High-Fidelity MR-Me: Lightweight Capture of Personalized Neural Animatable Avatars

Chen Guo

Motivation

Immersive Telepresence [TM & © Lucasfilm Ltd.]

NVIDIA

- High requirement of devices
- Tons of manual efforts

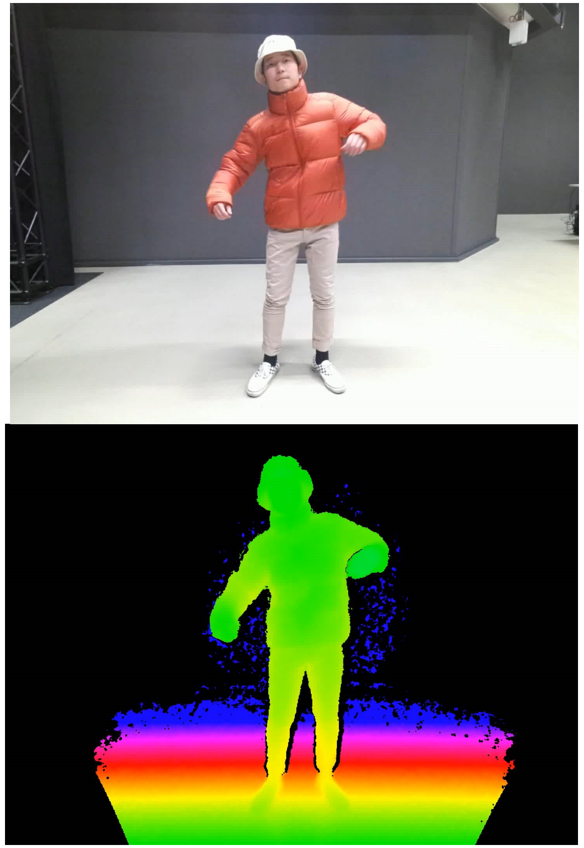
PINA: Learning a Personalized Implicit Neural Avatar from a Single RGB-D Video Sequence

Z. Dong*, C. Guo*, J. Song, X. Chen, A. Geiger, O. Hilliges

CVPR 2022

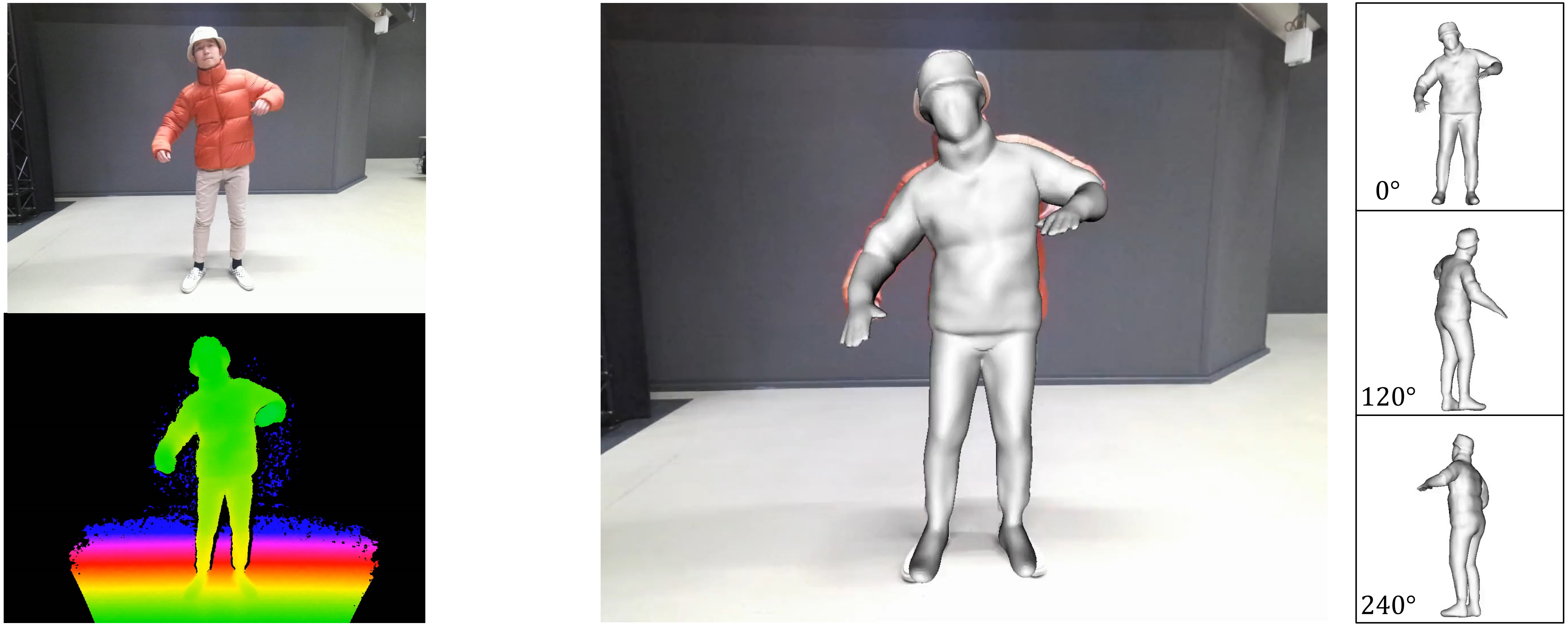
Problem Definition

Problem Definition



Input: RGB-D sequence

Problem Definition

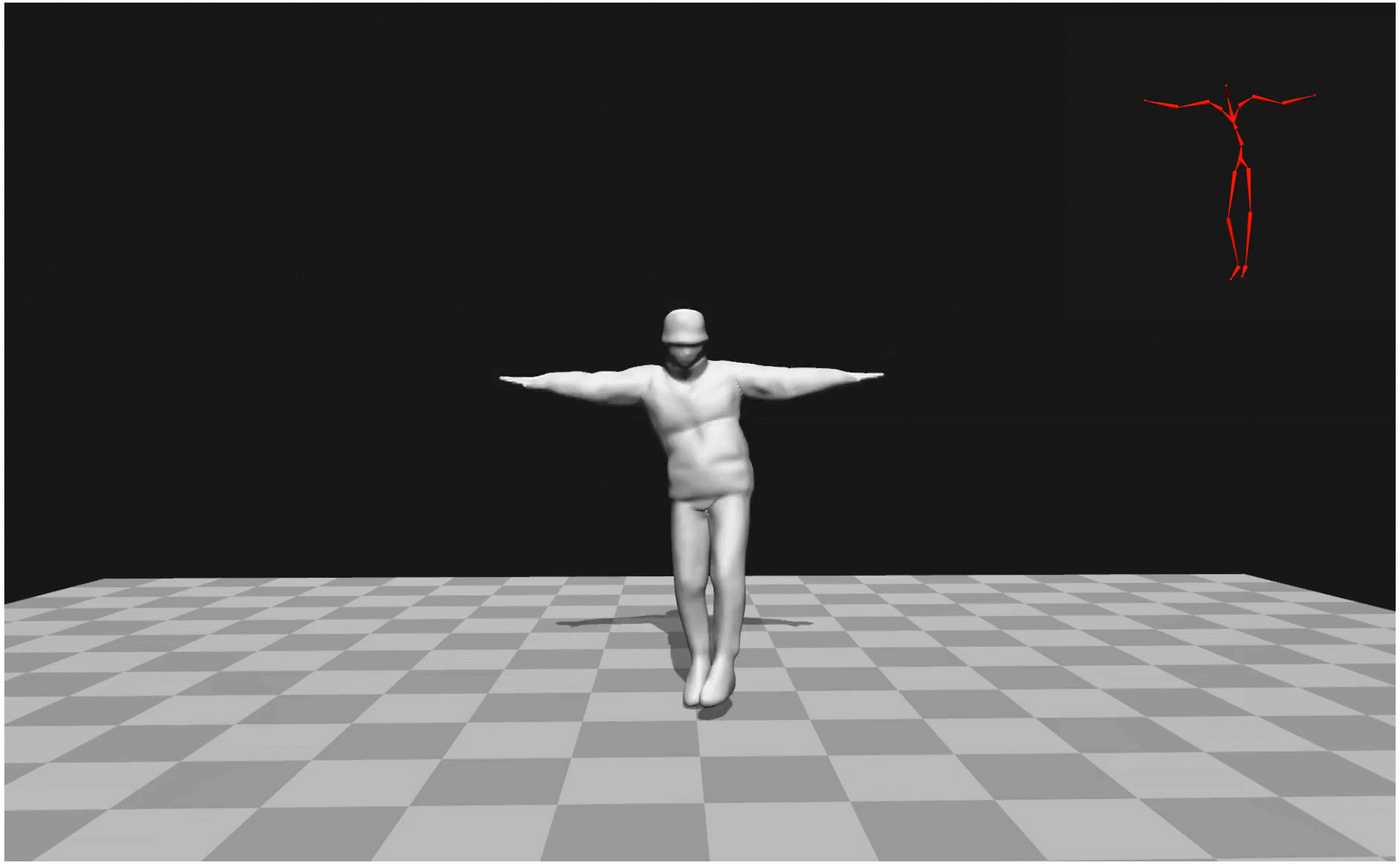


Input: RGB-D sequence

Reconstruction

Side View

Problem Definition

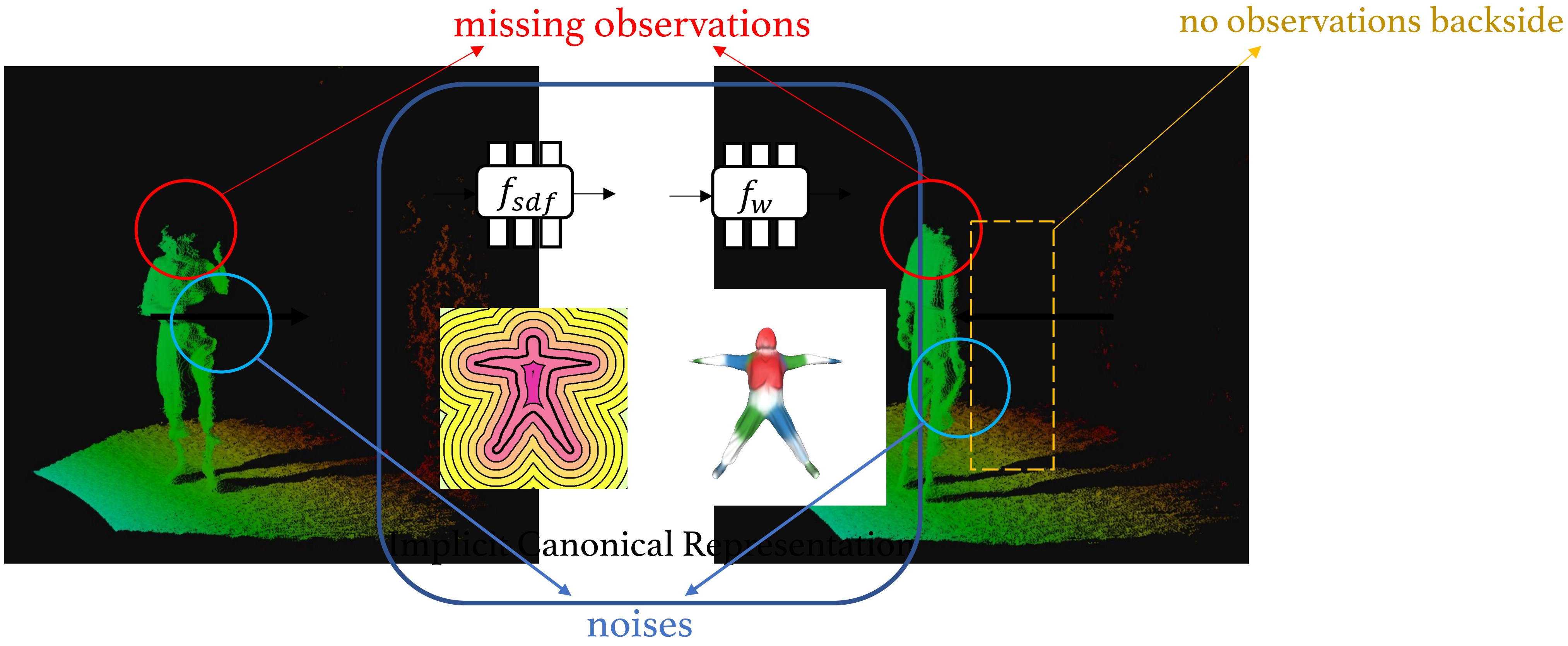


Animation

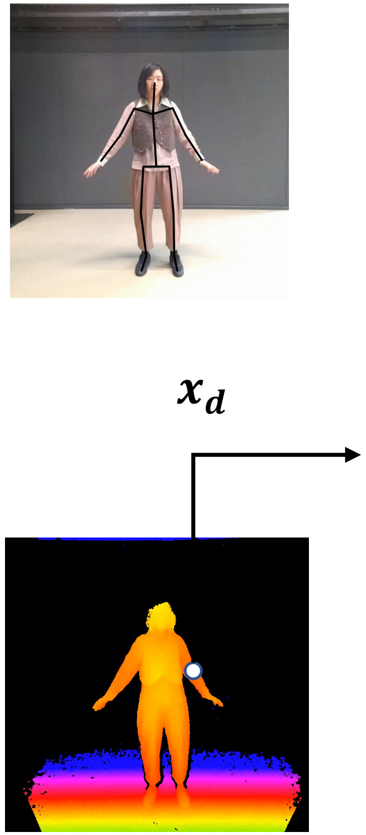
Method

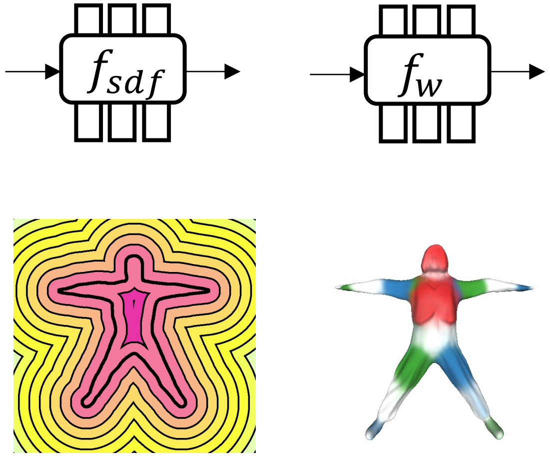
Challenge: raw depth frames are noisy and incomplete

Idea: globally fuse partial depth maps into a single consistent representation



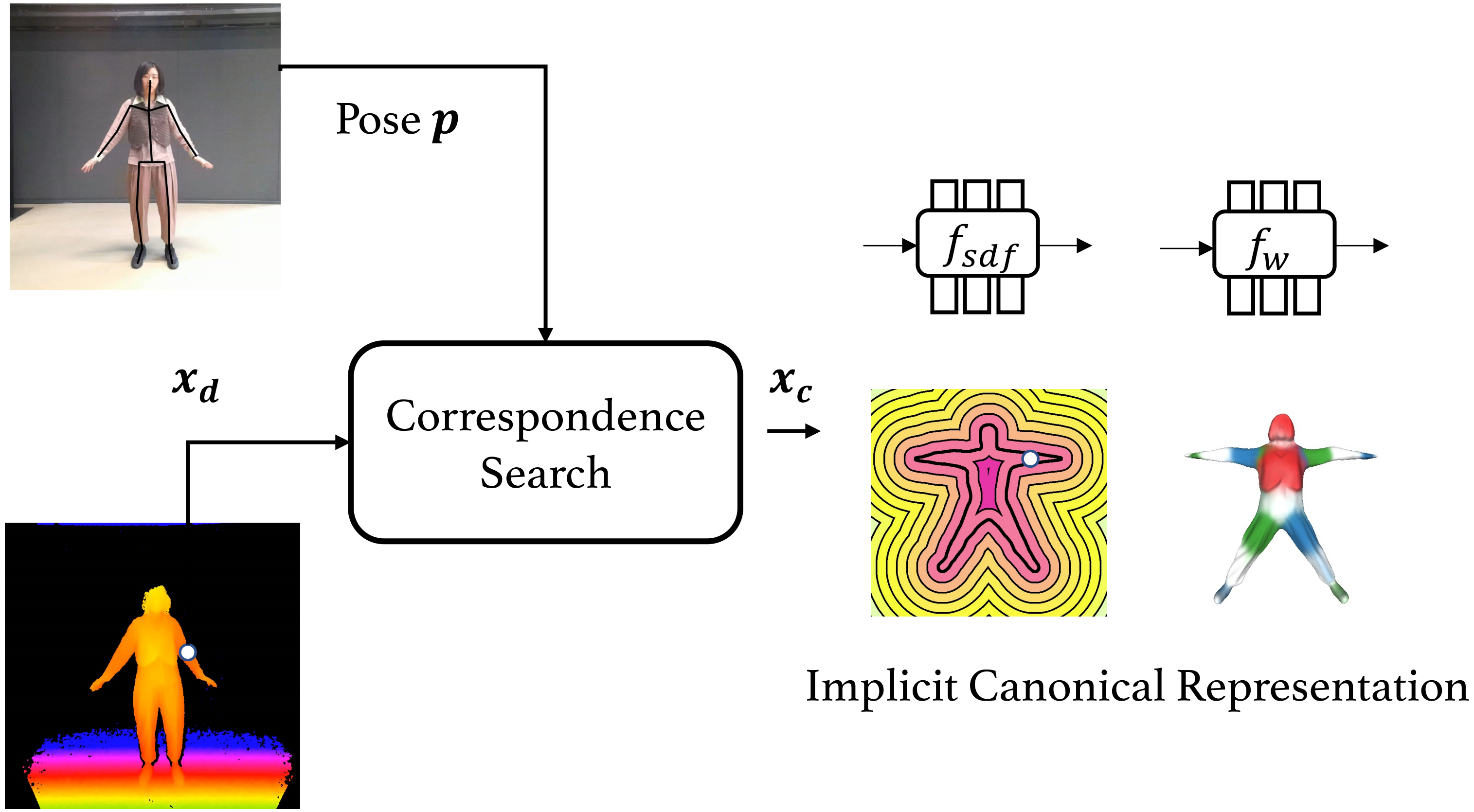
Method



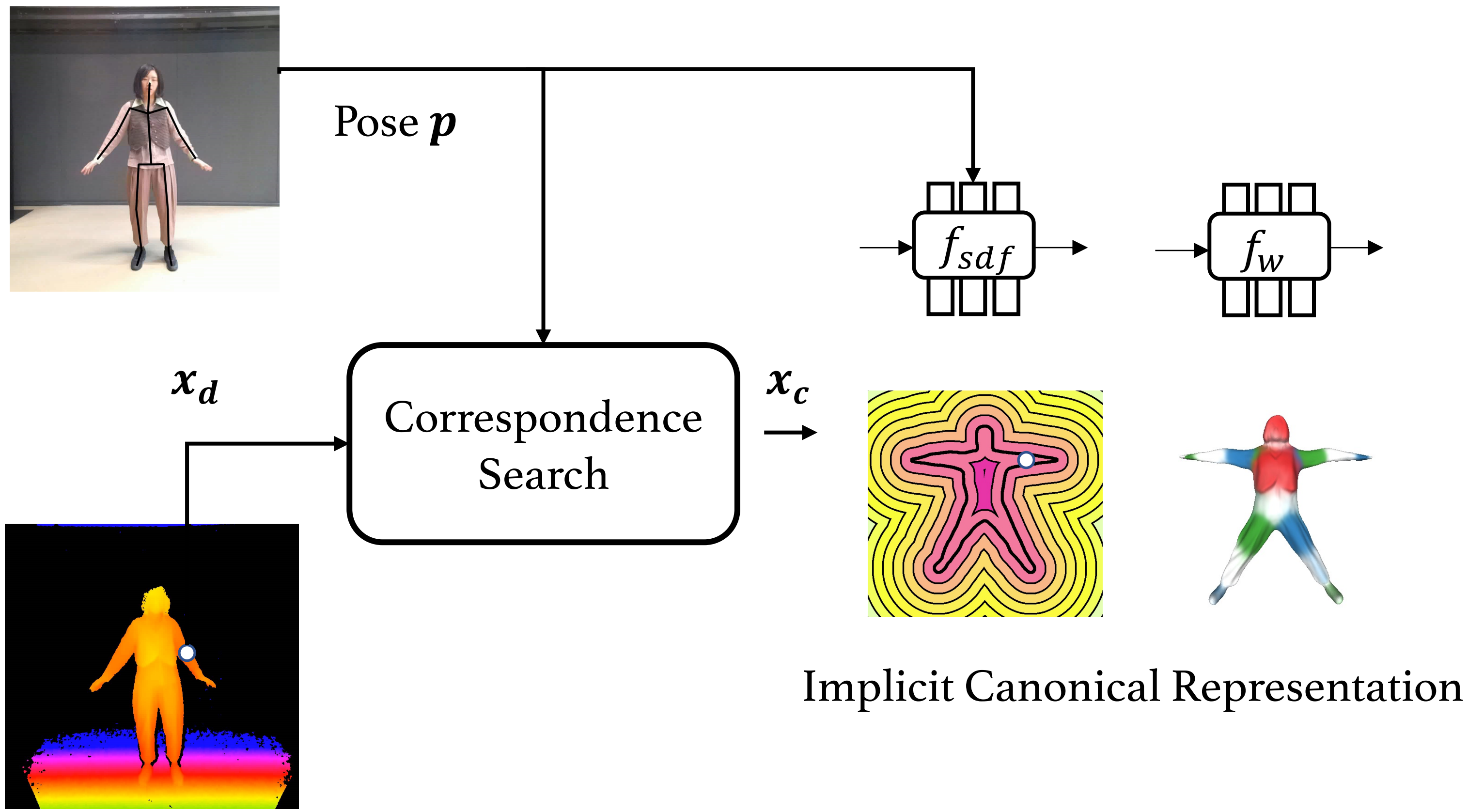


Implicit Canonical Representation

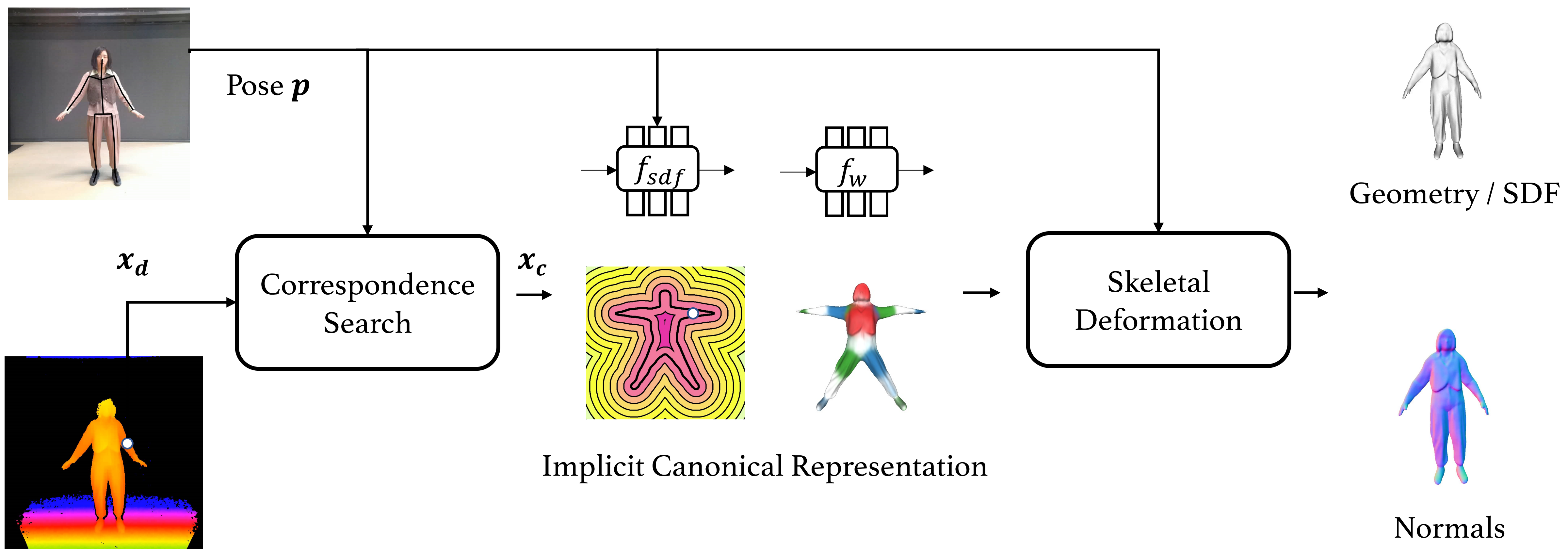
Method


Pose $\boldsymbol{p}$
$\boldsymbol{x_d}$
Correspondence Search
$\boldsymbol{x_c}$
f_{sdf}
f_w
Implicit Canonical Representation

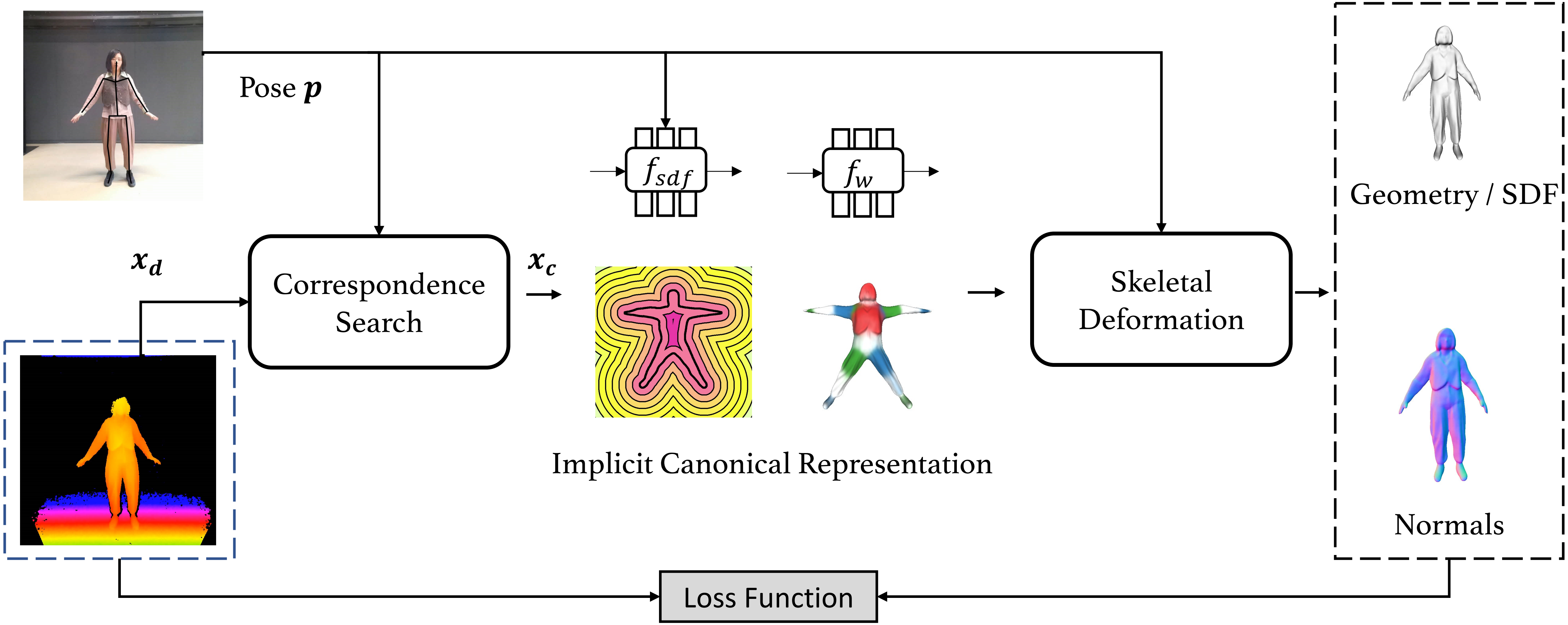
Method


Pose $\boldsymbol{p}$
$\boldsymbol{x_d}$
Correspondence Search
$\boldsymbol{x_c}$
f_{sdf}
f_w
Implicit Canonical Representation

Method

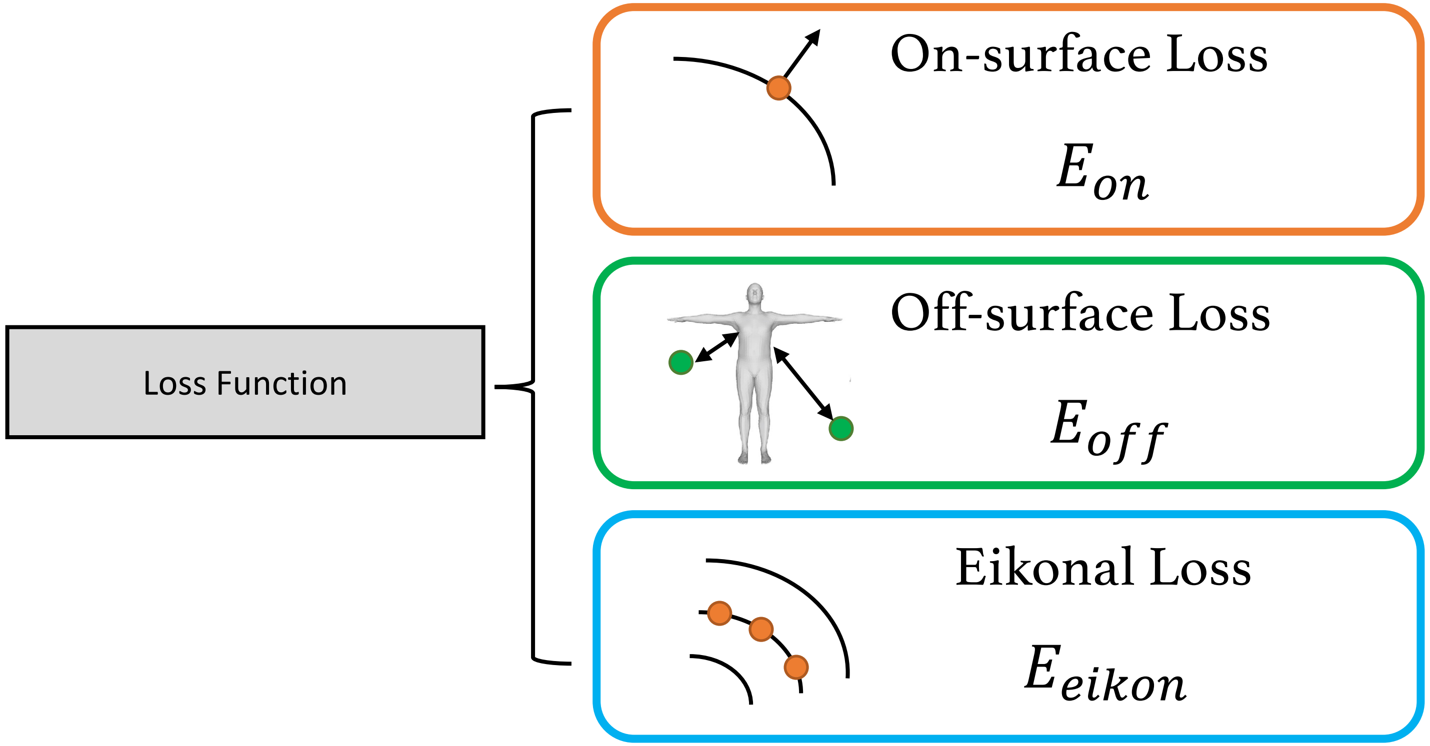

Pose $\boldsymbol{p}$
$\boldsymbol{x_d}$
Correspondence Search
$\boldsymbol{x_c}$
f_{sdf}
f_w
Implicit Canonical Representation
Skeletal Deformation
Geometry / SDF
Normals

Method



Pose $\boldsymbol{p}$
$\boldsymbol{x_d}$
Correspondence Search
$\boldsymbol{x_c}$
f_{sdf}
f_w
Implicit Canonical Representation
Skeletal Deformation
Geometry / SDF
Normals
Loss Function

Method – Loss



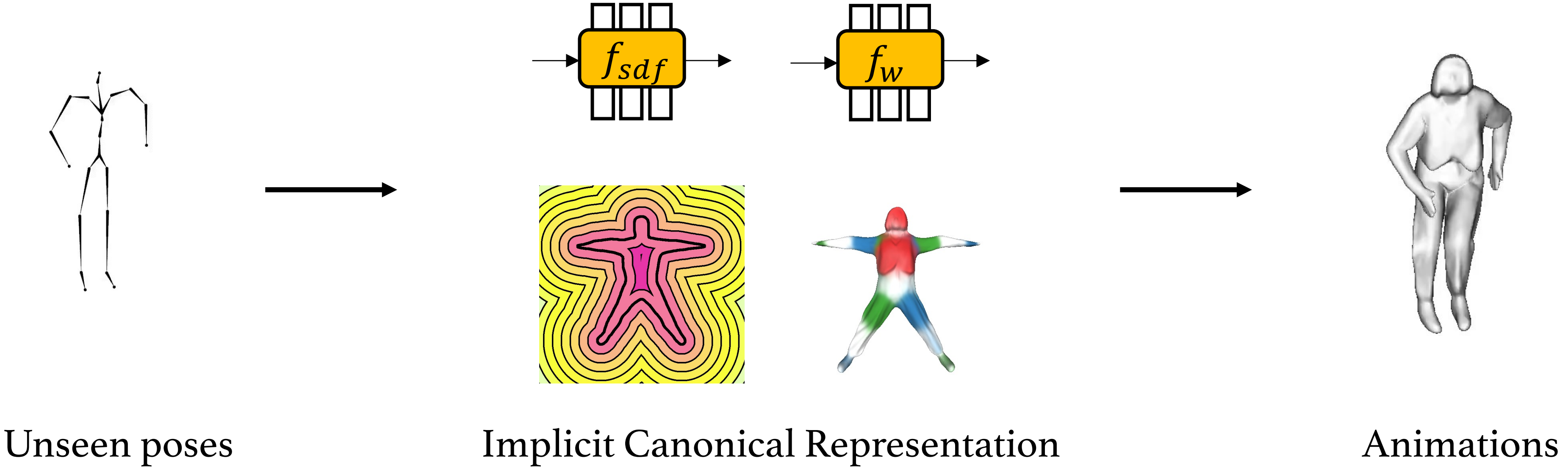
$$E_{sdf} = \sum_{\mathbf{x}_d \in \mathcal{P}^i_{\text{on}}} |SDF(\mathbf{x}_d)|$$

$$E_{normal} = \sum_{\mathbf{x}_d \in \mathcal{P}^i_{\text{on}}} |NC(\mathbf{x}_d)|$$

$$E_{off} = \sum_{\mathbf{x}_d \in \mathcal{P}^i_{\text{off}}} |SDF(\mathbf{x}_d) - SDF_{body}(\mathbf{x}_d)|$$

$$E_{eikon} = \mathbb{E}_{\mathbf{x}_c}(\|\nabla f_{\text{sdf}}(\mathbf{x}_c)\| - 1)^2$$

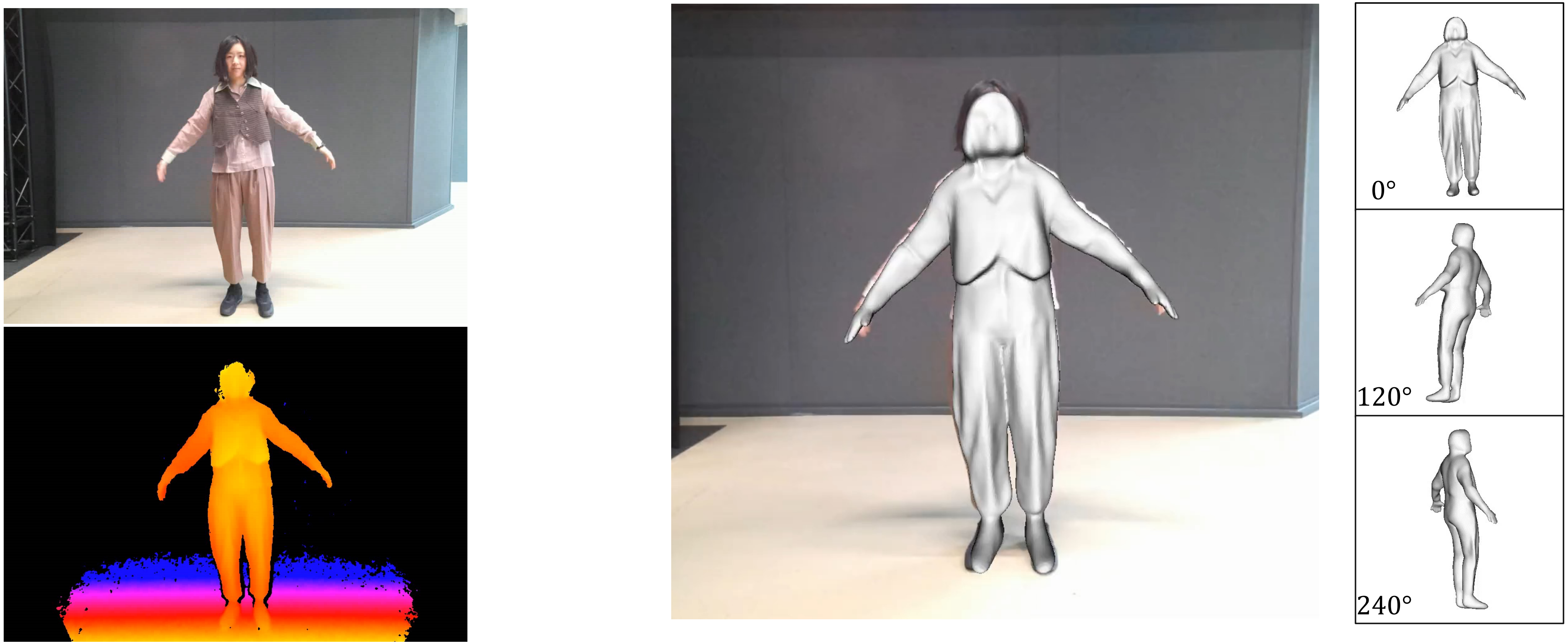
Animation



f_{sdf}
f_w
Unseen poses
Implicit Canonical Representation
Animations

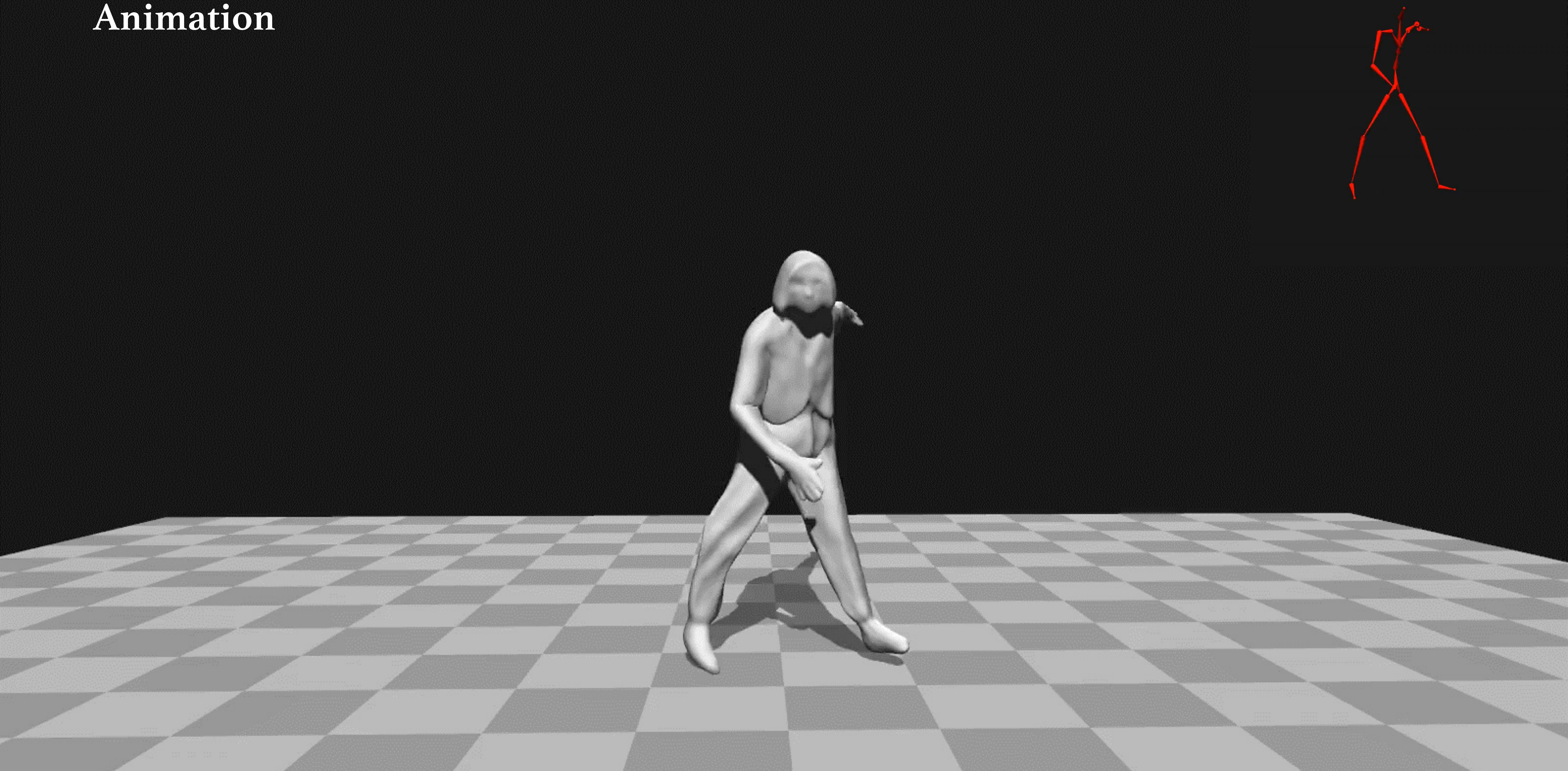
Results on Real-world Data

Subject 1:



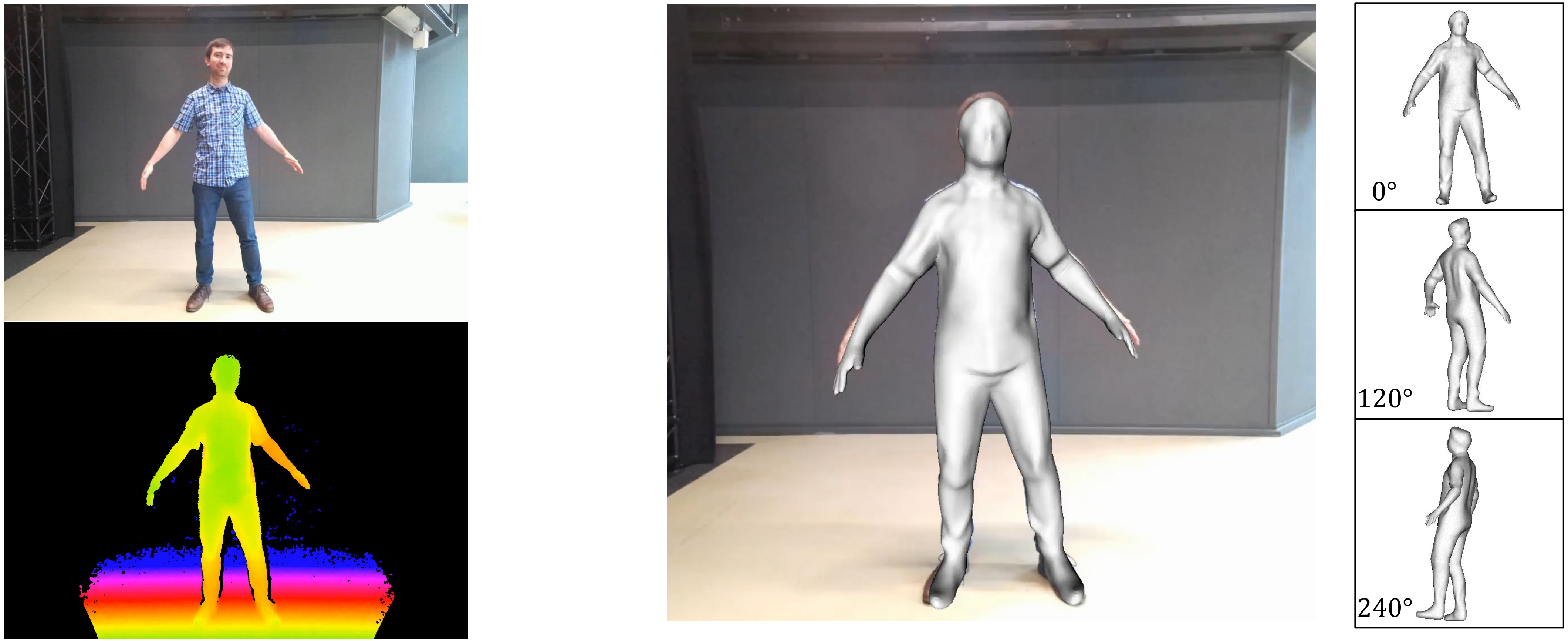
Input Reconstruction Side View

Animation

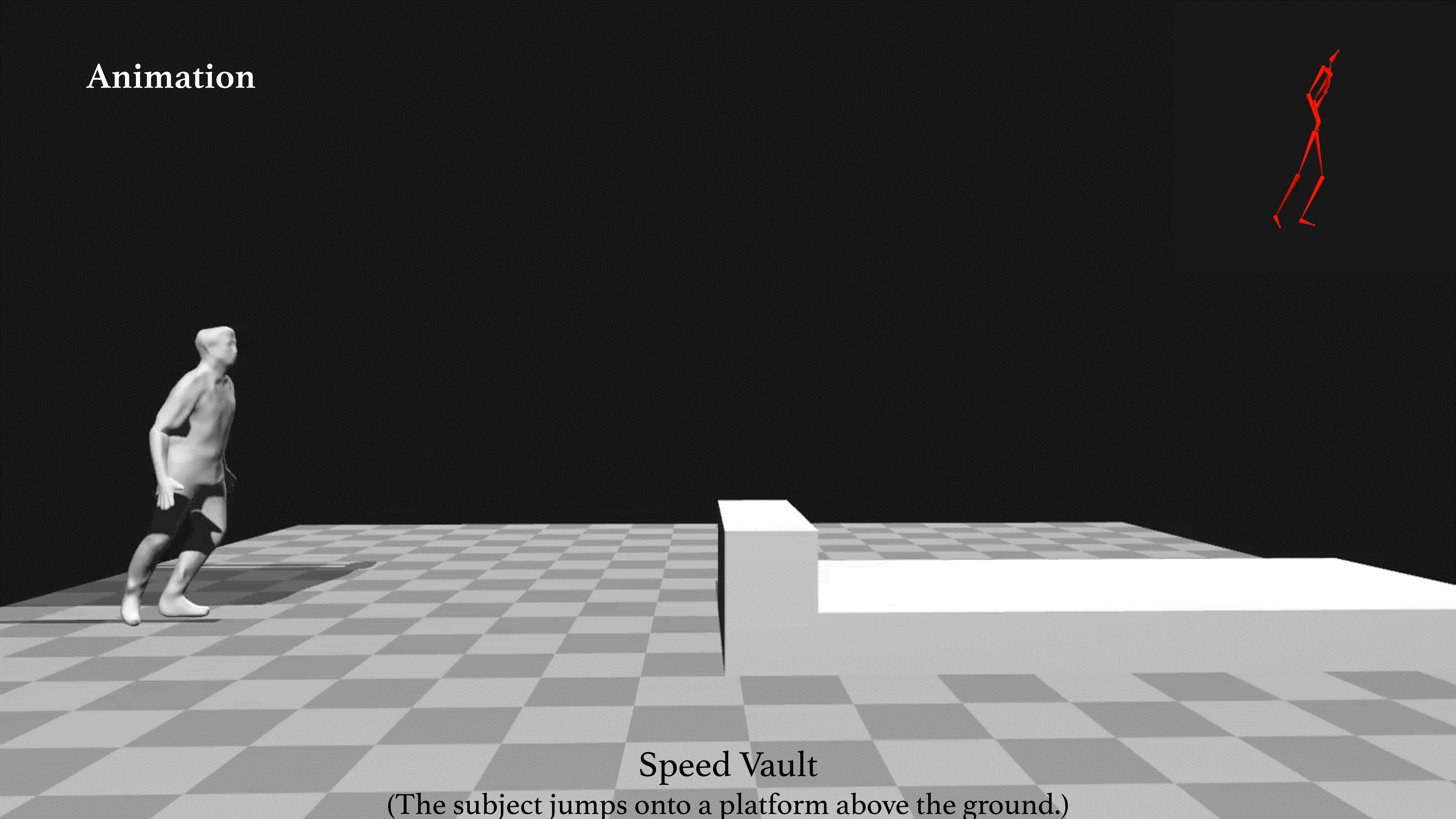


JAZZ Dance

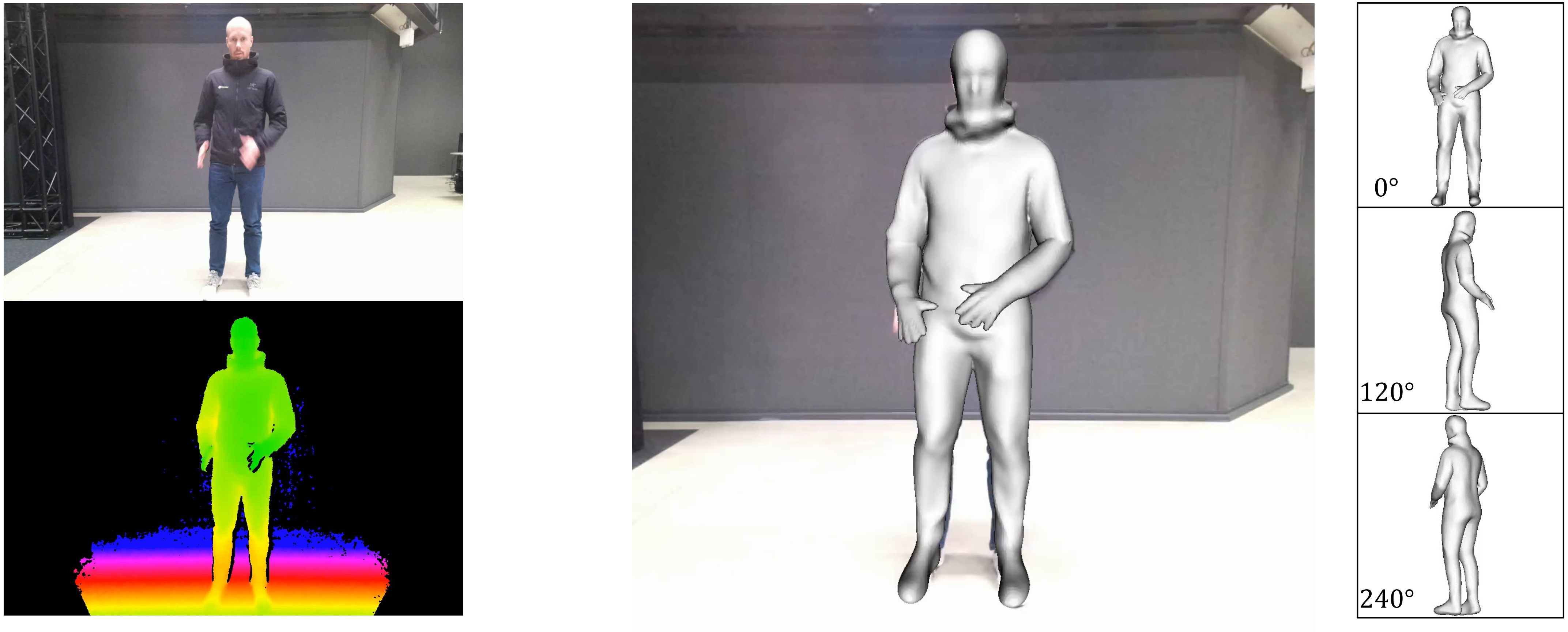
Subject 2:



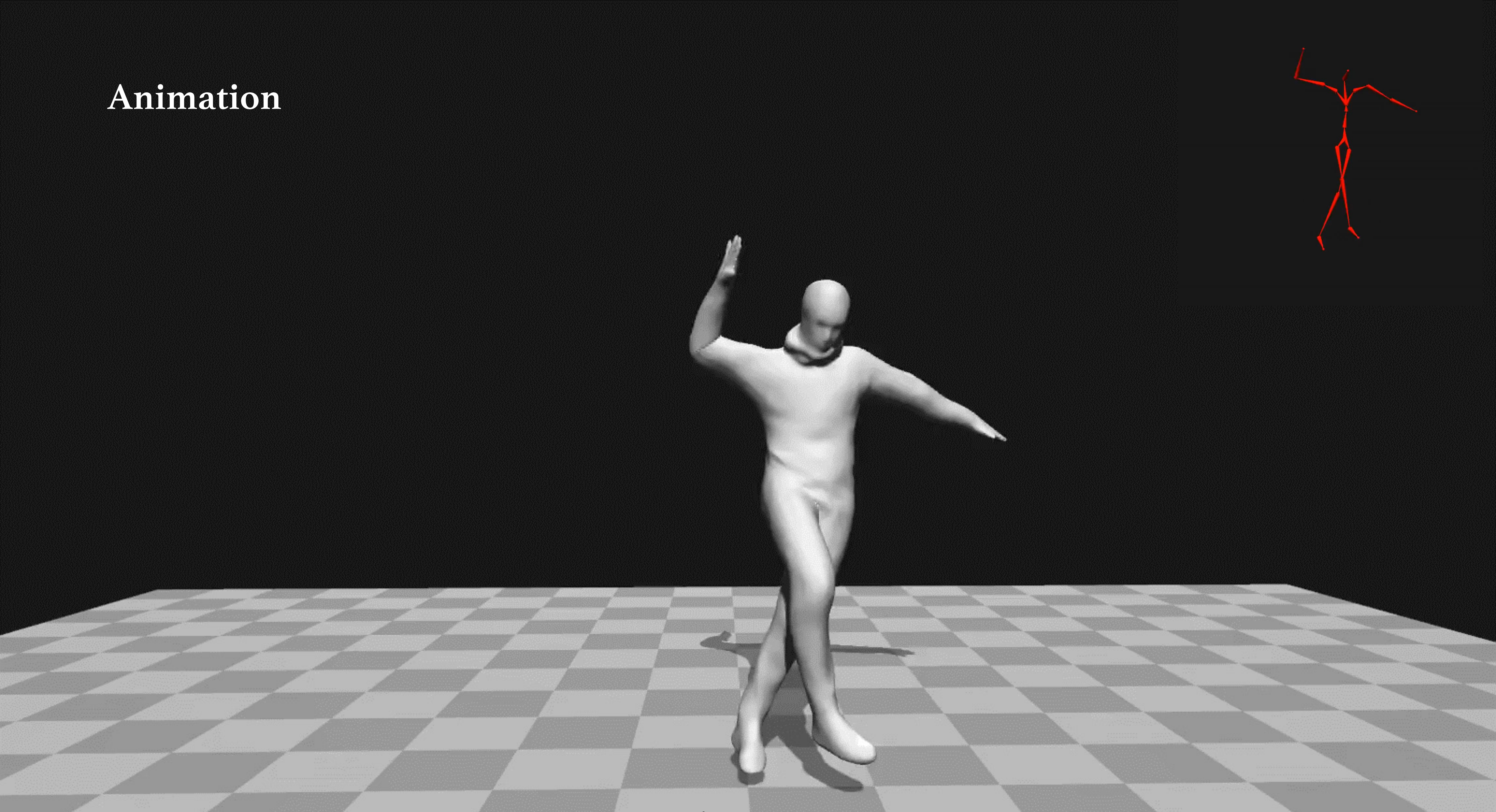
Input Reconstruction Side View


Animation
Speed Vault
(The subject jumps onto a platform above the ground.)

Subject 3



Input Reconstruction Side View



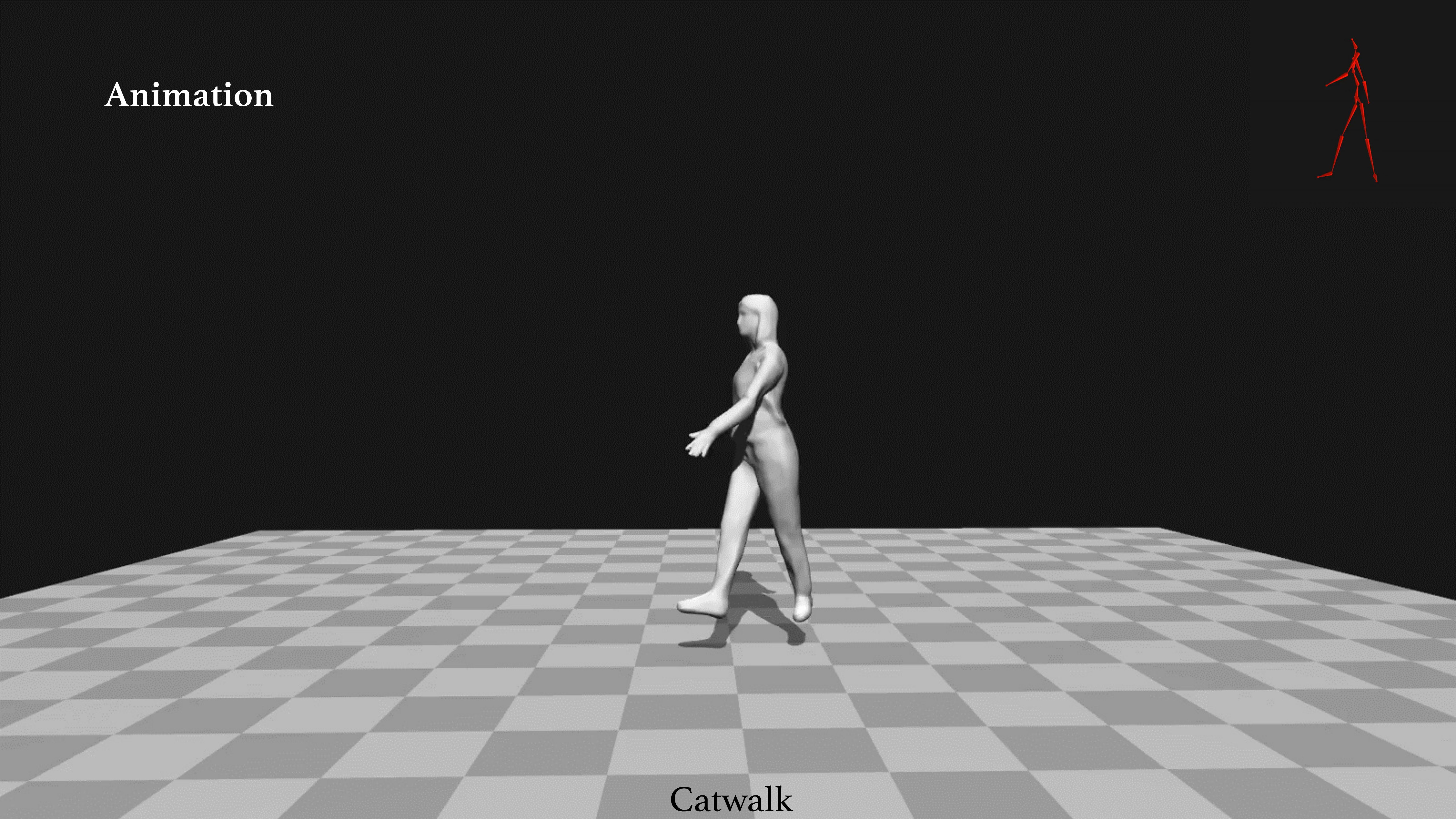
Hip-hop Dance

Subject 4

Input

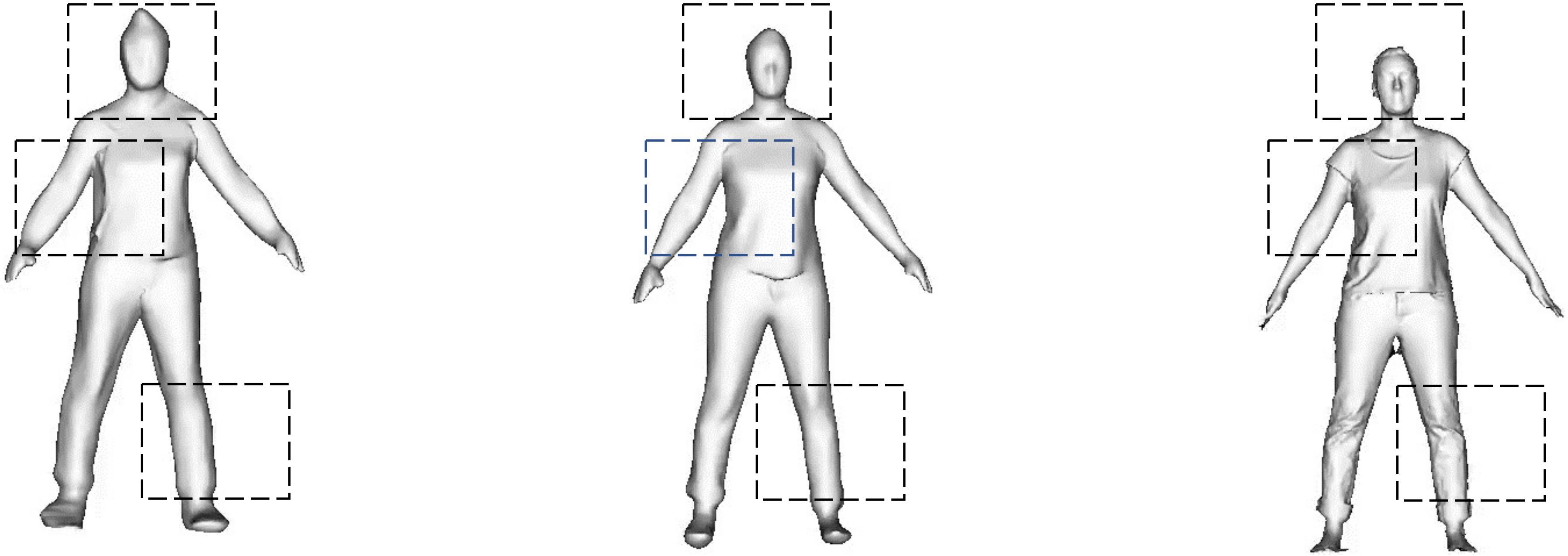
Reconstruction

Side View


Animation
Catwalk

Ablation Studies and Comparisons

Ablation – joint optimization of pose and shape

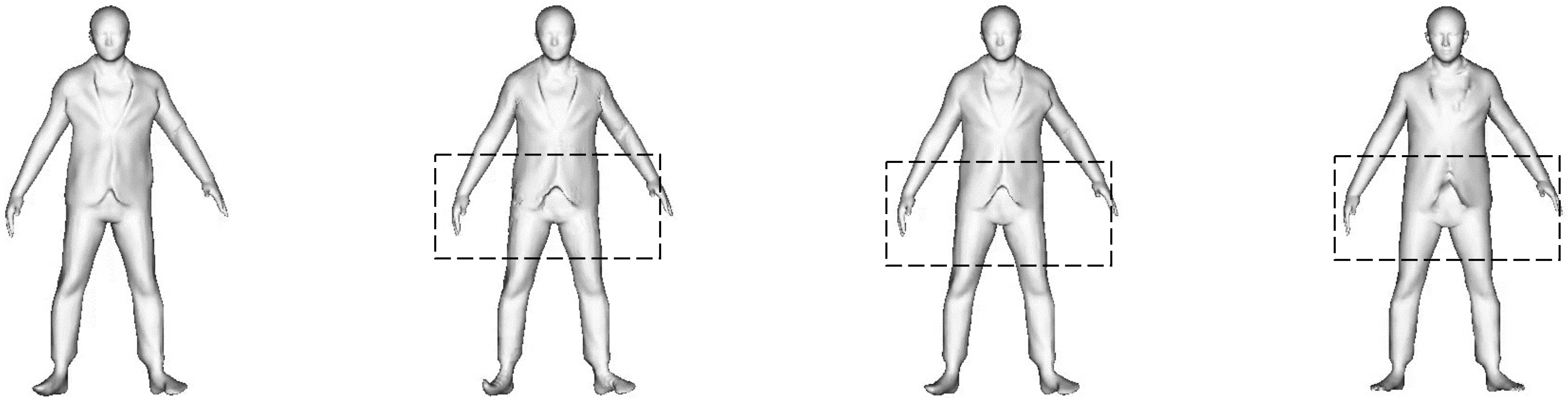


Ours
(w/o pose opt.)

Ours
(complete)

GT

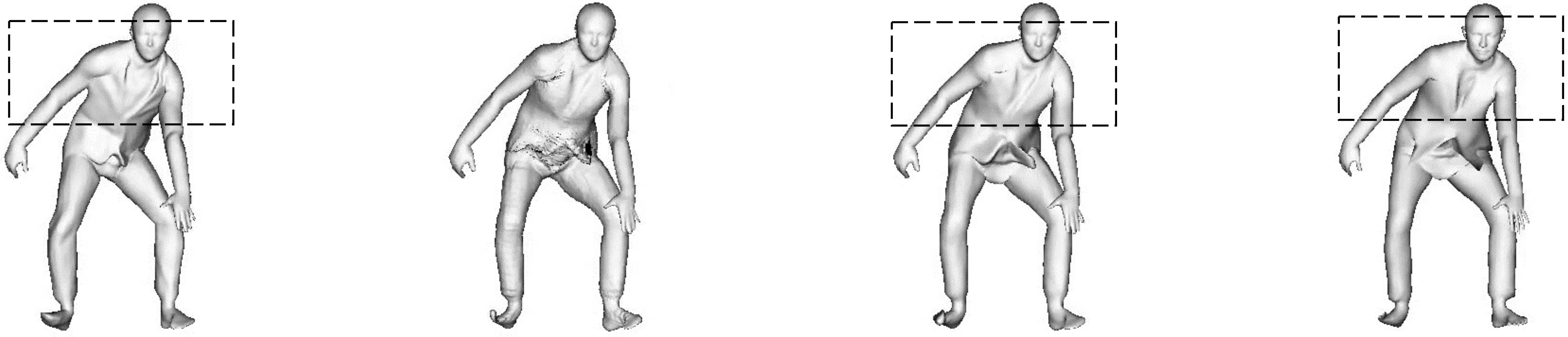
Ablation – deformation model



Ours (w/o pose condition) | Ours (w/ SMPL weights) | Ours (complete) | GT

Without learned skinning weights, the deformed regions can be noisy and display visible artifacts.

Ablation – deformation model



Ours
(w/o pose condition)

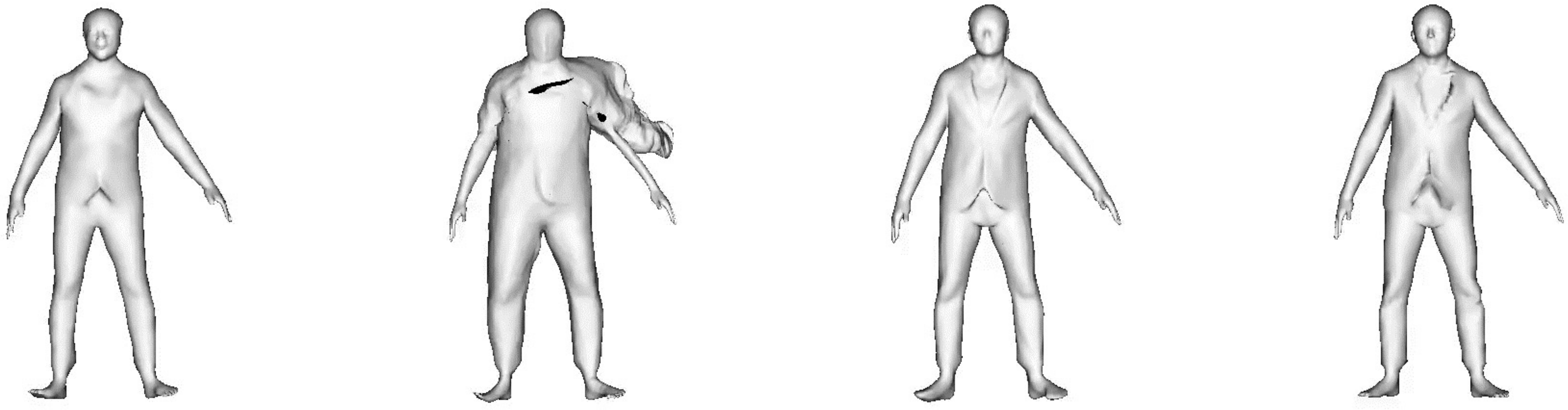
Ours
(w/ SMPL weights)

Ours
(complete)

GT

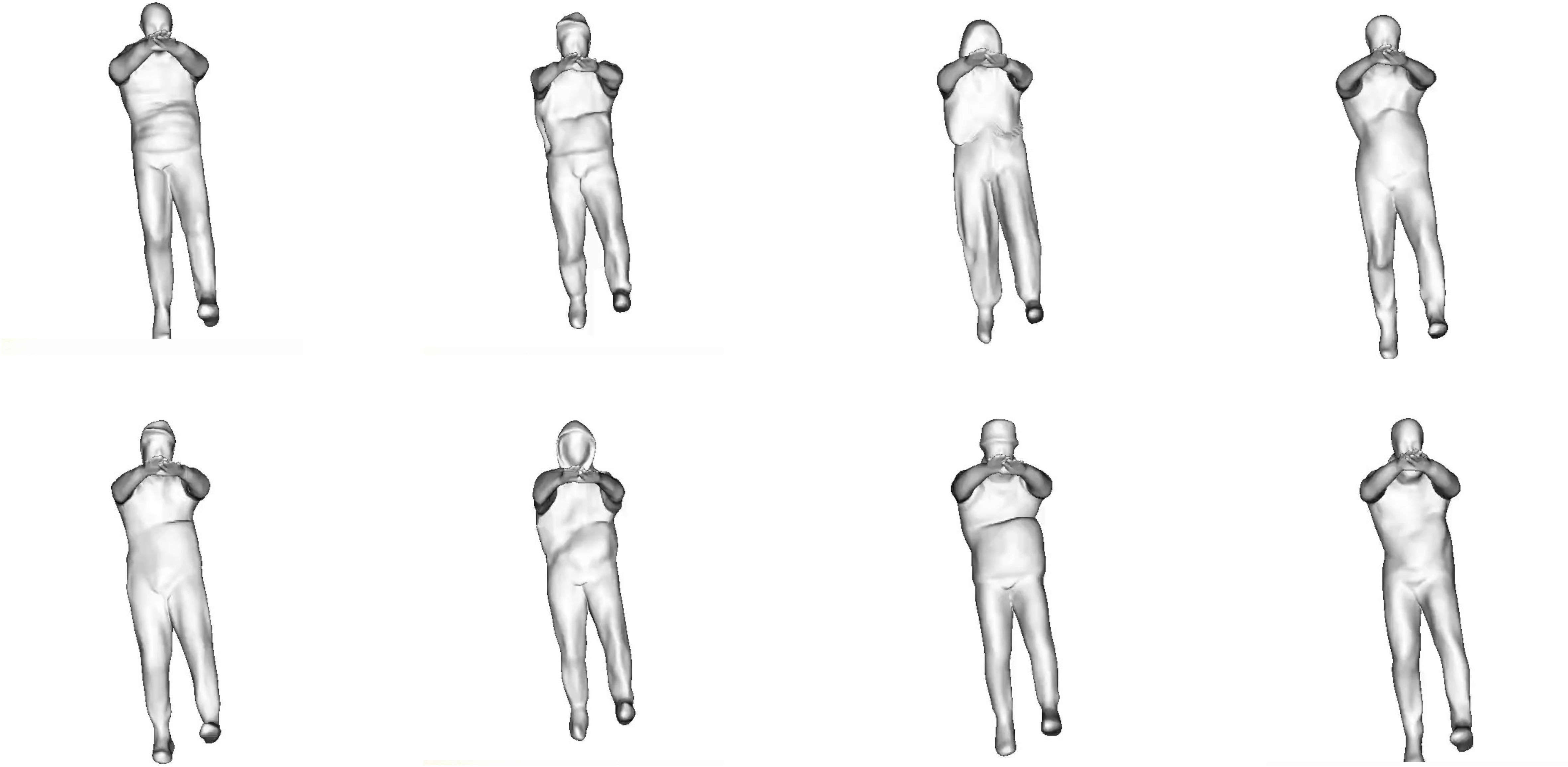
Without pose features, the shape network cannot represent dynamically changing surface details of the blazer.

Comparisons on CAPE (Animation)



IP-Net (3D) | SCANimate (2.5 D) | Ours (2.5 D) | GT

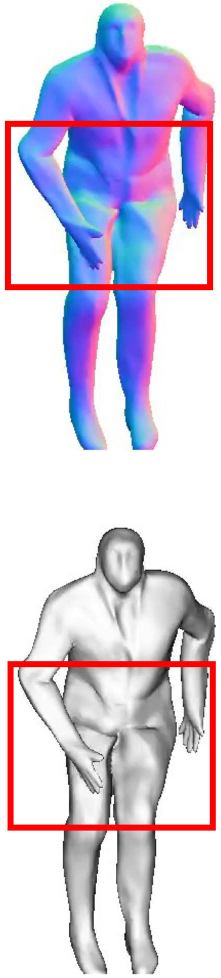
Avatar exhibition



Limitations and Future Research Directions

Limitations – handling loose clothing

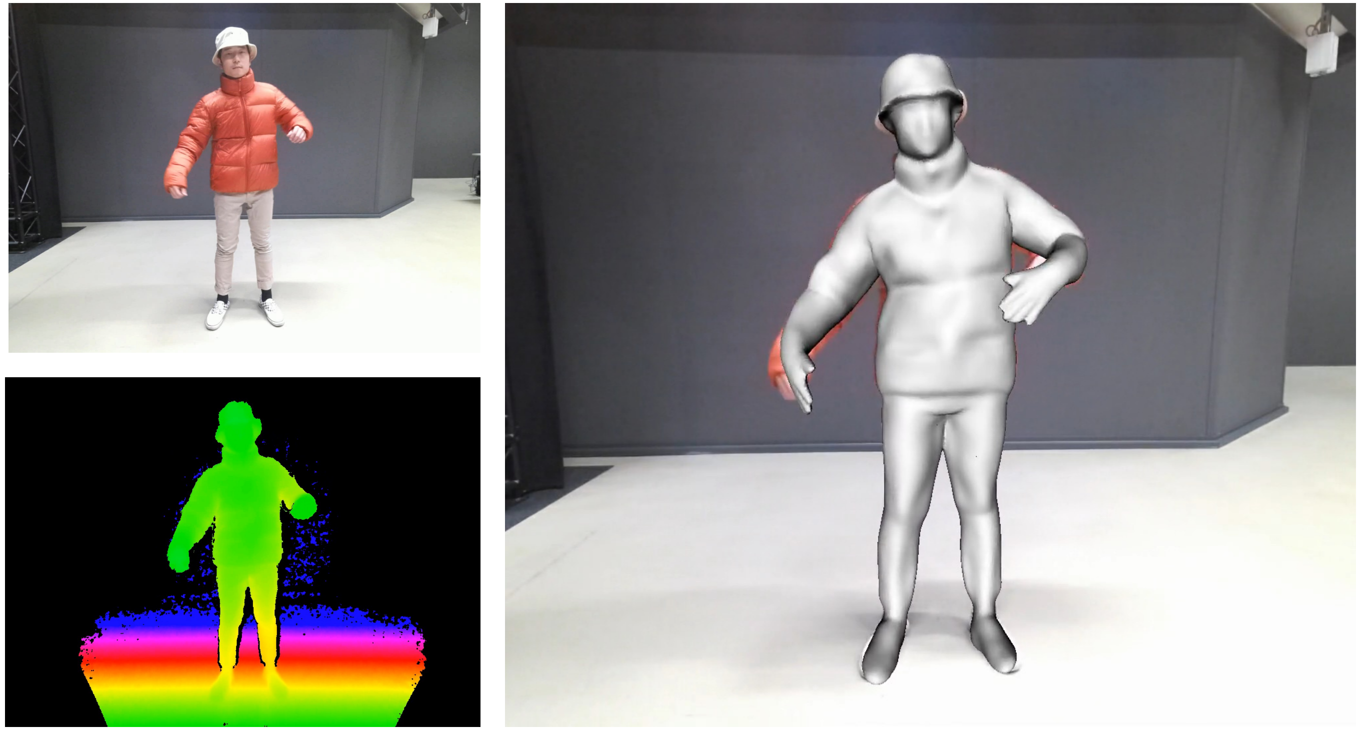
Reference image



Animation

Artifacts occur on this loose clothing parts
(large topological changes).

Limitations – learning textures

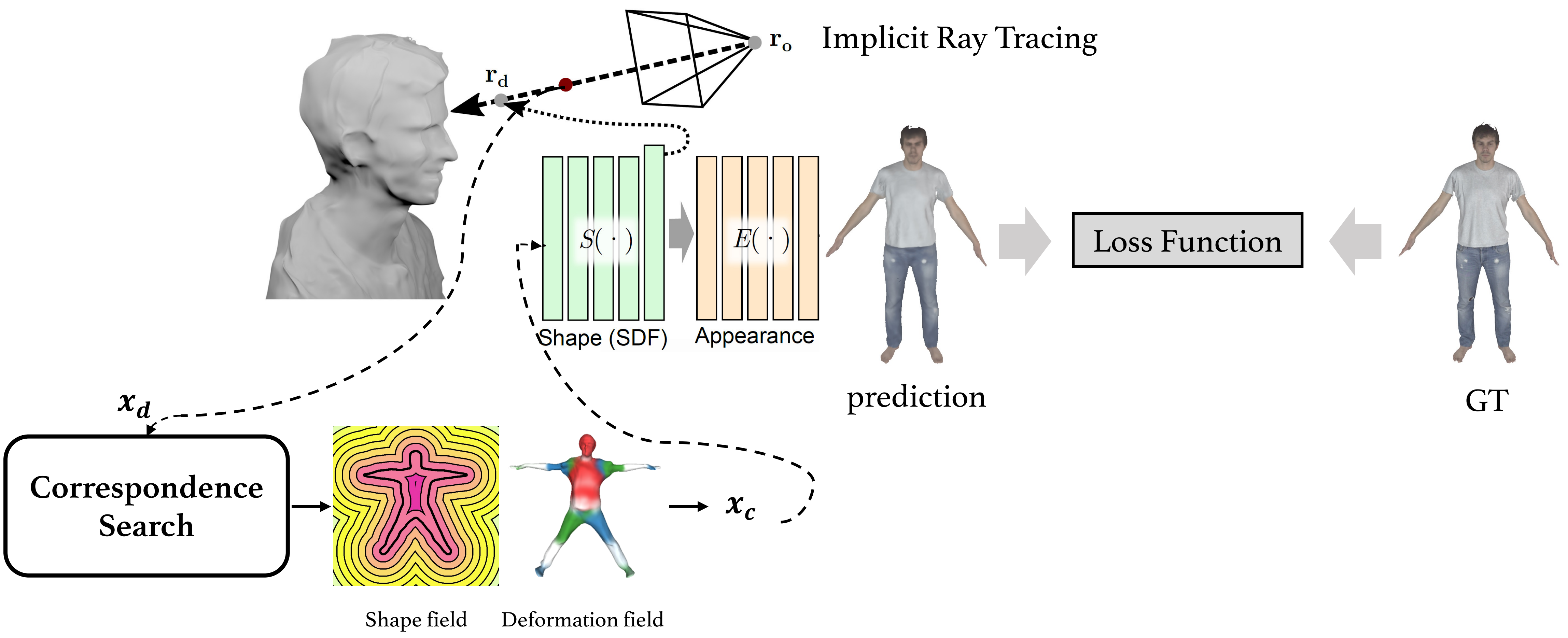


PINA

RenderPeople

RGB-PINA

Learning a Personalized Implicit Neural Avatar from a Single RGB Video Sequence :



RGB-PINA

Learning a Personalized Implicit Neural Avatar from a Single RGB Video Sequence :

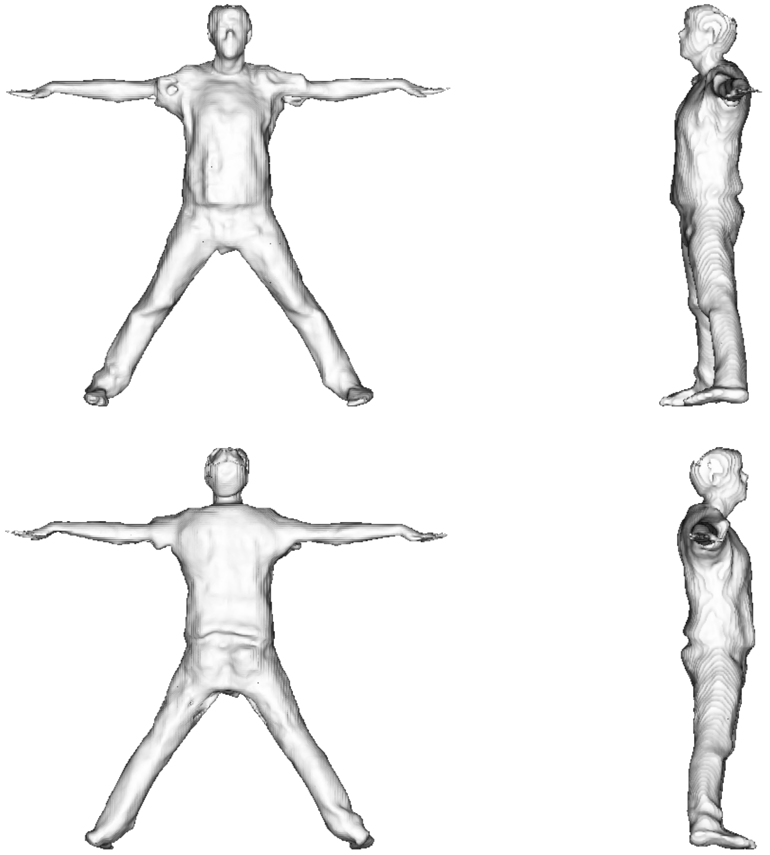
Experiment on synthetic data (rendered):

Input video

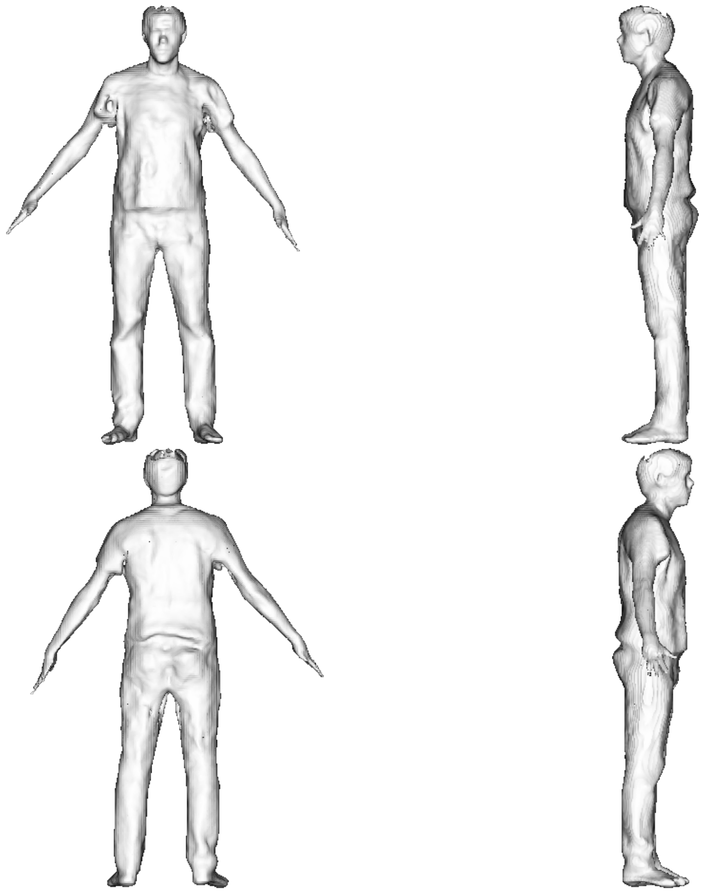
RGB-PINA

Learning a Personalized Implicit Neural Avatar from a Single RGB Video Sequence :

Geometry:



Canonical Space



Posed Space

RGB-PINA

Learning a Personalized Implicit Neural Avatar from a Single RGB Video Sequence :

Texture:

Ours

GT

Thank you!