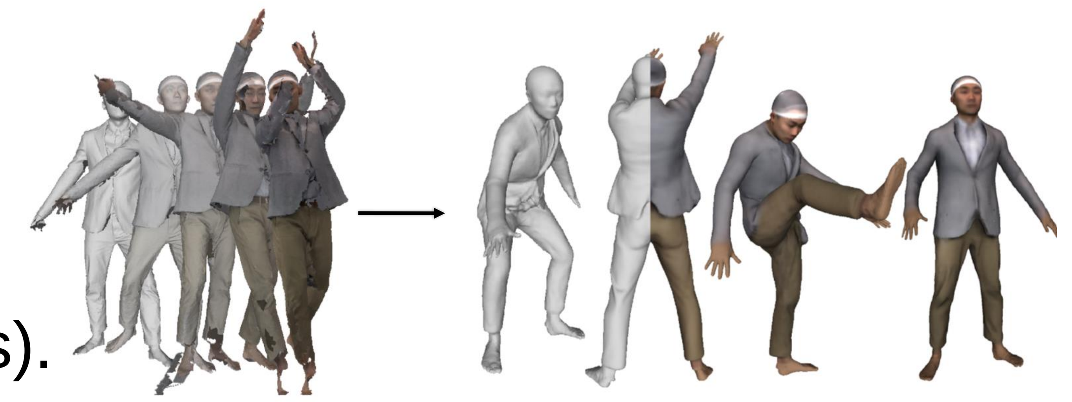
MAX PLANCK INSTITUTE
FOR INTELLIGENT SYSTEMS
ETH zürich

SCANimate: Weakly Supervised Learning of Skinned Clothed Avatar Networks

Shunsuke Saito[1], Jinlong Yang[1], Qianli Ma[1,2], Michael J. Black[1]
[1]Max Planck Institute for Intelligent Systems, [2]ETH Zurich

scanimate.is.tue.mpg.de
CVPR VIRTUAL JUNE 19-25

Goal



- Effortlessly build avatars driven by SMPL pose parameters with realistic clothing deformation.
- Existing approaches rely on template mesh registration [1] and/or physics-based simulation [2], which limits the scalability of clothed avatar modeling (e.g., clothing types, realistic deformations).

Contribution: the first framework to build a high-quality parametric clothed human model (scanimats) from raw scans without garment-specific templates or mesh registration.

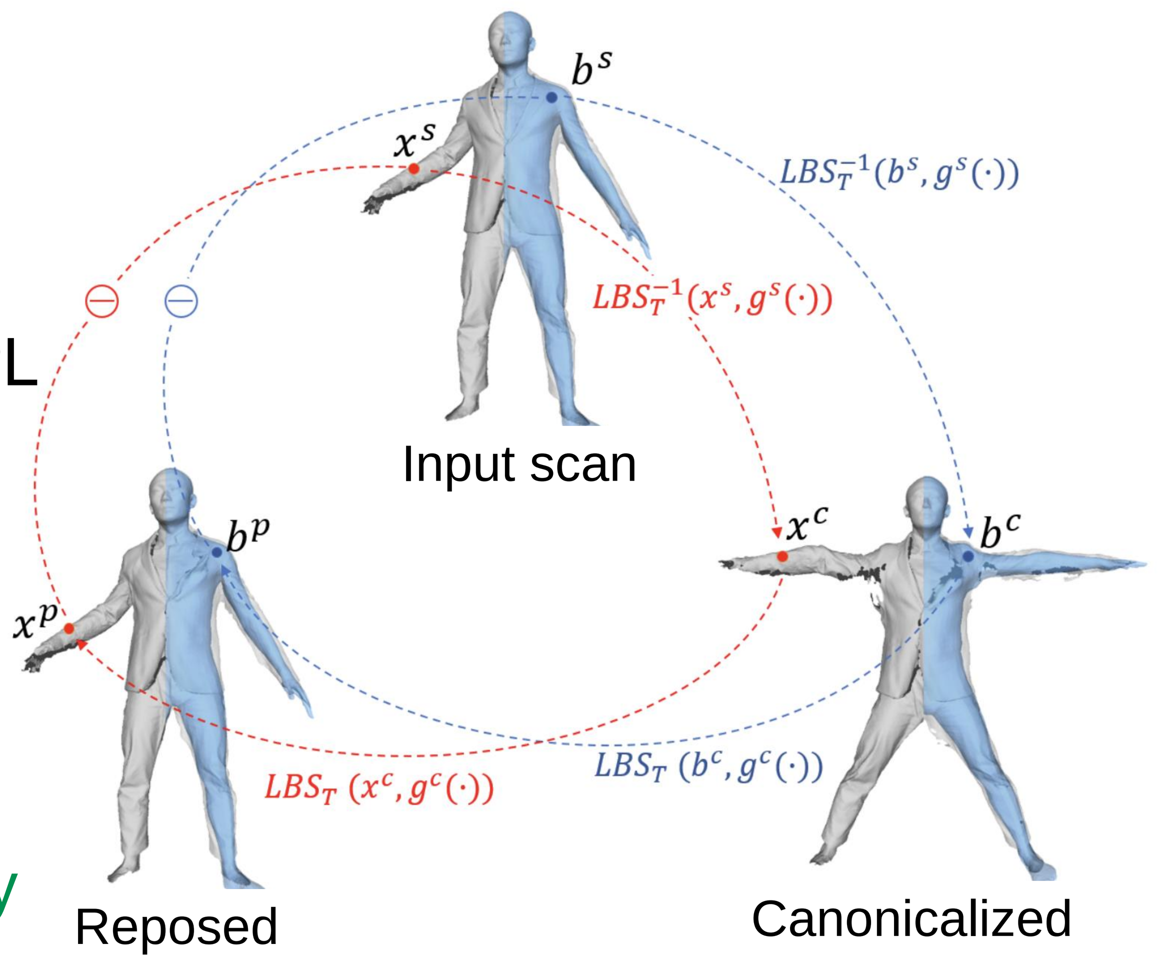
Our Approach

Weakly Supervised Canonicalization

- No ground-truth canonicalized scans
→ Weakly supervised learning with fitted SMPL

1. Propagating skinning weights on SMPL
2. Geometric cycle-consistency

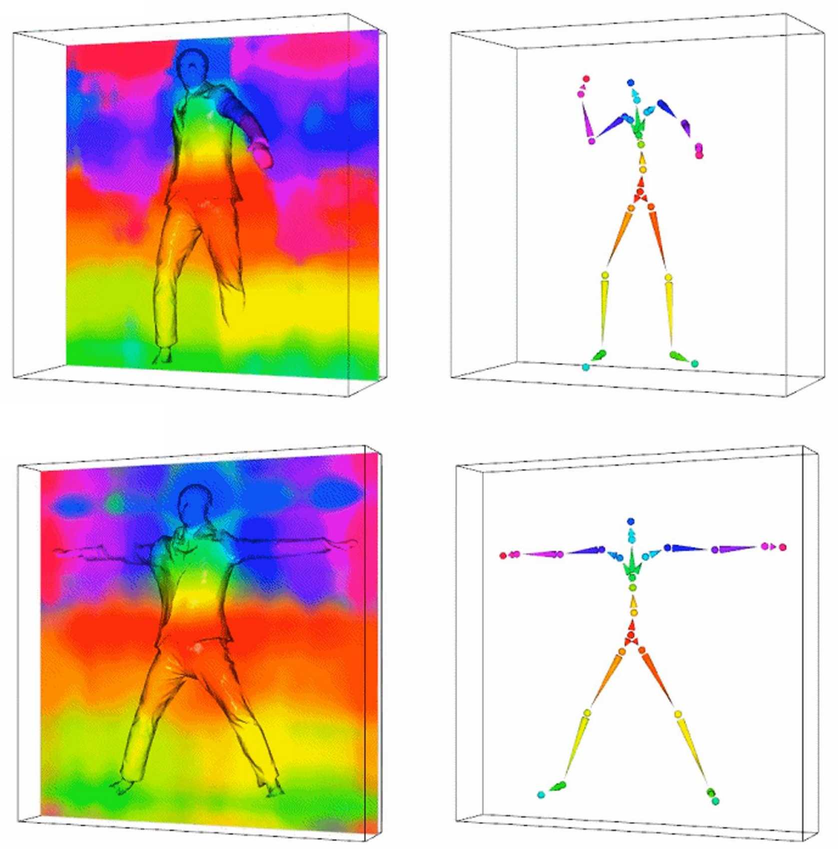
$$E_{cano}(\Theta_1, \Theta_2, \{z_i^s\}) = \overbrace{\sum_i (\underbrace{\lambda_B E_B + \lambda_S E_S}_{\text{SMPL-guided}} + \underbrace{E_C}_{\text{Cycle-consistency}} + E_R)}^{\text{Regularization}}$$



Implicit Skinning Fields

$$\boldsymbol{X}_i^p = LBS_{\mathbf{T}_i}(\boldsymbol{X}_i^c, \boldsymbol{w}(\boldsymbol{X}_i^c)) = (\sum w_j \mathbf{T}_{i,j})\boldsymbol{X}_i^c$$

$$\boldsymbol{X}_i^c = LBS_{\mathbf{T}_i}^{-1}(\boldsymbol{X}_i^s, \boldsymbol{w}(\boldsymbol{X}_i^s)) = (\sum w_j \mathbf{T}_{i,j})^{-1}\boldsymbol{X}_i^s$$



- Continuous extension of linear blend skinning (LBS)
  - Agnostic to underlying topology

$$\boldsymbol{w}(\boldsymbol{x}_i^c) = \boxed{g_{\Theta_1}^c}(\boldsymbol{x}_i^c) : \mathbb{R}^3 \rightarrow \mathbb{R}^J$$

$$\boldsymbol{w}(\boldsymbol{x}_i^s) = \boxed{g_{\Theta_2}^s}(\boldsymbol{x}_i^s, \boldsymbol{z}_i^s) : \mathbb{R}^3 \times \mathbb{R}^{\mathcal{Z}_s} \rightarrow \mathbb{R}^J$$

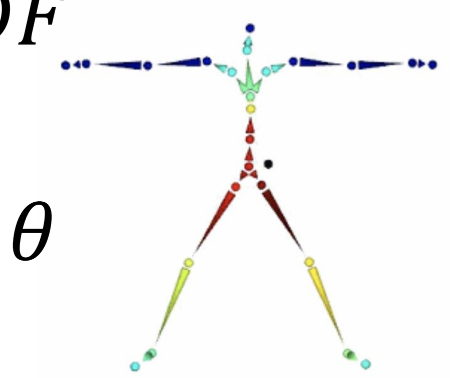
MLPs

Locally Pose-aware Shape Modeling

- Globally pose-conditioned Implicit Surface:

$$f(x, \theta) \rightarrow SDF$$

is prone to overfitting [3].

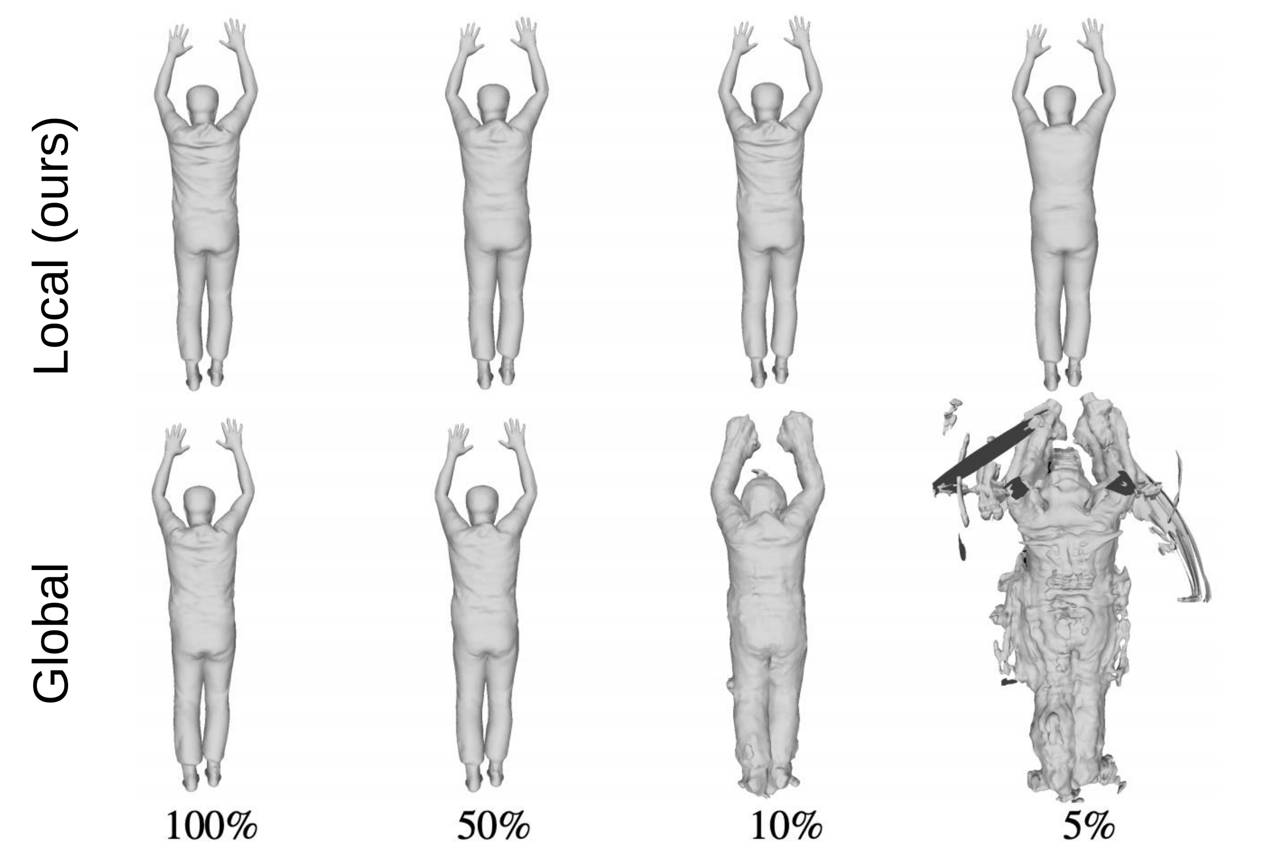


→ We localize pose encoding θ by the learned LBS.

Locally pose-conditioned Implicit Surface

$$f(x, (W \cdot g^c(x)) \circ \theta) \rightarrow SDF$$

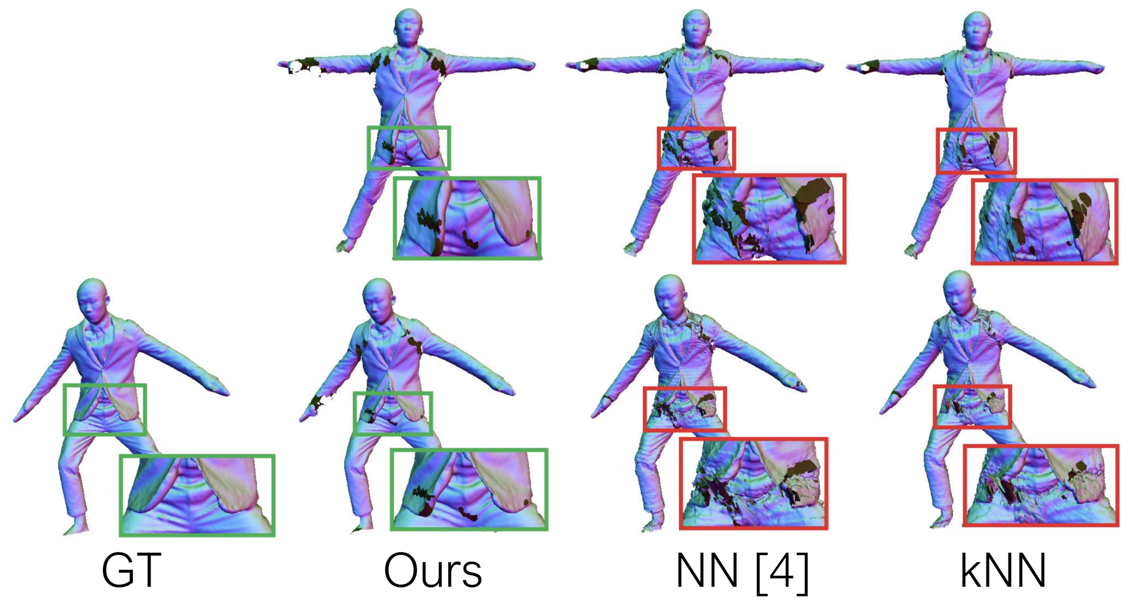
$W \in \mathbb{R}^{J \times J}$: Joint association matrix

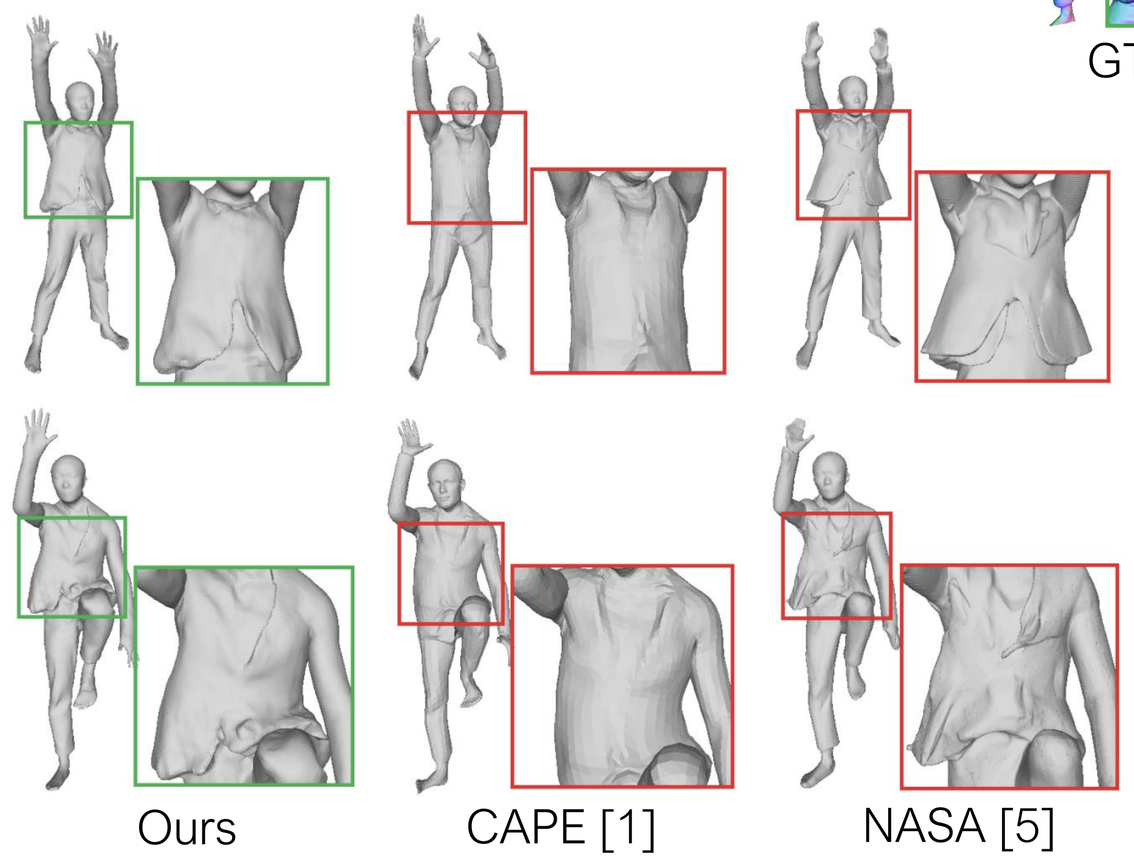


Experiments

- Quantitative and qualitative evaluation on CAPE dataset [1]

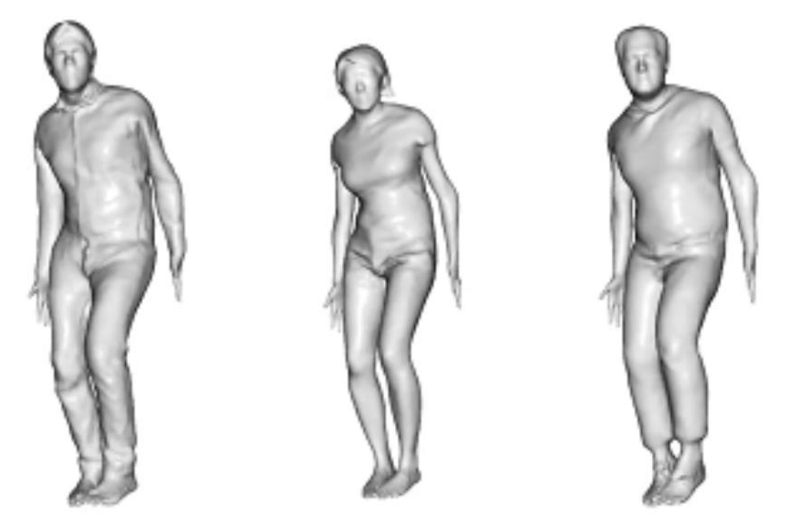
		Ours	N [4]]	KNN
Int.	$D_{s2m} \downarrow$	**0.570**	1.25	1.25
	$D_n \downarrow$	**0.253**	0.301	0.299
	$P_i \uparrow$	**0.5**	0.374	0.396
	$P_v \uparrow$	**0.5**	0.435	0.431
Ex.	$P_i \uparrow$	**0.5**	0.262	0.312
	$P_v \uparrow$	**0.5**	0.392	0.449





		Ours	CAPE [1]	NASA [5]
Int.	$D_{s2m} \downarrow$	**0.570**	0.970	1.12
	$D_n \downarrow$	**0.253**	0.308	0.289
	$P_i \uparrow$	**0.5**	0.268	0.432
	$P_v \uparrow$	**0.5**	0.455	0.457
Ex.	$P_i \uparrow$	**0.5**	0.214	0.343
	$P_v \uparrow$	**0.5**	0.422	0.395

Learned scanimats



Textured scanimats

References

[1] Learning to Dress 3D People in Generative Clothing, Ma et al., CVPR 2020
[2] TailorNet: Predicting Clothing in 3D as a Function of Human Pose, Shape and Garment Style, Patel et al., CVPR 2020
[3] STAR: Sparse Trained Articulated Human Body Regressor, Osman et al., ECCV 2020
[4] ARCH: Animatable Reconstruction of Clothed Humans, Huang et al., CVPR 2020
[5] NASA: Neural Articulated Shape Approximation, Deng et al., ECCV 2020