Digitalization of Clothed Humans with Expressive Behaviors from Videos

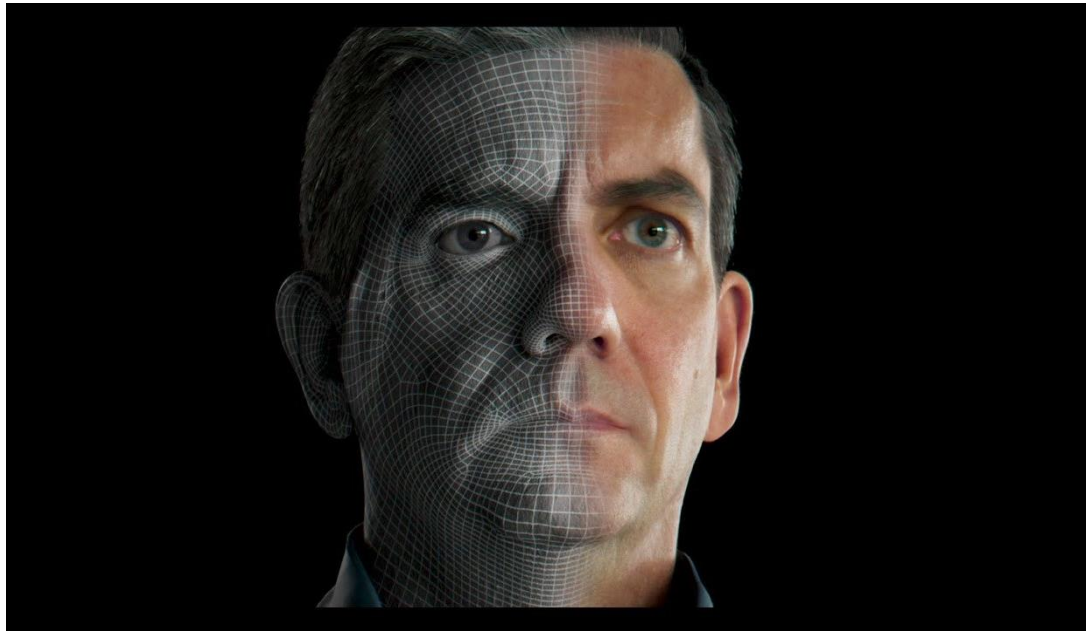
Chen Guo

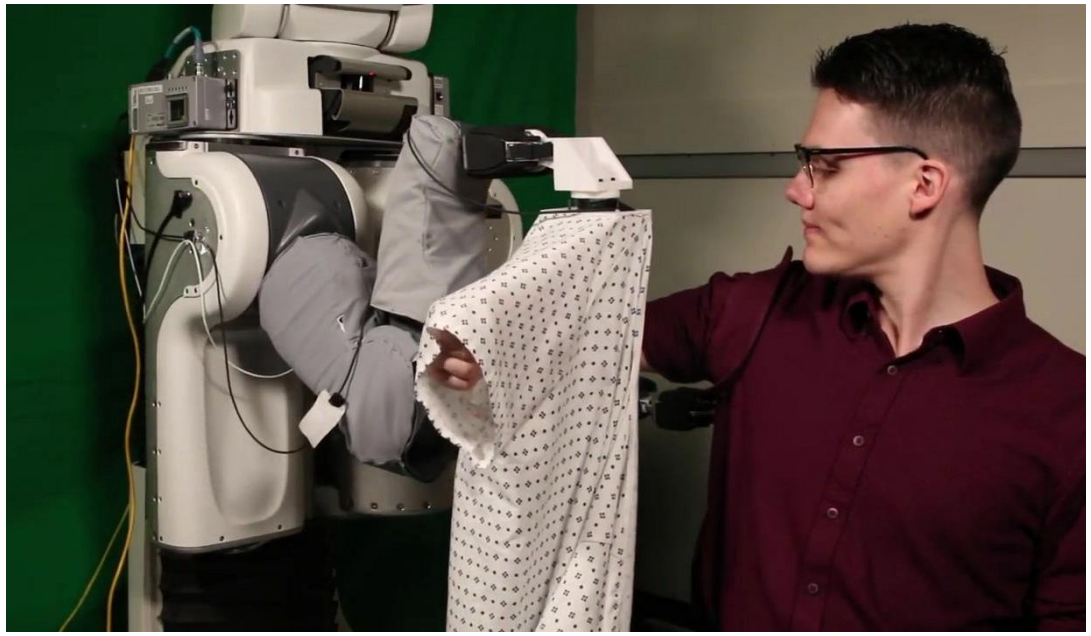
Wed., 10.04.2024

Advanced Interactive Technologies

ETH zürich

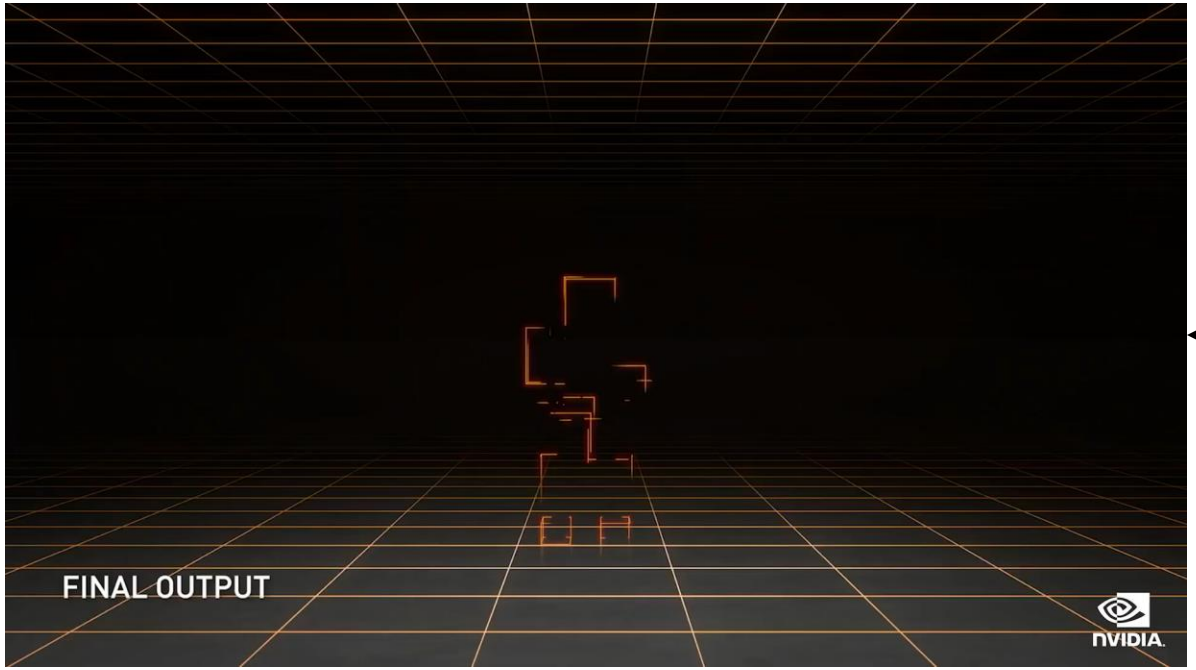
Microsoft





Credit: MetaHuman., Avatar, Gigazine, TM & © Lucasfilm Ltd.

Goal: **Automatic** reconstruction of **detailed** and **expressive** human model in **clothing** from **videos**



Immersive Telepresence

- High requirement of devices
- Tons of manual efforts

Credit: TM & © Lucasfilm Ltd., NVIDIA

Vid2Avatar: 3D Avatar Reconstruction from Video in the Wild via Self-supervised Scene Decomposition

Chen Guo Tianjian Jiang Xu Chen Jie Song Otmar Hilliges

CVPR 2023

ETH zürich

Advanced Interactive Technologies

Vid2Avatar: 3D Avatar Reconstruction from Video in the Wild via Self-supervised Scene Decomposition

Chen Guo Tianjian Jiang Xu Chen Jie Song Otmar Hilliges

CVPR 2023

ETH zürich

Advanced Interactive Technologies

Vid2Avatar: 3D Avatar Reconstruction from Video in the Wild via Self-supervised Scene Decomposition

Chen Guo Tianjian Jiang Xu Chen Jie Song Otmar Hilliges

CVPR 2023

ETH zürich

Advanced Interactive Technologies

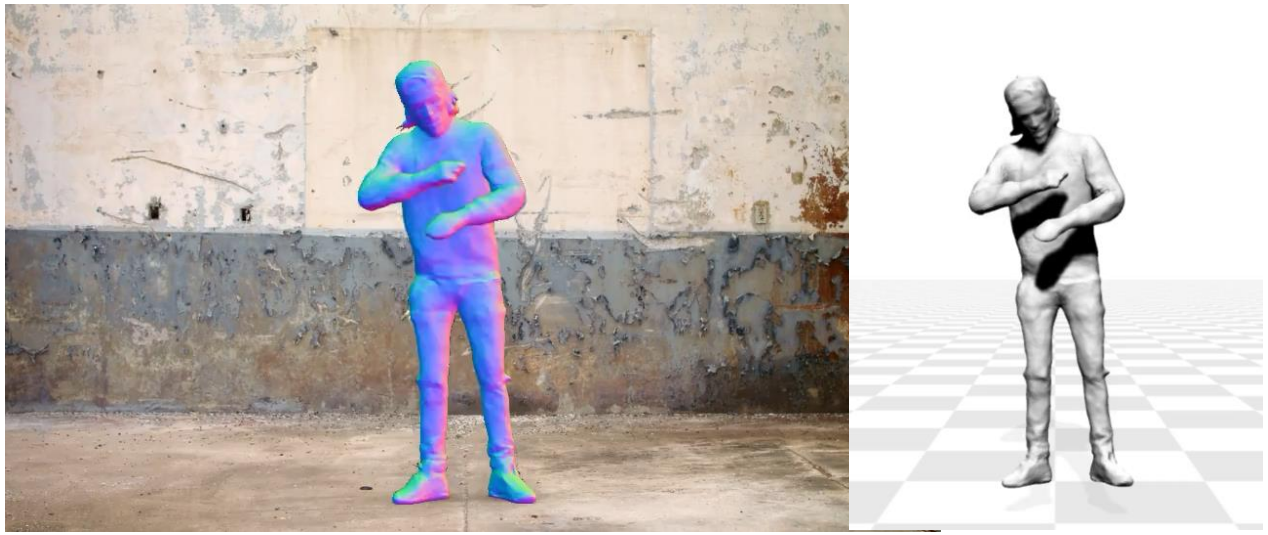
Problem Definition

Problem Definition

Input: RGB sequence

Problem Definition

Input: RGB sequence



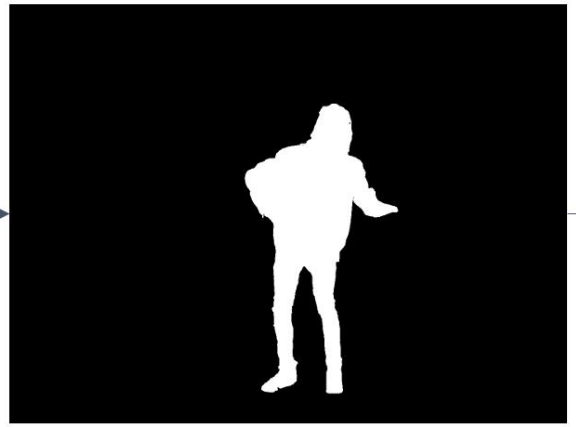
Reconstruction

- Accurate separation
- Detailed 3D surfaces

Motivation

Image

Off-the-shelf Segmentation Tool



Mask[1]

Reconstruction

SelfRecon[2]

[1] RVM: Lin et al. '21
[2] SelfRecon: Jiang et al. '22

Method Overview

Input: RGB sequence

Segmentation

Surface Reconstruction

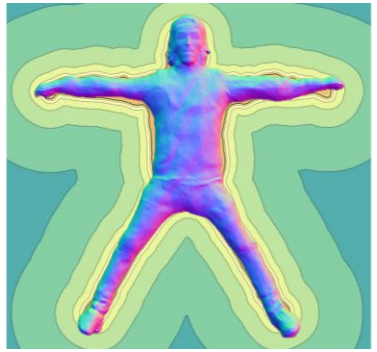
Method Overview

Input: RGB sequence

Segmentation

Surface Reconstruction

Joint 3D Optimization



Canonical SDF

Canonical Texture

Background Model

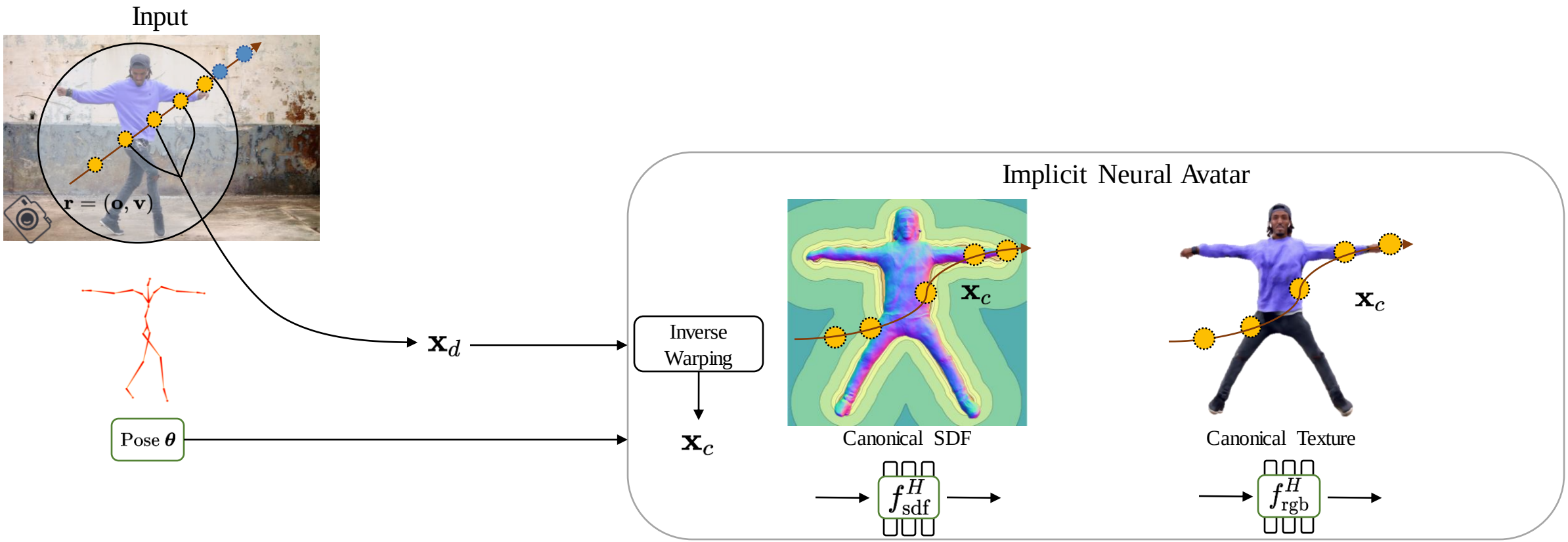
Method Overview

Input: RGB sequence

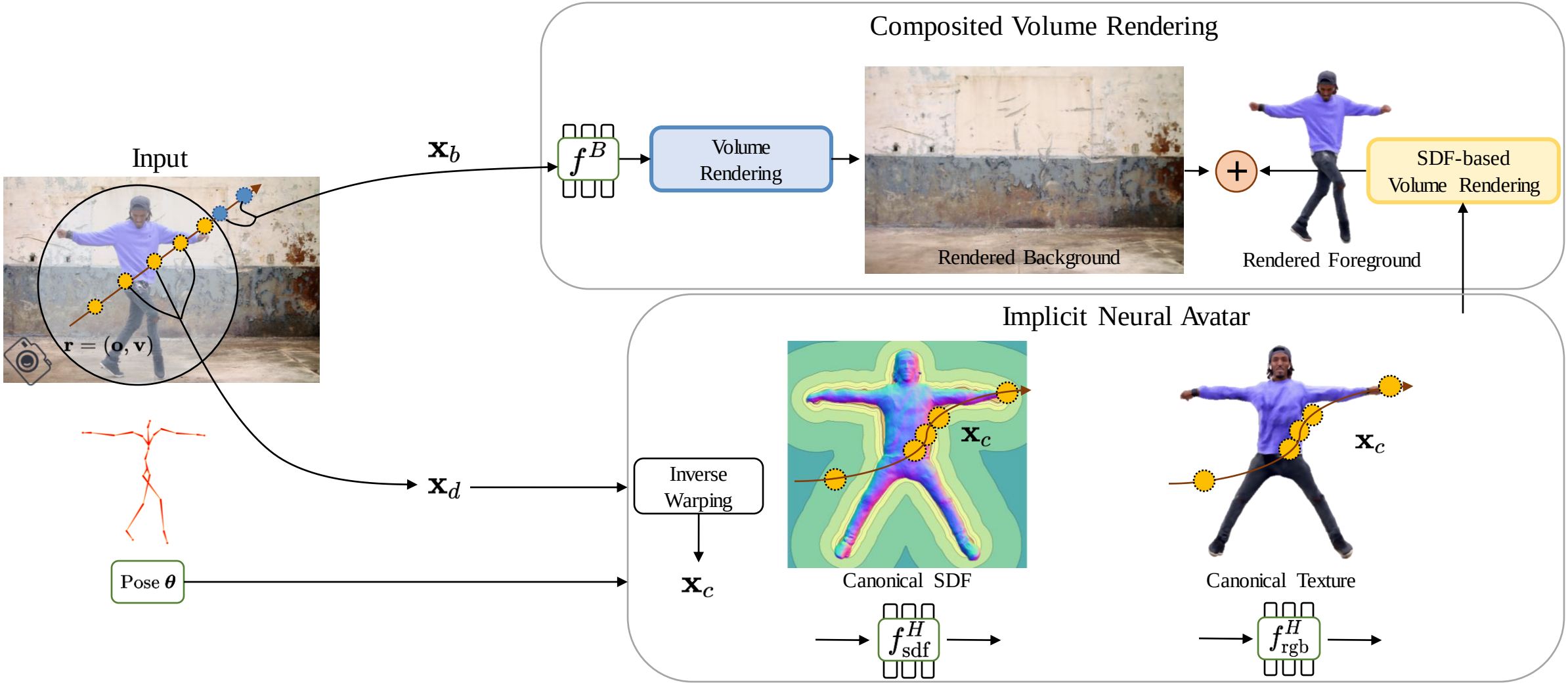
3D Avatar

Background

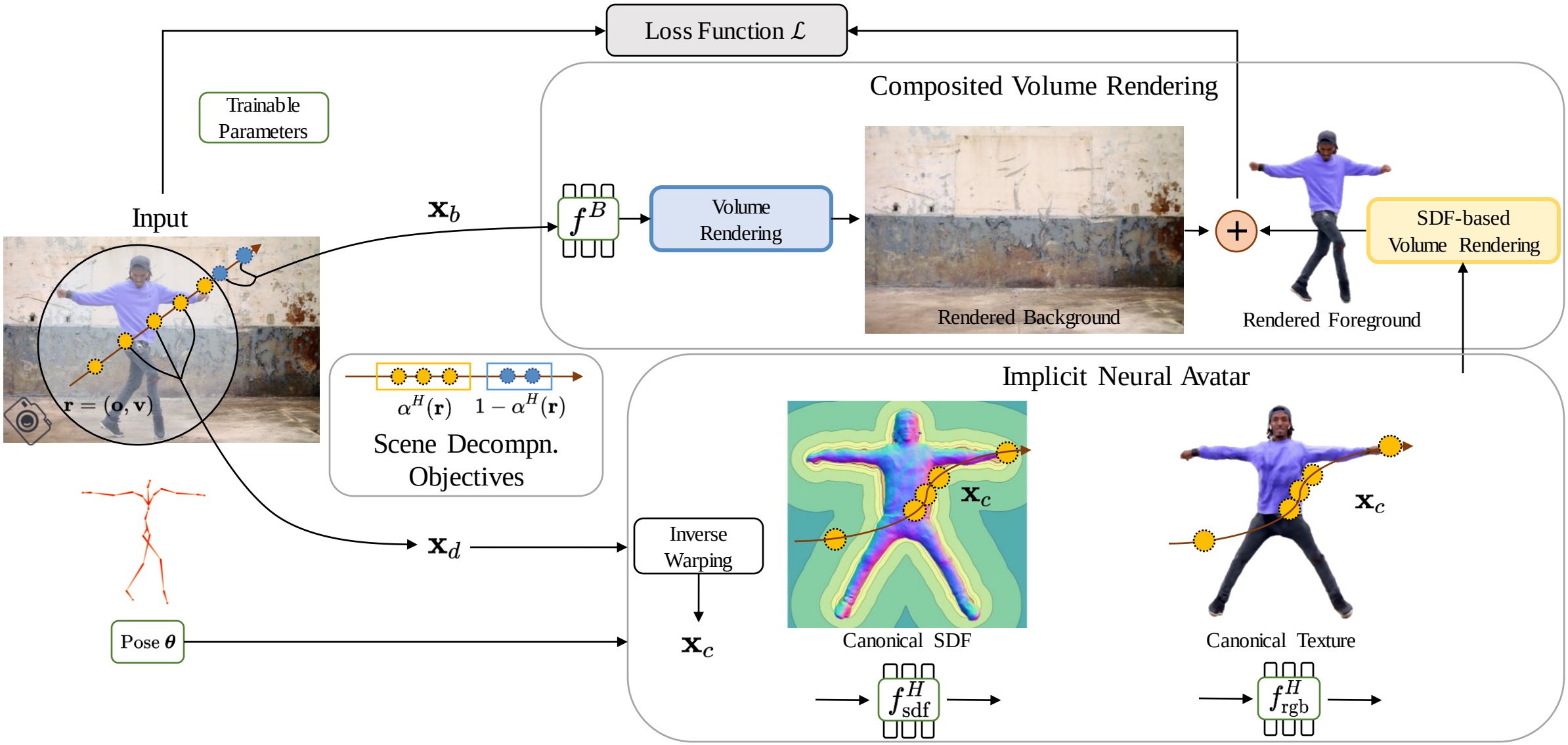
Method


Input
$\mathbf{r} = (\mathbf{o}, \mathbf{v})$
Pose $\boldsymbol{\theta}$
$\mathbf{x}_d$
Inverse Warping
$\mathbf{x}_c$
Implicit Neural Avatar
$\mathbf{x}_c$
Canonical SDF
f^H_{sdf}
$\mathbf{x}_c$
Canonical Texture
f^H_{rgb}

Method

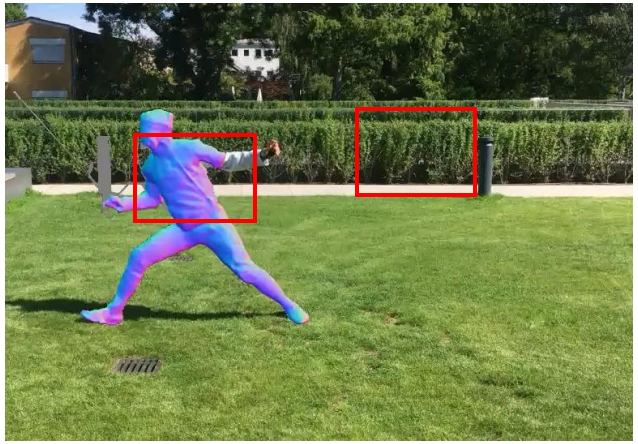

Composited Volume Rendering
Input
$\mathbf{x}_b$
f^B
Volume Rendering
Rendered Background
Rendered Foreground
SDF-based Volume Rendering
Implicit Neural Avatar
$\mathbf{r} = (\mathbf{o}, \mathbf{v})$
$\mathbf{x}_d$
Inverse Warping
$\mathbf{x}_c$
Pose $\boldsymbol{\theta}$
Canonical SDF
Canonical Texture
f^H_{sdf}
f^H_{rgb}

Method

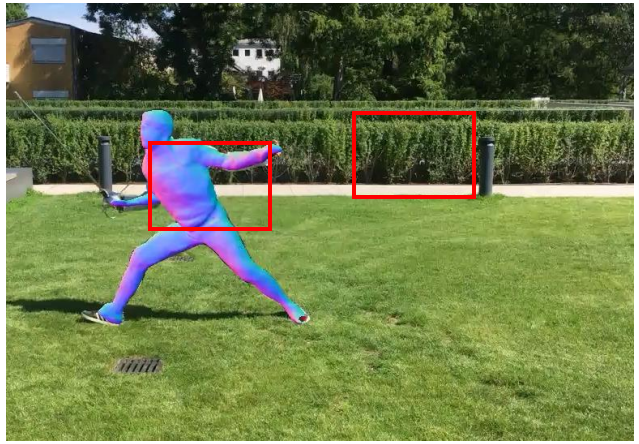

Loss Function $\mathcal{L}$
Trainable Parameters
Composited Volume Rendering
Input
$\mathbf{x}_b$
f^B
Volume Rendering
Rendered Background
SDF-based Volume Rendering
Rendered Foreground
$\mathbf{r} = (\mathbf{o}, \mathbf{v})$
$\alpha^H(\mathbf{r})$ $1 - \alpha^H(\mathbf{r})$
Scene Decompn. Objectives
Implicit Neural Avatar
$\mathbf{x}_d$
Inverse Warping
$\mathbf{x}_c$
Canonical SDF
Canonical Texture
Pose $\boldsymbol{\theta}$
f^H_{sdf}
f^H_{rgb}

Reconstruction Comparison on 3DPW

Video



ICON



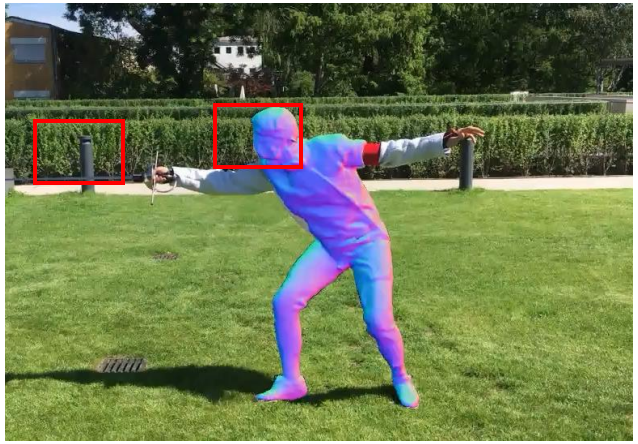
SelfRecon

Ours

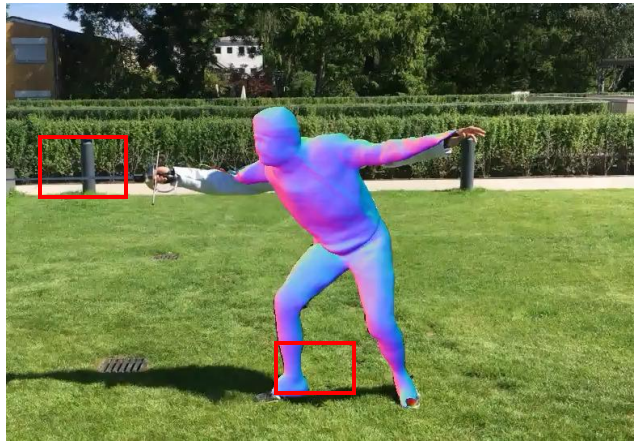
[ICON: Xiu et al. ‘22]
[SelfRecon: Jiang et al. ‘22]

Reconstruction Comparison on 3DPW

Video



ICON



SelfRecon

Ours

[ICON: Xiu et al. ‘22]
[SelfRecon: Jiang et al. ‘22]

Reconstruction Comparison on Online Video

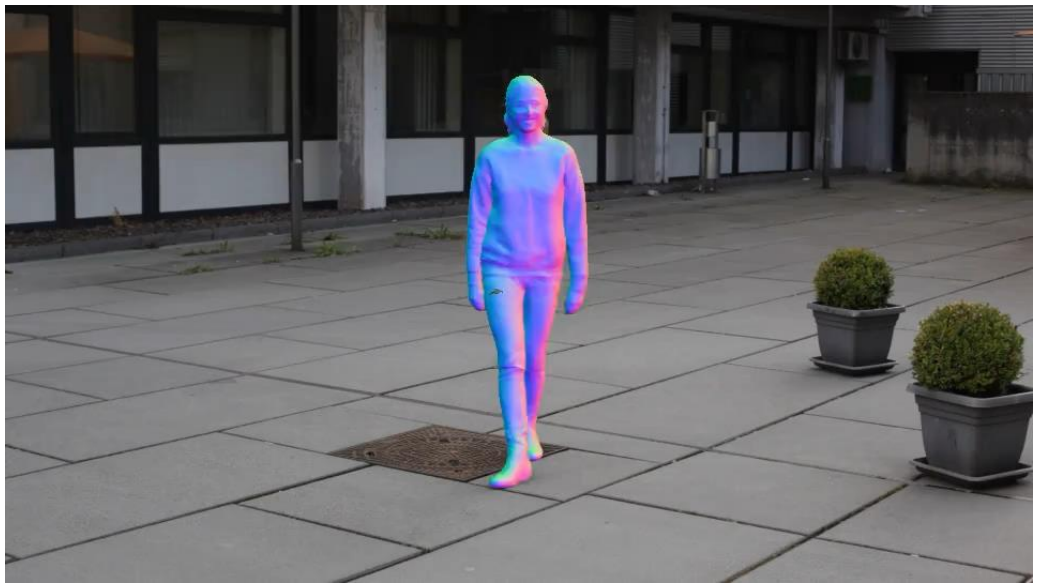
Video

ICON

SelfRecon

Ours

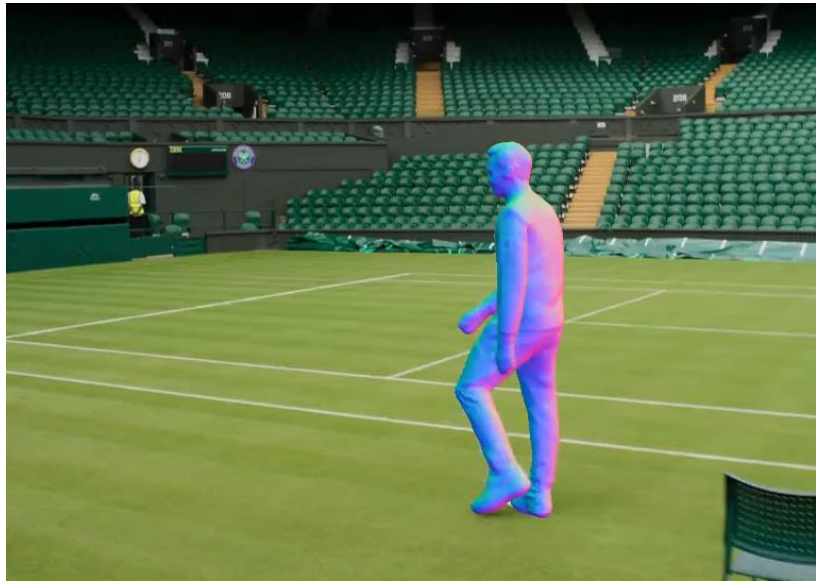
Qualitative Results on Monocular in-the-wild Videos – Datasets

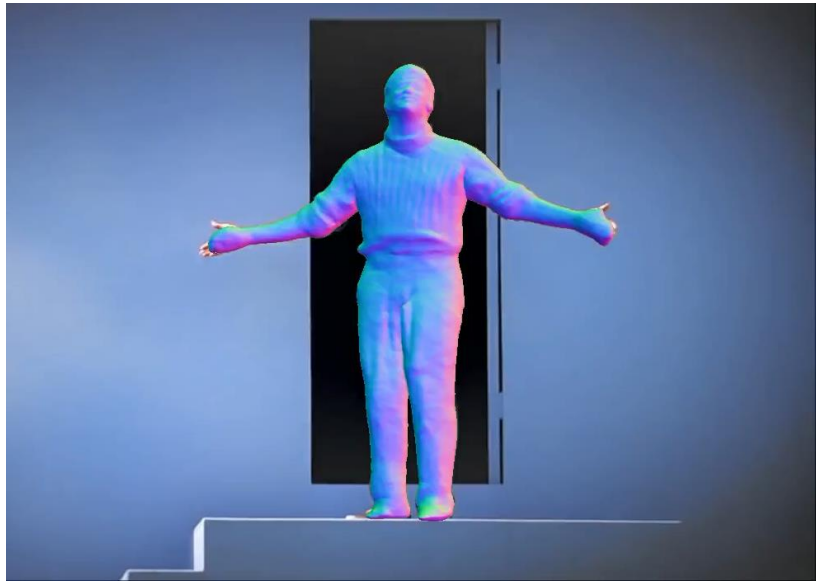


Video

Reconstruction

Qualitative Results on Monocular in-the-wild Videos – Online Videos



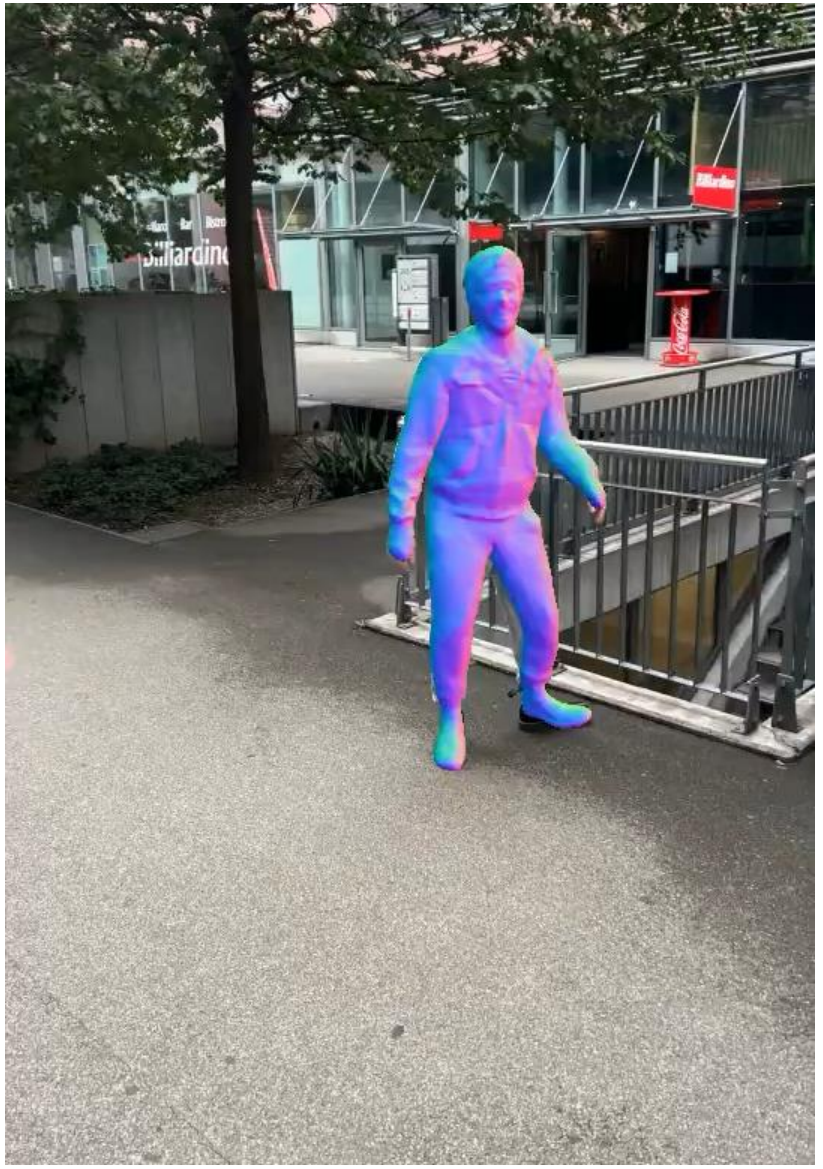


Video

Reconstruction

Qualitative Results on Monocular in-the-wild Videos – Self-captured Videos

Video



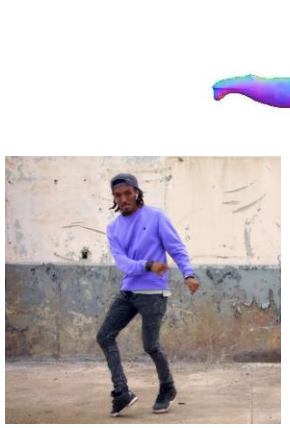
Reconstruction

Qualitative Results on Monocular in-the-wild Videos – Self-captured Videos

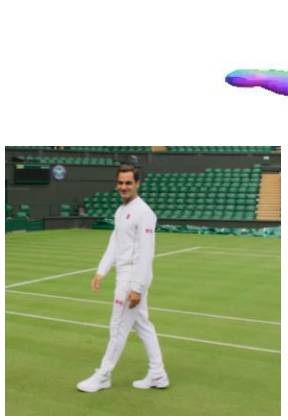
Video

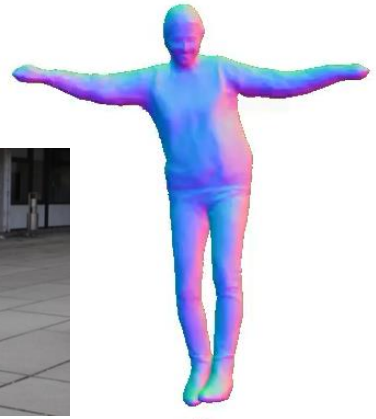
Reconstruction

Animation







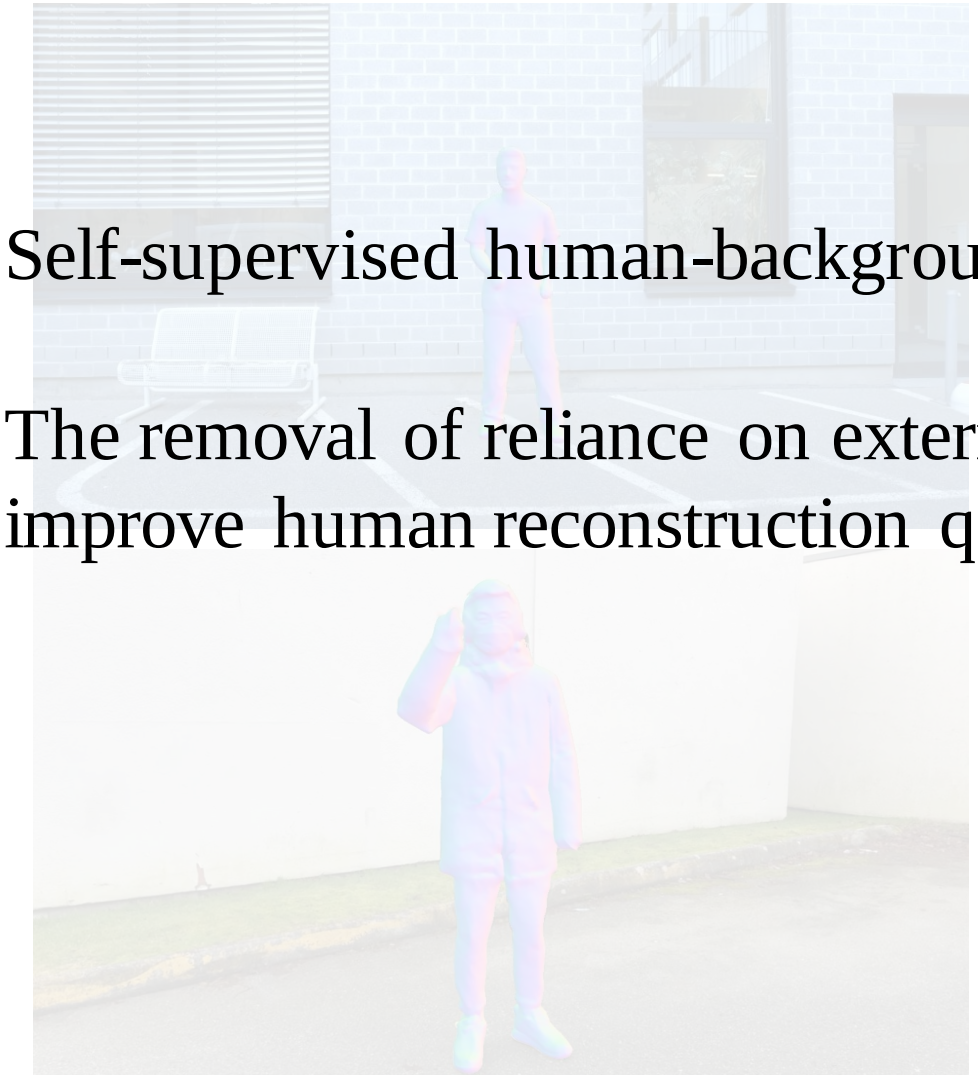



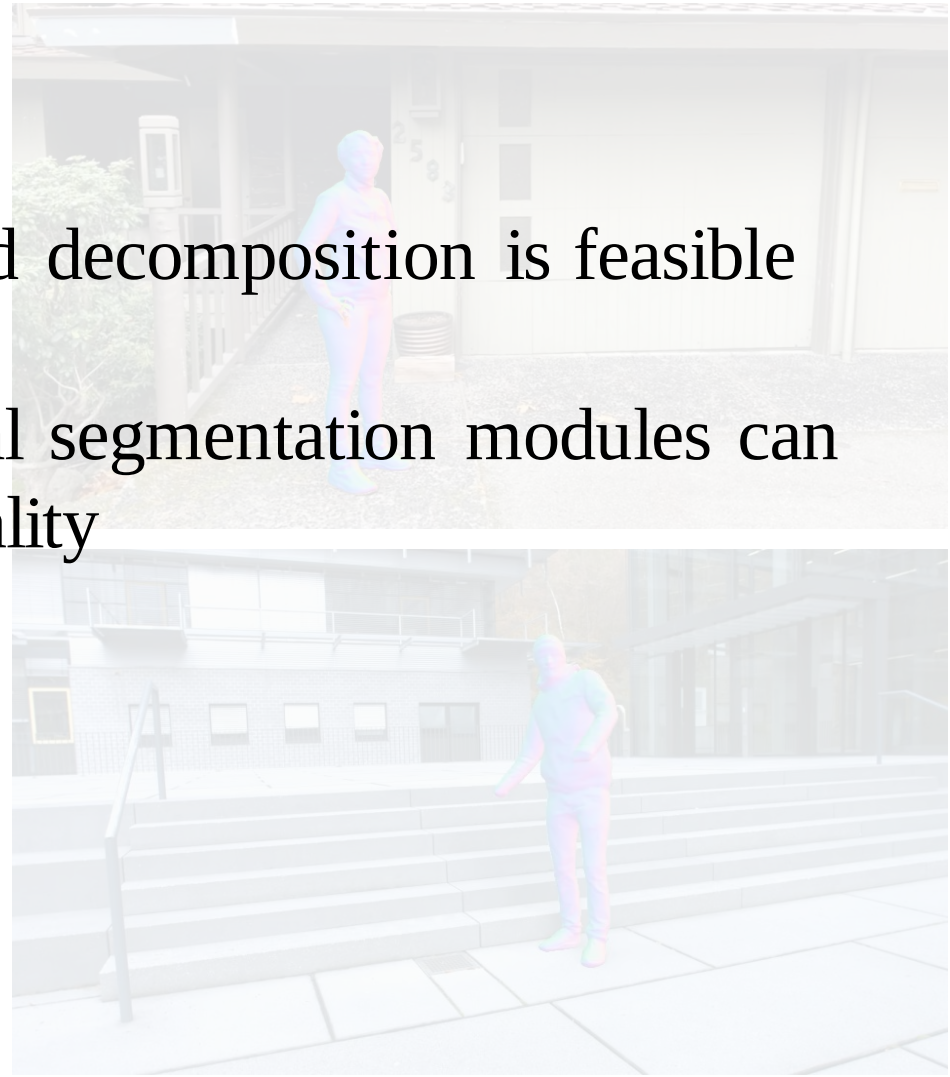






Summary





- Self-supervised human-background decomposition is feasible
- The removal of reliance on external segmentation modules can improve human reconstruction quality

X-Avatar

Expressive Human Avatars

Kaiyue Shen[1*] Chen Guo[1*] Manuel Kaufmann[1] Juan Jose Zarate[1]

Julien Valentin[2] Jie Song[1] Otmar Hilliges[1]

*Equal Contribution [1]ETH Zurich [2]Microsoft

CVPR 2023

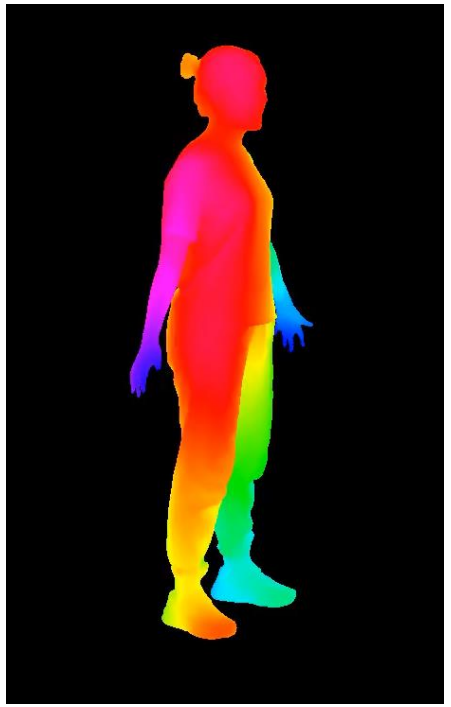
ETH zürich

Advanced Interactive Technologies

Microsoft

Problem Setting

OR

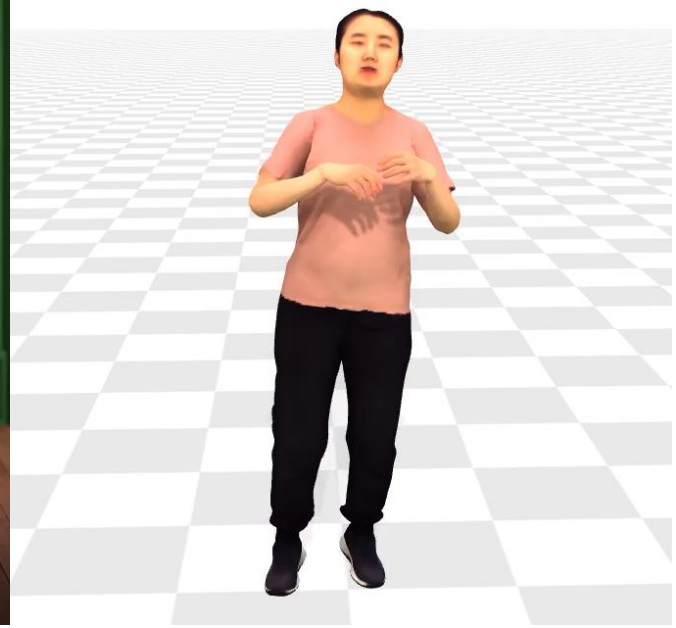


Scan input

RGB-D input

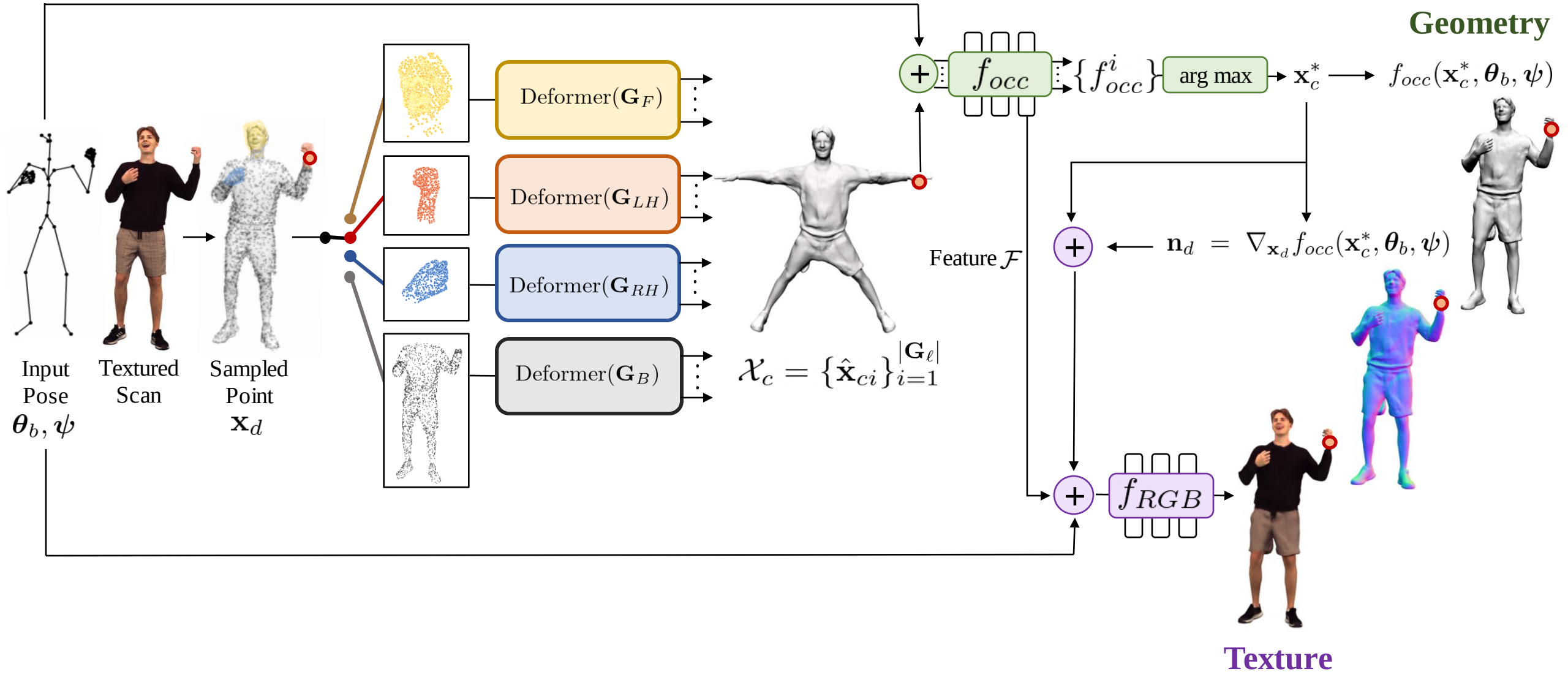
Built X-Avatar

Driving pose



Animation

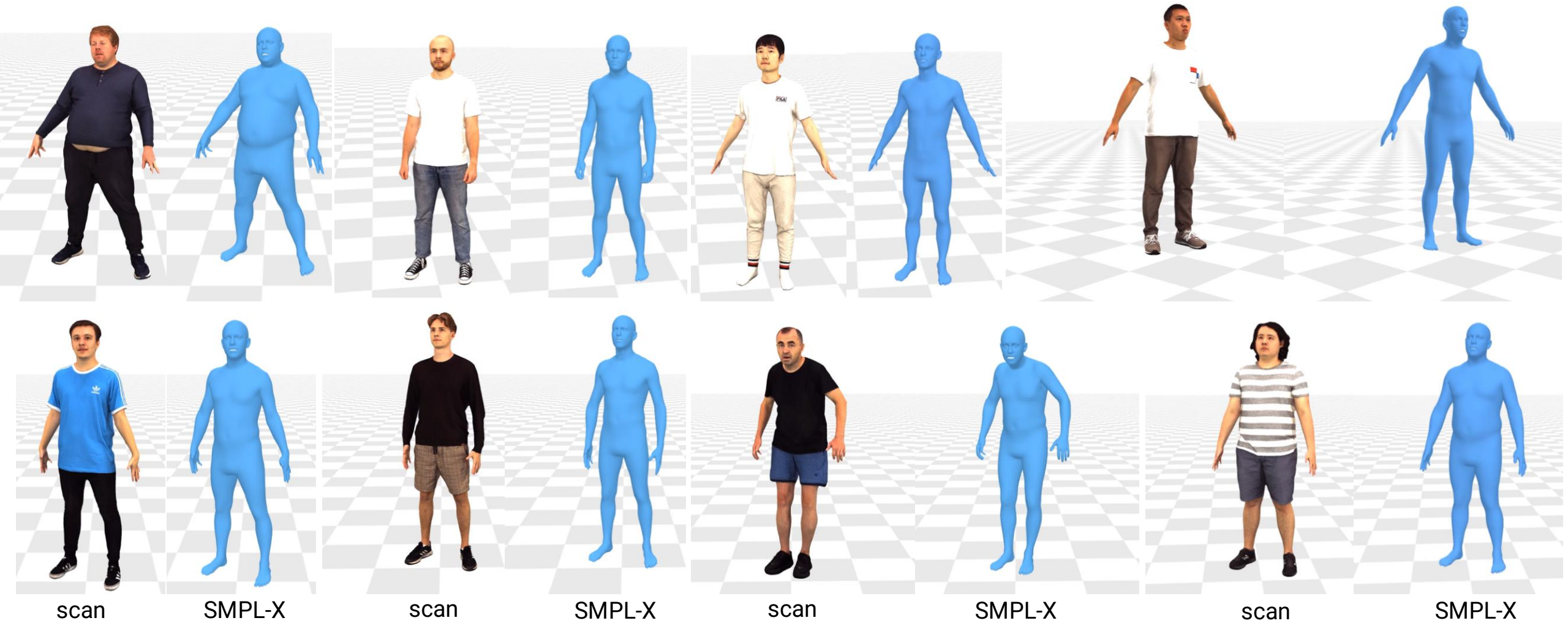
Zoom in


Geometry
Input Pose $\boldsymbol{\theta}_b, \boldsymbol{\psi}$
Textured Scan
Sampled Point $\mathbf{x}_d$
Deformer($\mathbf{G}_F$)
Deformer($\mathbf{G}_{LH}$)
Deformer($\mathbf{G}_{RH}$)
Deformer($\mathbf{G}_B$)
$\mathcal{X}_c = \{\hat{\mathbf{x}}_{ci}\}_{i=1}^{|\mathbf{G}_\ell|}$
Feature $\mathcal{F}$
f_{occ}
$\{f_{occ}^i\}$
arg max
$\mathbf{x}_c^*$
$f_{occ}(\mathbf{x}_c^*, \boldsymbol{\theta}_b, \boldsymbol{\psi})$
$\mathbf{n}_d = \nabla_{\mathbf{x}_d} f_{occ}(\mathbf{x}_c^*, \boldsymbol{\theta}_b, \boldsymbol{\psi})$
f_{RGB}
Texture

X-Humans

Male Subjects (shown 8 of 11)

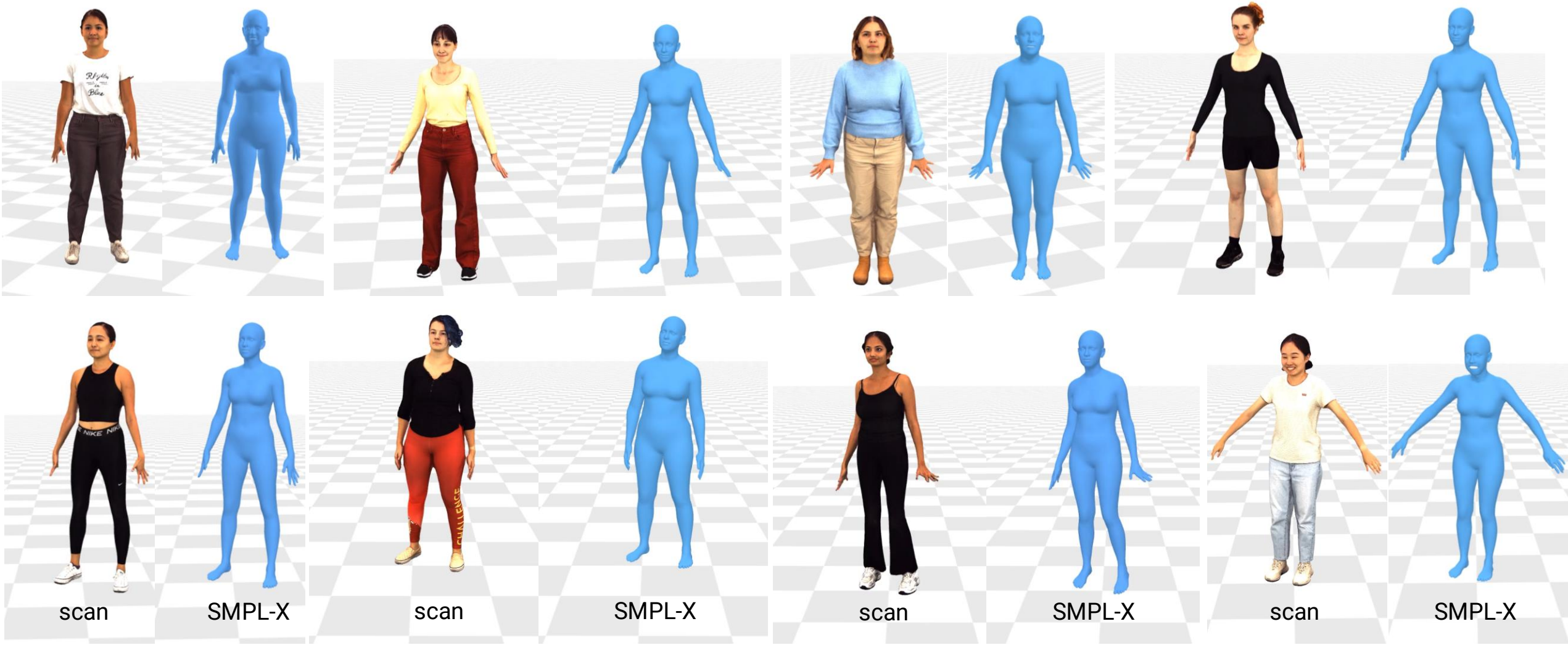
- 20 subjects, 234 sequences, 34,663 frames
- Textured scans, SMPL[-X] registrations



- 20 subjects, 234 sequences, 34,663 frames
- Textured scans, SMPL[-X] registrations

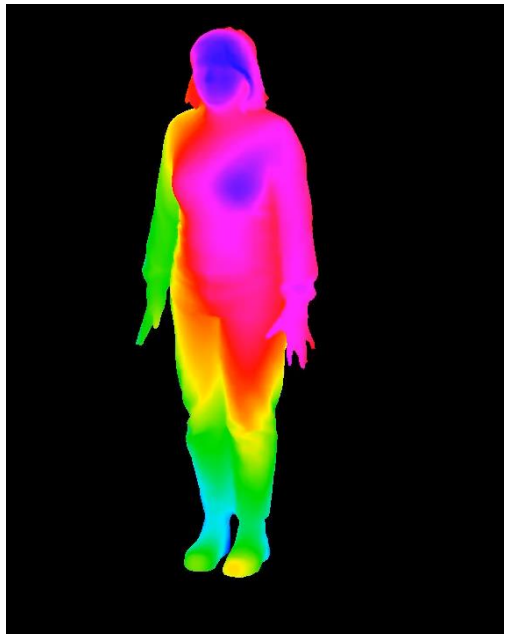
Female Subjects (shown 8 of 9)

- 20 subjects, 234 sequences, 34,663 frames
- Textured scans, SMPL[-X] registrations
- Body pose + hand gesture + facial expression
- Various clothing types, hair styles, genders and ages



Creating X-Avatars from RGB-D

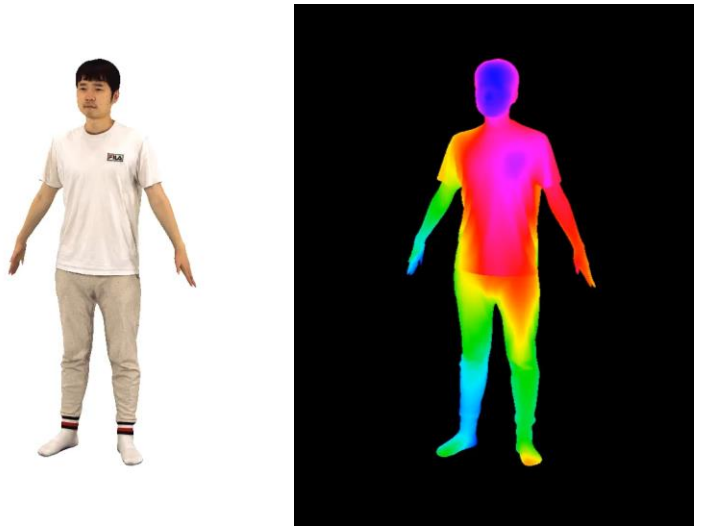
RGB-D input



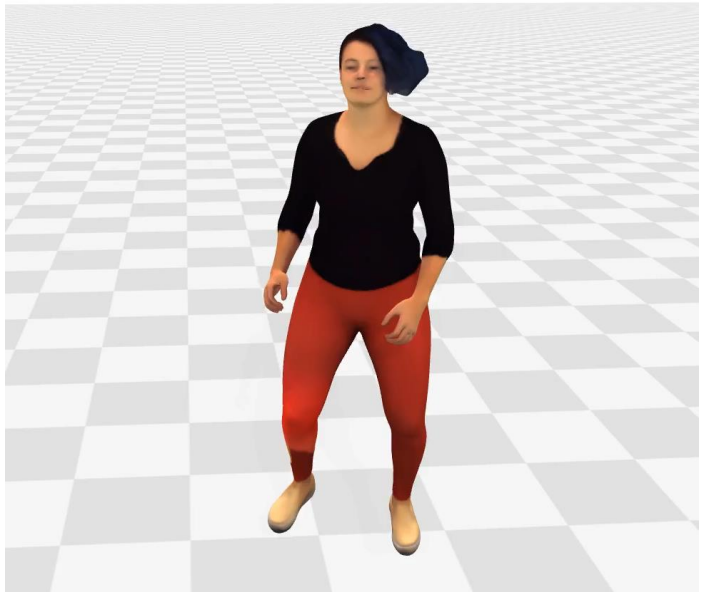
Built X-Avatar

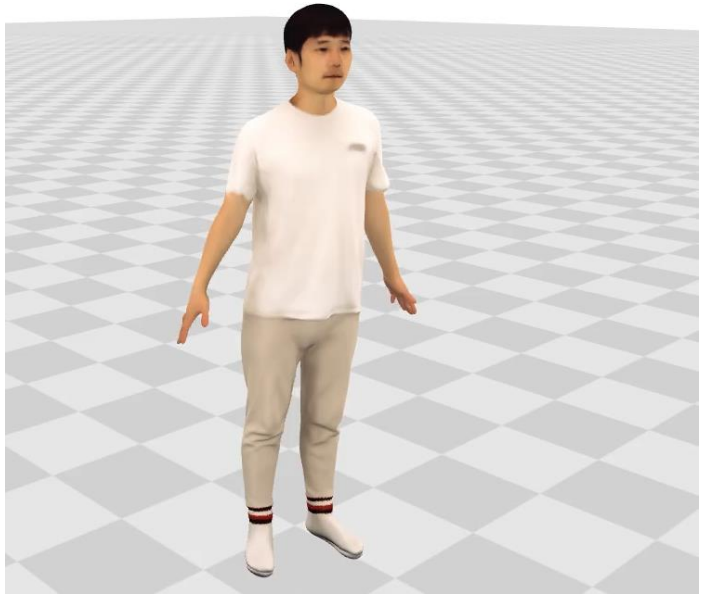
Animation

RGB-D input



Animation





Re-targeting (tennis)

Re-targeting (dance)

Re-targeting with 3D consistency

Summary

- **X-Avatar**, one of the first expressive implicit human avatar models.
- An unified approach to efficiently build X-Avatars from scans and RGB-D data.
- **X-Humans**, a dataset of high-quality textured scans of clothed people performing varied body and hand movements and facial expressions.

Future Work

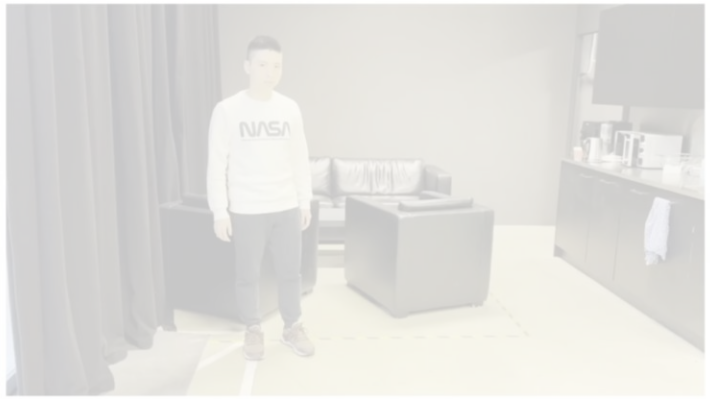
Reconstruct Anything

Vid2Avatar

↓

Vid2Anything

humans, scene, objects, etc.

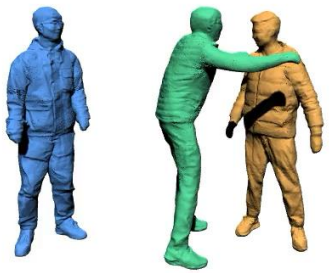


Input: RGB video

Input

Reconstructions

Input



Reconstructions

Accepted as CVPR 2024 Oral

Thank you!

Co-authors: Otmar Hilliges, Kaiyue Shen, Yifei Yin, Jie Song, Julien Valentin, Xu Chen,, Tianjian Jiang, Manuel Kaufmann, Juan Zarate