**62.1**	**12.5**	**15.5**



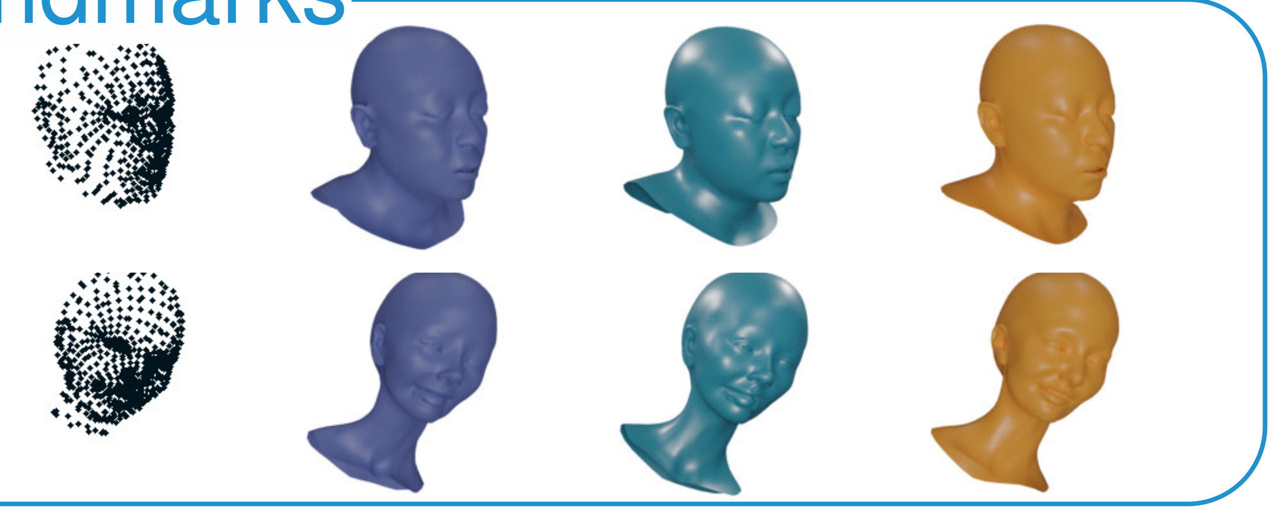
Body fitting to 2D keypoints

Method	Type	Image	2D keypoints	Part segmentation	PA-MPJPE
SMPLify [1]	O	✗	✓	✗	106.1
SCOPE [7]	O	✗	✓	✗	68.0
VIBE [6]	R	✓	✗	✗	55.9
Neural Descent [8]	R+O	✓	✓	✓	57.5
LGD [4]	R+O	✗	✓	✗	55.9
Ours (full)	R+O	✗	✓	✗	**52.2**

Face fitting to 2D landmarks



Method	V2V (mm) ↓ Face -	Face PA	Head -	Head PA	LdmkErr (mm) ↓ -	LdmkErr (mm) ↓ PA
LM	34.4	3.7	33.8	5.3	33.8	**3.4**
Ours	**7.9**	**3.5**	**8.5**	**4.1**	**8.0**	3.7



References

[1] Bogo et al., Keep it SMPL: Automatic Estimation of 3D Human Pose and Shape from a Single Image, ECCV, 2016
[2] Dittadi et al., Full-Body Motion from a Single Head-Mounted Device: Generating SMPL Poses from Partial Observations, ICCV, 2021
[3] Wood et al., 3D face reconstruction with dense landmarks , ECCV, 2022
[4] Song et al., Human Body Model Fitting by Learned Gradient Descent, ECCV, 2020
[5] Choutas et al., Monocular Expressive Body Regression through Body-Driven Attention, ECCV, 2020
[6] Kocabas et al., VIBE: Video Inference for Human Body Pose and Shape Estimation, CVPR, 2020
[7] Fan et al., Revitalizing Optimization for 3D Human Pose and Shape Estimation: A Sparse Constrained Formulation, ICCV, 2021
[8] Zanfir et al., Neural Descent for Visual 3D Human Pose and Shape , CVPR, 2021