Accurate 3D Body Shape Regression using Metric and Semantic Attributes

Vasileios Choutas*, [1,2], Lea Müller*, [1], Chun-Hao Paul Huang [1], Siyu Tang[2], Dimitrios Tzionas[1,3], Michael J. Black[1]

[1]MPI for Intelligent Systems, Tübingen, DE, [2]ETH Zürich, CH, [3]University of Amsterdam, NL

* Equal contribution, alphabetical order

{vchoutas, lea.mueller, paul.huang, black}@tue.mpg.de, d.tzionas@uva.nl ,siyu.tang@inf.ethz.ch

CVPR JUNE 19-24 2022 NEW ORLEANS LOUISIANA

shapy.is.tue.mpg.de

Models, Data, Code

MAX PLANCK INSTITUTE FOR INTELLIGENT SYSTEMS

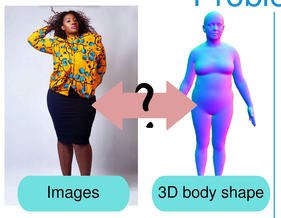
ETH zürich

UNIVERSITY OF AMSTERDAM

Goal

3D body shape estimation from RGB images

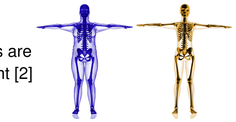
Problem



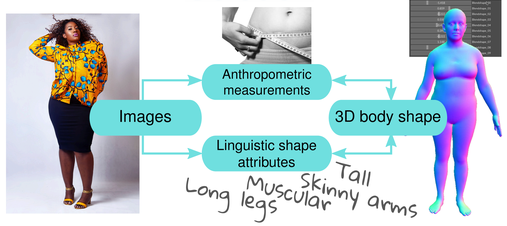
Impractical to scan 1000s of subjects

Silhouettes unsuitable for loose clothing [5]

Keypoints are insufficient [2]



Solution



References

[1] Madadi et al., SMPLR: Deep learning based SMPL reverse for 3D human pose and shape recovery, PR, 2020

[2] Sengupta et al., Synthetic training for accurate 3D human pose and shape estimation in the wild, BMVC, 2020

[3] Kolotouros et al., Learning to reconstruct 3D human pose and shape via model-fitting in the loop, ICCV, 2019

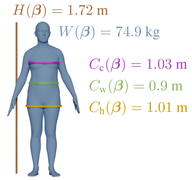
[4] Müller et al., On self-contact and human pose, CVPR, 2021

[5] Sengupta et al., Hierarchical kinematic probability distributions for 3D human shape and pose estimation from images in the wild, ICCV, 2021

[6] Choutas et al., Monocular expressive body regression through body-driven attention, ECCV 2020

Data

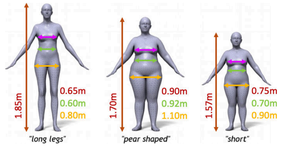
Male & Female		Male only	Female only
short	long neck	skinny arms	pear shaped
big	long legs	average	petite
tall	long torso	rectangular	slim waist
muscular	short arms	delicate build	large breasts
	broad shoulders	soft body	skinny legs
		masculine	feminine



Linguistic shape attributes: $A \in \{1,2,3,4,5\}^{N \times A \times K}$
- N subjects
- A attributes
- K annotators

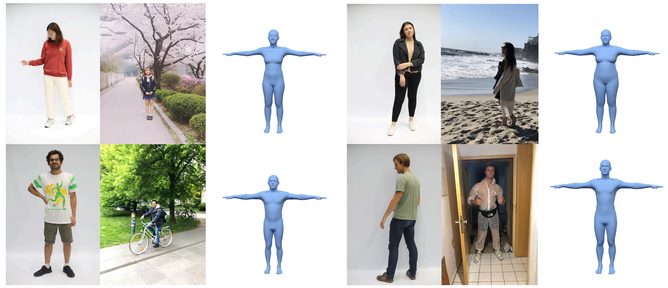
Anthropometric measurements:
- Height: $H(\beta)$
- Weight: $W(\beta)$
- Chest, waist and hips circumference: $C_c(\beta)$, $C_w(\beta)$ and $C_h(\beta)$

CAESAR:
- Linguistic shape attributes
- Anthropometric measurements
- SMPL-X shape parameters, β



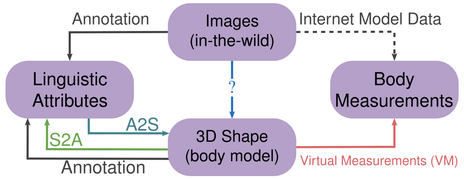
Model agency dataset:
- Images
- Linguistic shape attributes
- Anthropometric measurements

Human Bodies in the Wild (HBW)



- 3D body scans for 35 subjects (20 female, 15 male) and SMPL-X registrations
- 1318 lab and 1225 in-the-wild photos for each subject
- Validation set: 10 subjects (6 female, 4 male), 781 images
- Test set: 25 subjects (14 female, 11 male), 1,762 images

Shape Mappings



Attributes to Shape (A2S): $\beta = \text{A2S}(\bar{A}), \bar{A}_{ij} = \frac{1}{K}\sum_{k=1}^{K} A_{ijk}$

Shape to Attributes (S2A): $\hat{A} = \text{S2A}(\beta), \hat{A} \in [0,5]$

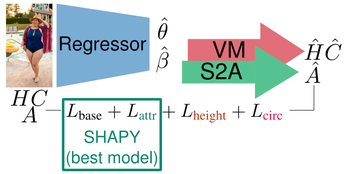
Virtual Measurements (VM): $H, W, C_c, C_w, C_h = \text{VM}(\beta)$

Attributes & Measurements to Shape (AHWC2S):
- $\beta = \text{AHWC2S}(\bar{A}, H, W, C_c, C_w, C_h)$
- $\beta = \text{HC2S}(H, C_c, C_w, C_h)$

Method	P2P$_{20K}$ (mm) ↓ Female	Male
A2S	10.9 ± 5.2	11.1 ± 5.2
H2S	12.8 ± 7.0	12.1 ± 6.1
AH2S	7.2 ± 2.8 (-5.6 mm)	6.8 ± 2.3 (-5.3 mm)
HW2S	7.9 ± 3.2	8.1 ± 2.7
AHW2S	6.4 ± 2.5 (-1.5 mm)	6.3 ± 2.1 (-1.8 mm)
C2S	19.5 ± 10.8	19.7 ± 11.1
AC2S	9.6 ± 4.3 (-9.9 mm)	9.6 ± 4.4 (-10.1 mm)
HC2S	7.3 ± 2.8	7.7 ± 2.6
AHC2S	6.3 ± 2.4 (-1.0 mm)	6.0 ± 2.0 (-1.7 mm)
HWC2S	7.2 ± 2.9	7.3 ± 2.6
AHWC2S	6.2 ± 2.4 (-1.0 mm)	5.8 ± 2.0 (-1.5 mm)

Predicting 3D shape *with* attributes is *always more accurate*.

SHAPY



$$L_{\text{attr}} = ||A - \hat{A}||_2^2$$

$$L_{\text{height}} = ||H - \hat{H}||_2^2$$

$$L_{\text{circ}} = \sum_{i \in \{c,w,h\}} ||C_i - \hat{C}_i||_2^2$$

Results

Ground-truth SHAPY Sengupta et al. [5]

Method	Model	Height	Chest	Waist	Hips	P2P$_{20K}$
SMPLR [1]	SMPL	182	267	309	305	69
STRAPS [2]	SMPL	135	167	145	102	47
SPIN [3]	SMPL	59	92	78	101	29
TUCH [4]	SMPL	58	89	75	**57**	26
Sengupta et al. [5]	SMPL	82	133	107	63	32
ExPose [6]	SMPL-X	85	99	92	94	35
SHAPY (ours)	SMPL-X	**51**	**65**	**69**	**57**	**21**

HBW test set results. All errors in mm.