AGORA: Avatars in Geography Optimized for Regression Analysis

Priyanka Patel[1], Chun-Hao P. Huang[1], Joachim Tesch[1], David T. Hoffmann[2,3], Shashank Tripathi[1], Michael J. Black[1]
Max Planck Institute for Intelligent Systems, Tübingen, Germany, University of Freiburg, Bosch Center for Artificial Intelligence

PERCEIVING SYSTEMS
MAX PLANCK INSTITUTE FOR INTELLIGENT SYSTEMS

CVPR VIRTUAL JUNE 19-25

Overview

Goal

Provide challenging image dataset to evaluate and improve the SOTA in 3D human pose and shape estimation

Existing datasets limitation:

Method	Complex Clothing	Multi-Person	Occlusion	Full body GT
Human3.6M[1]	✗	✗	✗	✗
HUMBI [2]	✓	✗	✗	✓
3DPW [3]	✓	✓	✓	✗
EFT [4]	✓	✓	✓	✗
SMPLy [5]	✓	✓	✓	✗
AGORA	✓	✓	✓	✓

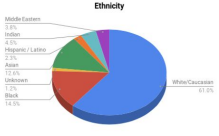
AGORA preparation

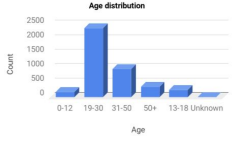
- ➔ fitted 4240 SMPL-X [7] to commercial scans with high accuracy of 5mm shape error
- ➔ render 5-15 scans per image using Unreal Engine
- ➔ used image based lighting and 3D environment
- ➔ span 350 unique subjects; include 257 kid scans

AGORA

- ➔ 14529 training and 1225 validation images with corresponding masks and SMPL-X [7]/SMPL [8] fits
- ➔ 3387 test images with web-based evaluation server
- ➔ 173K individual person crops in total

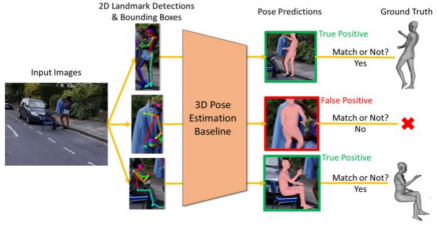
Age and Ethnicity





Baseline Evaluation

	Method	MPJPE ↓ B	MPJPE ↓ LH/RH	MPJPE ↓ F	MPJPE ↓ FB	MVE ↓ B	MVE ↓ LH/RH	MVE ↓ F	MVE ↓ FB	NMJE ↓ B	NMJE ↓ FB	NMVE ↓ B	NMVE ↓ FB	F1 score↑
SMPL	HMR [6]	180.5	N/A	N/A	N/A	173.6	N/A	N/A	N/A	226.0	N/A	217.0	N/A	0.80
	CenterHMR [10]	168.1	N/A	N/A	N/A	161.4	N/A	N/A	N/A	242.3	N/A	233.9	N/A	0.69
	EFT [4]	165.4	N/A	N/A	N/A	159.0	N/A	N/A	N/A	203.6	N/A	196.3	N/A	**0.81**
	SPIN [9]	175.1	N/A	N/A	N/A	168.7	N/A	N/A	N/A	223.1	N/A	216.3	N/A	0.78
	SPIN-ft (ours)	**153.4**	N/A	N/A	N/A	**148.9**	N/A	N/A	N/A	**199.2**	N/A	**193.4**	N/A	0.77
SMPL-X	SMPLify-X [7]	182.1	**46.5/49.6**	**52.9**	231.8	187.0	**48.3/51.4**	**48.9**	236.5	256.5	326.5	263.3	333.1	0.71
	ExPose [11]	**150.4**	72.5/68.8	55.2	**215.9**	**151.5**	74.9/71.3	51.1	**217.3**	**183.4**	**263.3**	**184.8**	**265.0**	**0.82**
	Frankmocap [12]	165.2	52.3/53.1	N/A	N/A	168.3	54.7/55.7	N/A	N/A	204.0	N/A	207.8	N/A	0.81



Evaluation Metric and Protocol:

- ➔ TP, FP and FN are detected by matching the projected GT and prediction keypoints
- ➔ MPJPE/MVE: pelvis-aligned joint and vertex error
- ➔ NMJE/NMVE to facilitate multiperson 3DHPS analysis

`NMJE or NMVE = (MPJPE or MVE)/F1`

- ➔ Low NMJE/NMVE -> low false positive and misses
- ➔ SMPL-X evaluated on body(B), faces(F) and hand(LH/RH) individually and together (FB)

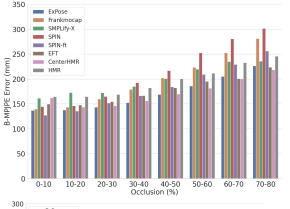
`FB = B+(LH+RH+F)/3`

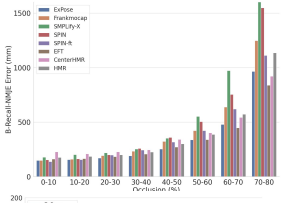
Baseline Finetuning

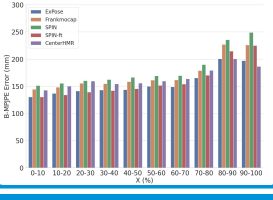
Models	3DPW (14) MPJPE	3DPW (14) PA-MPJPE	3DPW (24) MPJPE	3DPW (24) PA-MPJPE	AGORA (24) MPJPE
SPIN-pt	96.9	59.3	95.5	65.5	175.1
SPIN-ft (ours)	**85.7**	**55.3**	**83.7**	**61.8**	**153.4**

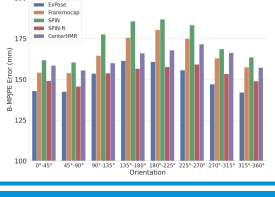
- ➔ finetuning SPIN [9] with AGORA training images improves MPJPE by 12% for 3DPW [3] and ~13% for AGORA
- ➔ improved accuracy on 3DPW (natural image dataset) confirms the realism of AGORA

Baseline Analysis









Occlusion:

- ➔ high occlusion leads to high MPJPE error and high misses
- ➔ methods that miss lots of people suffer under recall NMJE

Distance from center:

- ➔ weak perspective cameras assumption by baselines
- ➔ error increase for off-center people in the image for large fov images

Orientation:

- ➔ error increase for people facing opposite to the camera

References

1. Human3.6M; Ionescu et al., TPAMI 2014.
2. HUMBI; Yu et al., CVPR 2020.
3. 3DPW; von Marcard et al., ECCV 2018.
4. EFT; Joo et al., arXiv 2020.
5. SMPLy; Leroy et al., 3DV 2020.
6. HMR; Kanazawa et al., CVPR 2018.
7. SMPLify-X; Pavlakos et al., CVPR 2019.
8. SMPL; Loper et al., SIGGRAPH Asia 2015.
9. SPIN; Kolotouros et al., ICCV 2019.
10. CenterHMR; Sun et al., arXiv 2020.
11. ExPose; Choutas et al., ECCV 2020.
12. FrankMocap; Rong et al., arXiv 2020.