HuManiFlow: Ancestor-Conditioned Normalising Flows on SO(3) Manifolds for Human Pose and Shape Distribution Estimation

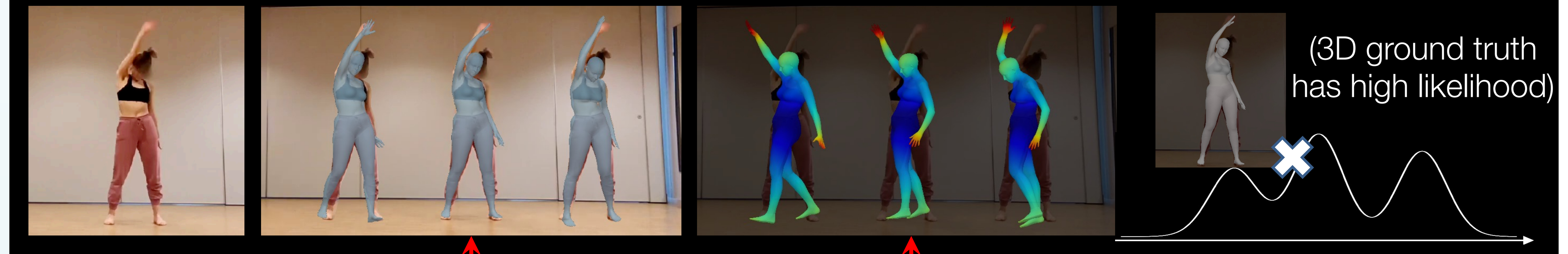
UNIVERSITY OF CAMBRIDGE

Akash Sengupta Ignas Budvytis Roberto Cipolla

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

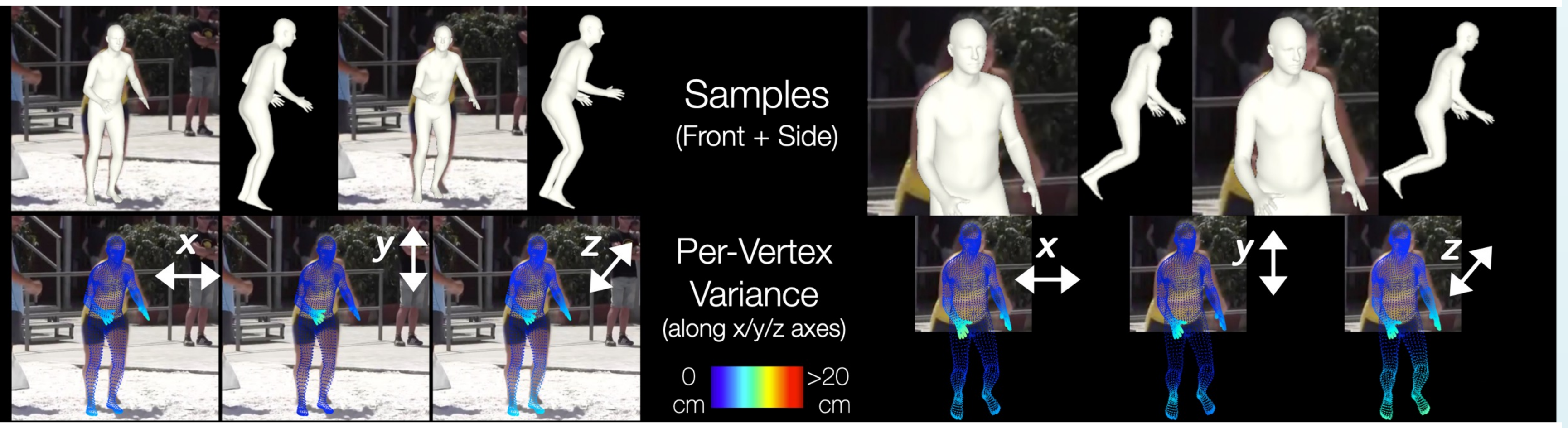
Motivation

Multiple 3D human reconstructions can correspond to a 2D image due to depth ambiguity, occlusion and truncation. → Motivates a *probability distribution over 3D pose and shape*, which should exhibit 3 properties…

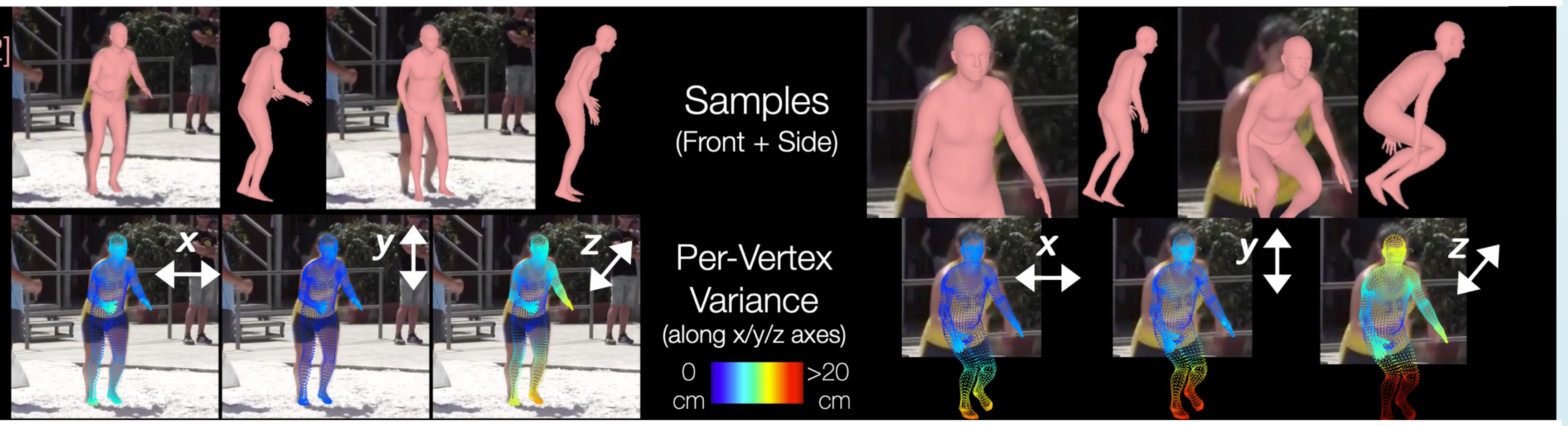


Trade-off in current probabilistic methods[1,2]

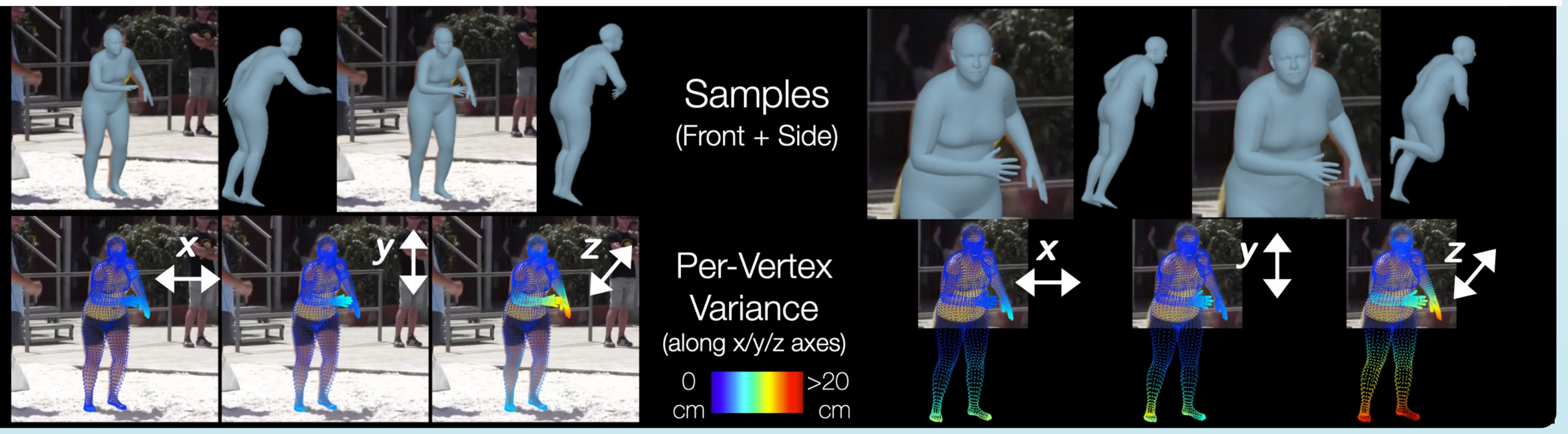
ProHMR[1]
Input-Consistent in 2D
Not Diverse in 3D



3D Multibodies[2]
Not Input-Consistent in 2D
Diverse in 3D



HuManiFlow
Input-Consistent in 2D
Diverse in 3D

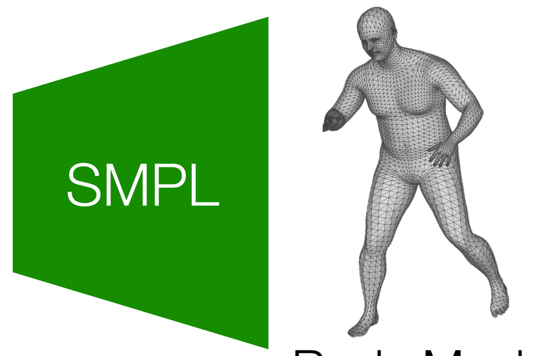


Method

We use the SMPL[3] 3D body model.

Shape $\boldsymbol{\beta}$

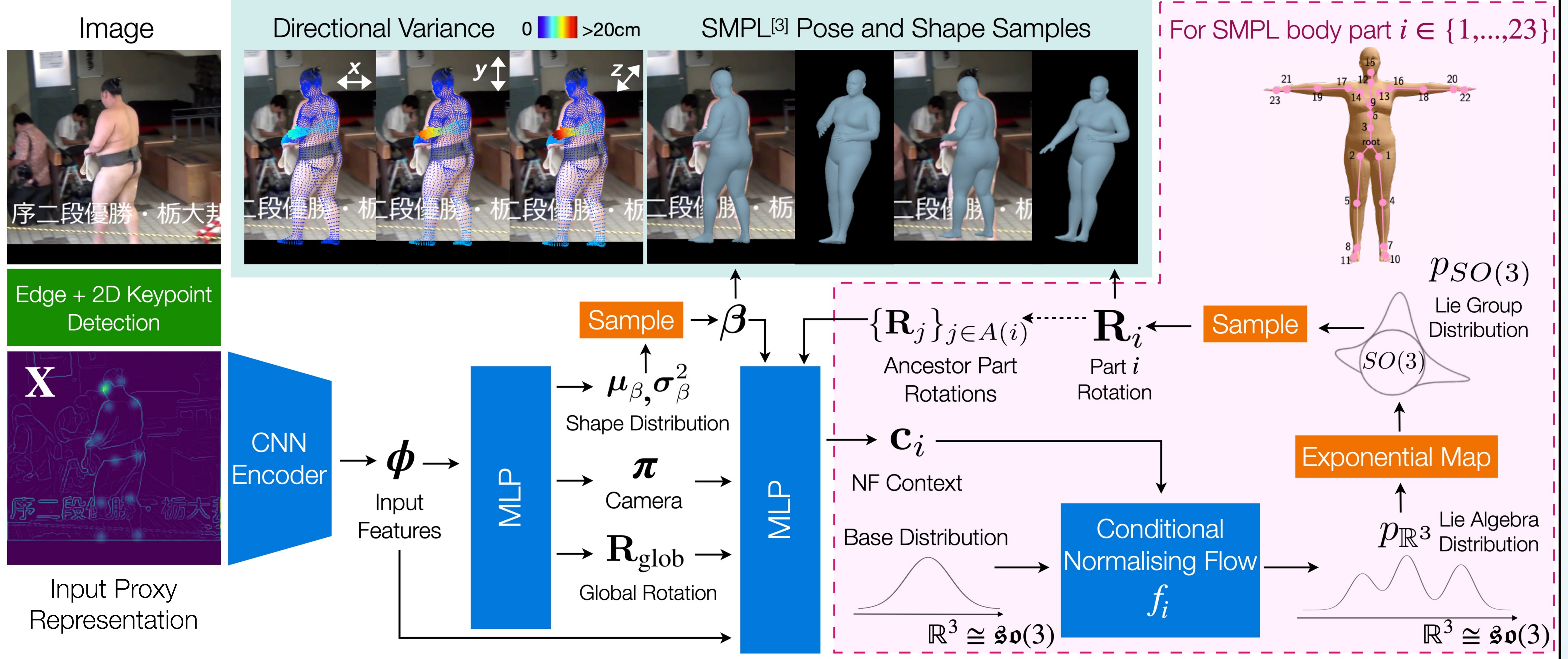
Pose $\{\mathbf{R}_i\}_{i=1}^{23}$



3D rotation of each body-part about its parent joint.
Body-part rotations belong to the Lie group $SO(3)$.

We predict a distribution over SMPL pose and shape conditioned on a 2D input $\mathbf{X}$.

$$p_{\text{joint}}(\{\mathbf{R}_i\}_{i=1}^{23}, \boldsymbol{\beta}|\mathbf{X}) = p_{\text{shape}}(\boldsymbol{\beta}|\mathbf{X})\, p_{\text{pose}}(\{\mathbf{R}_i\}_{i=1}^{23}|\boldsymbol{\beta}, \mathbf{X})$$



Full-body pose is factorised into per-body-part rotation distributions *conditioned on ancestor body part rotations.*

$$p_{\text{pose}}(\{\mathbf{R}_i\}_{i=1}^{23}|\boldsymbol{\beta}, \mathbf{X}) = \prod_{i=1}^{23} p_{SO(3)}(\mathbf{R}_i|\{\mathbf{R}_j\}_{j\in A(i)}, \boldsymbol{\beta}, \mathbf{X})$$

Aggregated into context $\mathbf{C}_i$

Ancestors of part i

Per-body-part distributions are normalising flows on the Lie algebra $\mathfrak{so}(3) \cong \mathbb{R}^3$.
These are *pushed through the exp map* onto $SO(3)$ using change-of-variables.

$$p_{SO(3)}(\mathbf{R}_i|\mathbf{c}_i) = \sum_{k\in\mathbb{Z}} p_{\mathbb{R}^3}(\mathbf{v}_i^k|\mathbf{c}_i)\,|\det J_{\exp}(\mathbf{v}_i^k)|^{-1}$$

where $\mathbf{R}_i = \exp(\mathbf{v}_i^k)$ and $\mathbf{v}_i^k = (\theta_i + 2\pi k)\mathbf{u}_i$

Rodrigues' rotation formula Angle Axis

Results

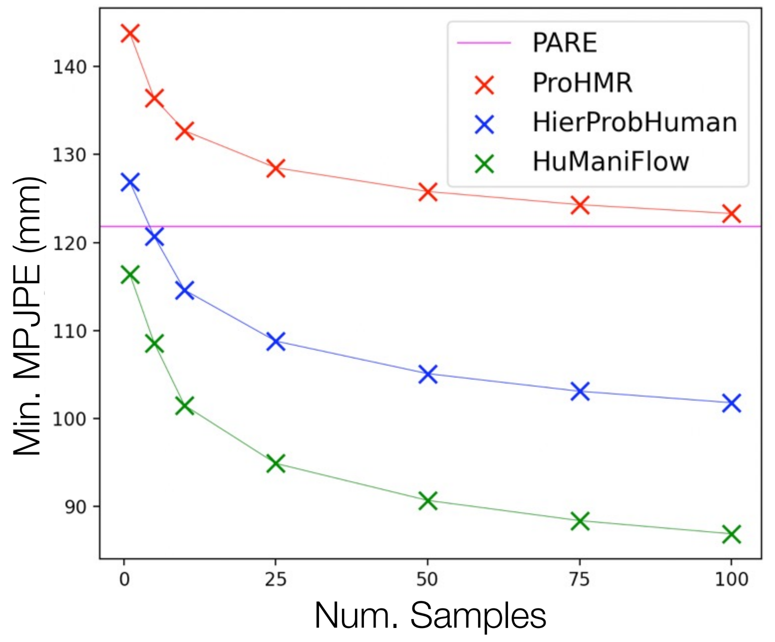
2D input-consistency and 3D sample diversity of probabilistic pose and shape methods on 3DPW.

Method	*Sample Consistency* 2DKP Error (pixels)	*Sample Diversity* 3DKP Spread (mm) Visible / Invisible
3D Multibodies	7.8	80.1 / **126.9**
ProHMR	7.5	35.1 / 60.8
HierProbHuman	7.2	47.6 / 101.4
HuManiFlow	**6.2**	42.8 / 116.0

Input-consistent in 2D because we:
(i) Use kinematic tree to factorise body pose.
(ii) Consider domain of body-part rotations $SO(3)$.

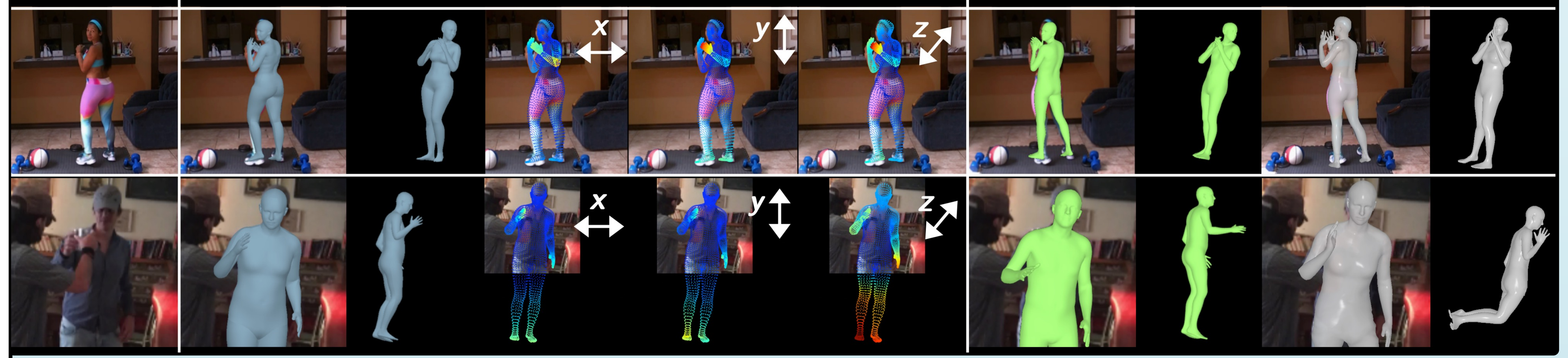
Diverse in 3D because we:
(i) Use expressive distribution models (flows).
(ii) Don't use ill-posed loss functions (e.g. 3D MSE).

Min. sample MPJPE on 3DPW Cropped



Faster rate of improvement → 3D ground-truth has higher likelihood under the predicted distribution.

Comparison between HuManiFlow and recent deterministic 3D pose and shape predictors. HuManiFlow handles occlusion and truncation. Per-vertex variance indicates directional uncertainty → useful for downstream tasks.



References

[1] N. Kolotouros et al. Probabilistic modeling for human mesh recovery. ICCV 2021.
[2] B. Biggs et al. 3D Multibodies: Fitting sets of plausible 3D models to ambiguous image data. NeurIPS 2020.
[3] M. Loper et al. SMPL: A Skinned Multi-Person Linear model. ACM SIGGRAPH Asia 2015.
[4] M. Kocabas et al. PARE: Part attention regressor for 3D human body estimation. ICCV 2021.
[5] Li et al. Hybrik: A hybrid analytical-neural inverse kinematics solution for 3d human pose and shape estimation. CVPR 2021