ECON: Explicit Clothed humans Optimized via Normal Integration (Highlight)





Yuliang Xiu[1] Jinlong Yang[1] Xu Cao[2] Dimitrios Tzionas[3] Michael J. Black[1]

[1]Max Planck Institute for Intelligent Systems *[2]Osaka Univeristy* *[3]University of Amsterdam*

Poster ID: TUE-AM-049

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

clipe clipe-itn.eu

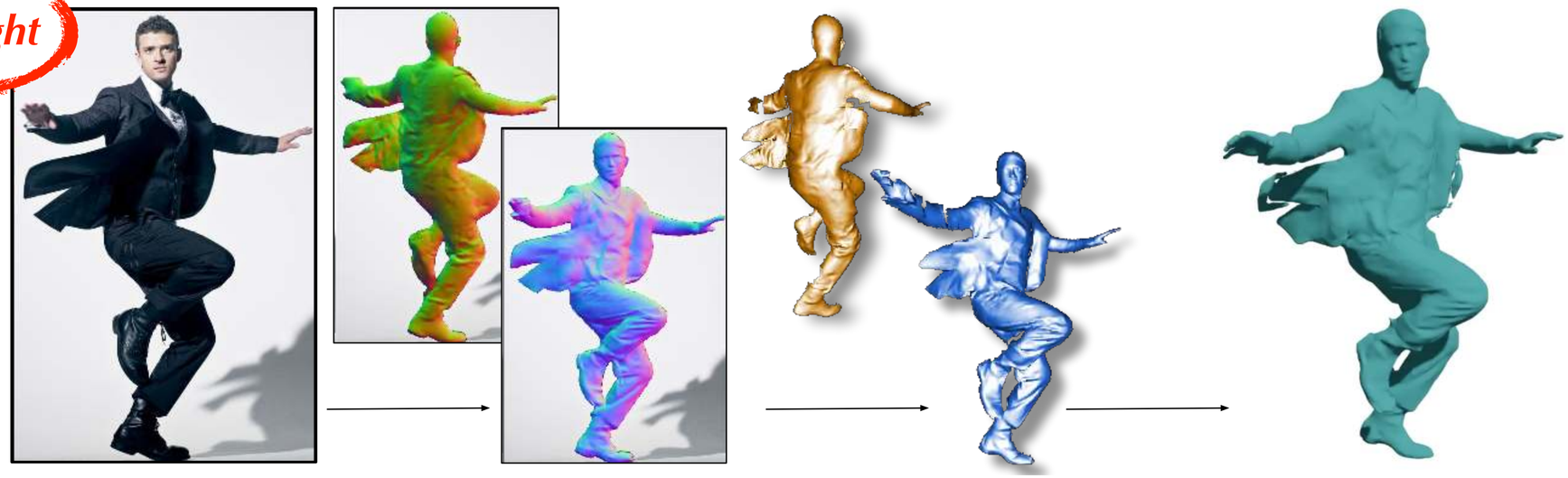
大阪大学 OSAKA UNIVERSITY

UNIVERSITY OF AMSTERDAM

PERCEIVING SYSTEMS MAX PLANCK INSTITUTE FOR INTELLIGENT SYSTEMS

ECON: Explicit Clothed humans Optimized via Normal Integration

Highlight



Yuliang Xiu[1]

Jinlong Yang[1]

Xu Cao[2]

Dimitrios Tzionas[3]

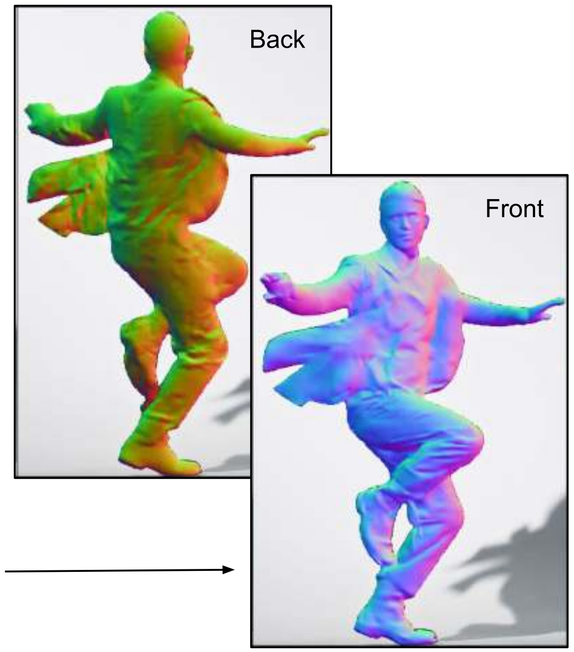
Michael J. Black[1]

[1]*Max Planck Institute for Intelligent Systems* [2]*Osaka Univeristy* [3]*University of Amsterdam*

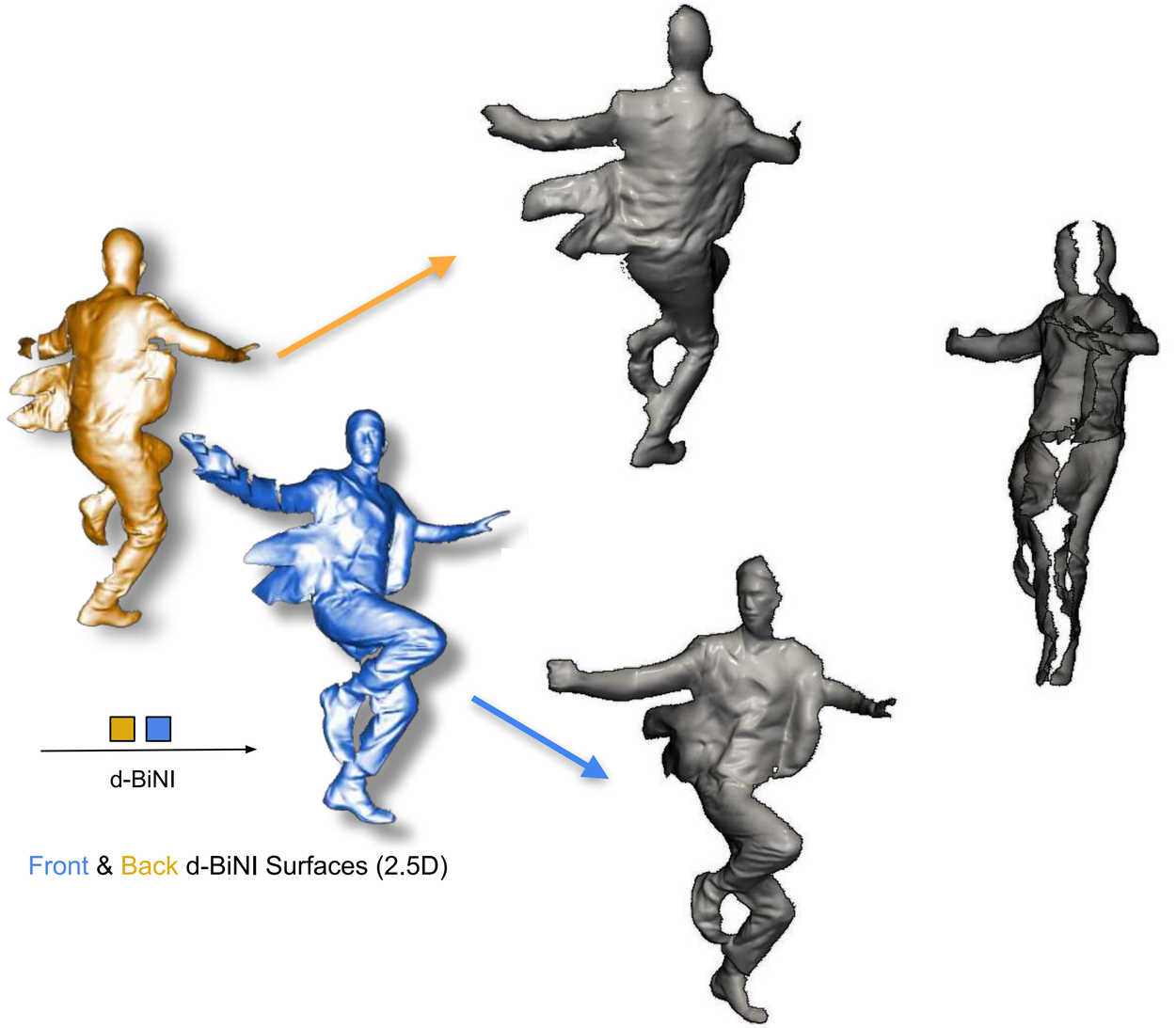
Pose Robustness + Topological Flexibility

Overview of ECON

Input Image



Front & Back Normal Estimation (2D)

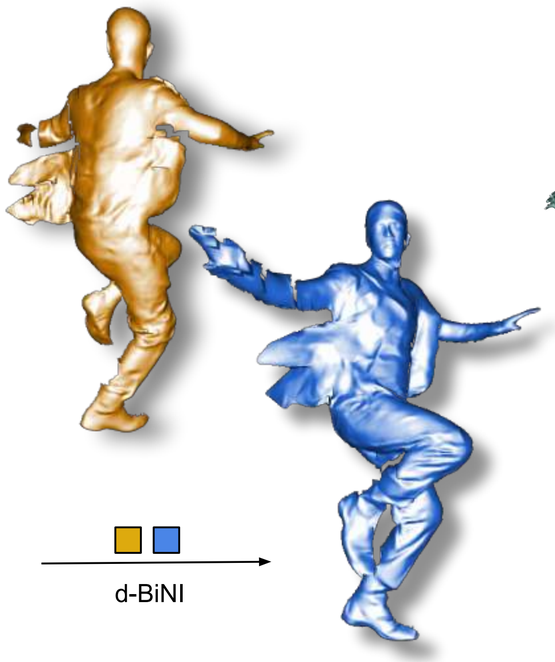


Front & Back d-BiNI Surfaces (2.5D)

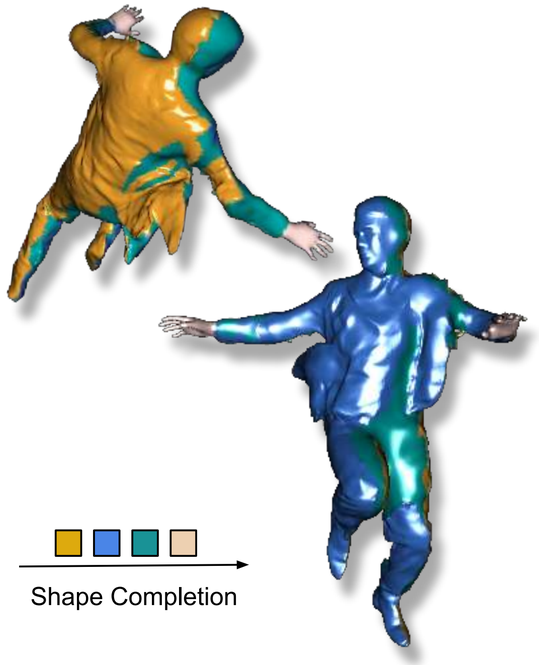
Overview of ECON

Input Image

Front & Back Normal Estimation (2D)

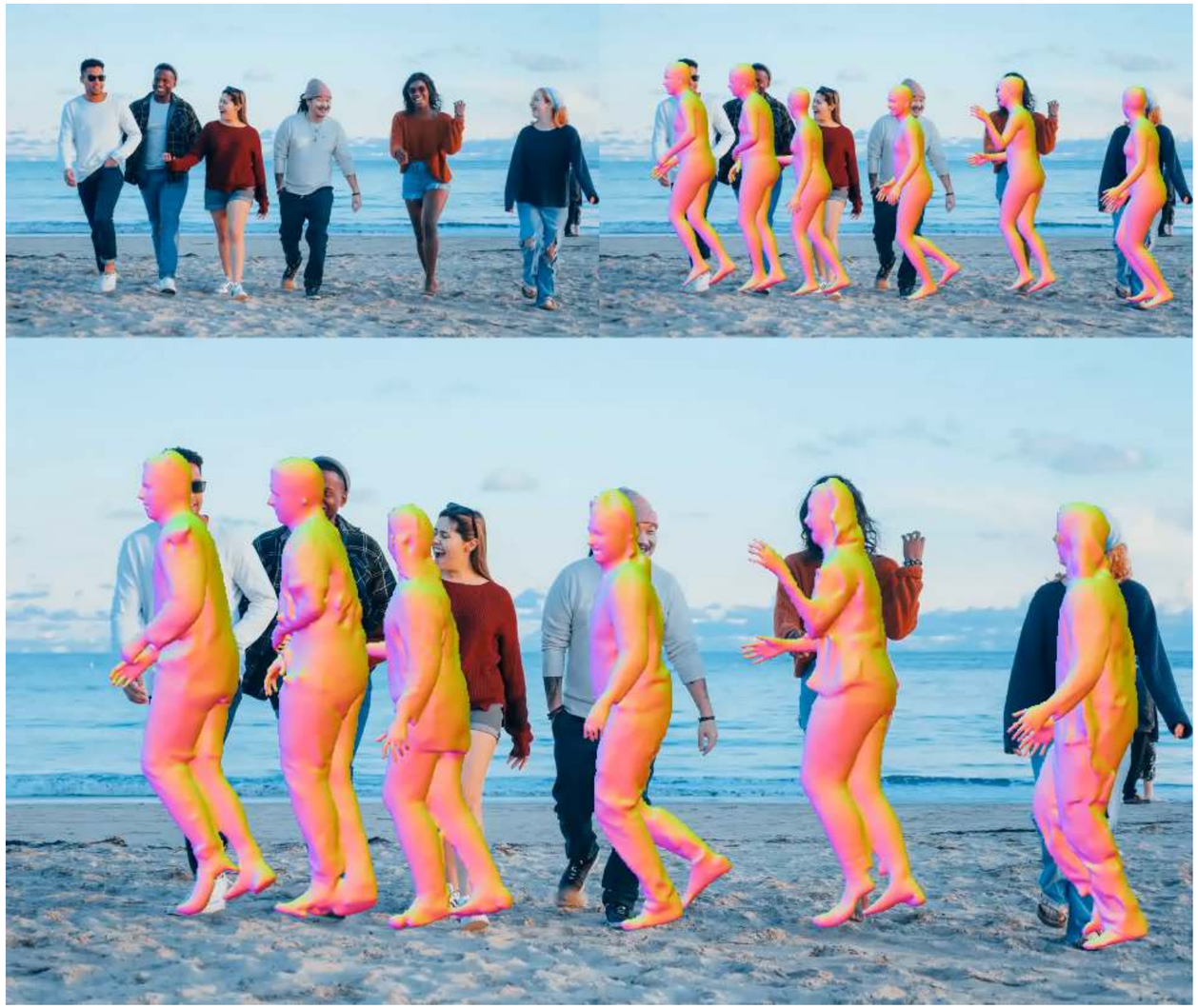


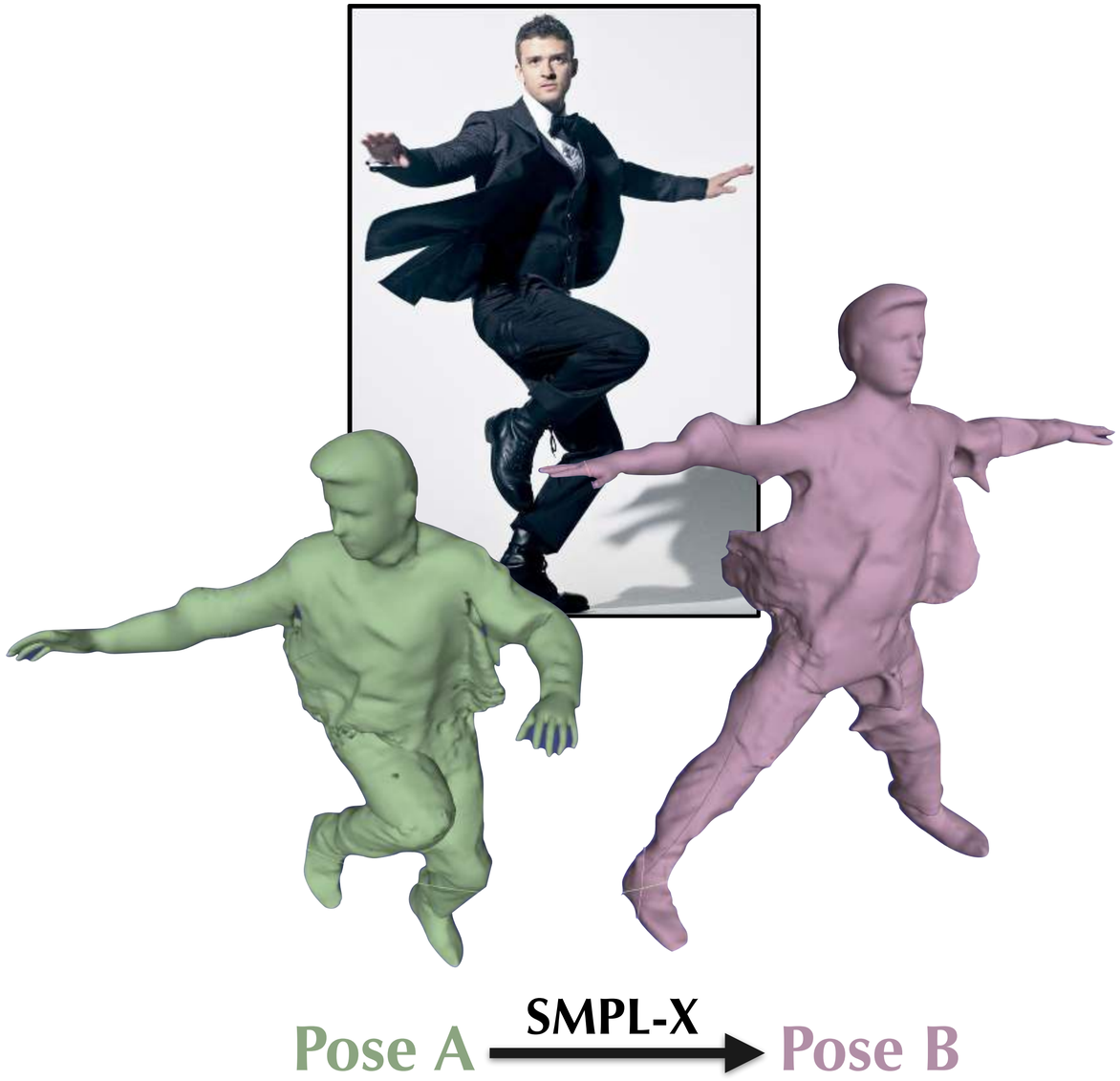
Front & Back d-BiNI Surfaces (2.5D)



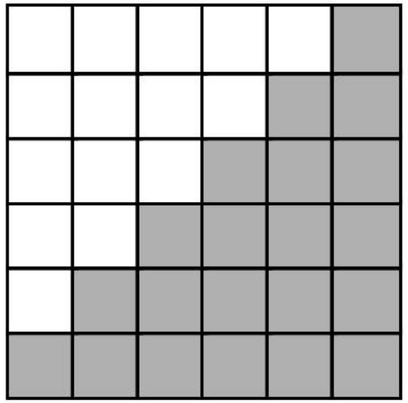
Full Clothed Human Mesh (3D)

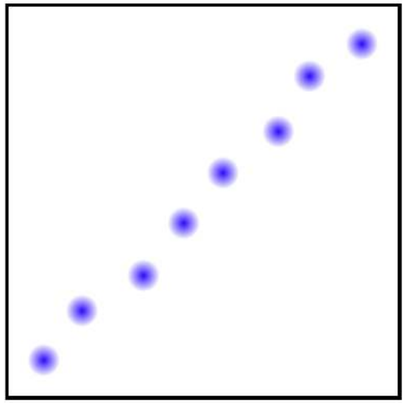
Multi-Person & Animatable

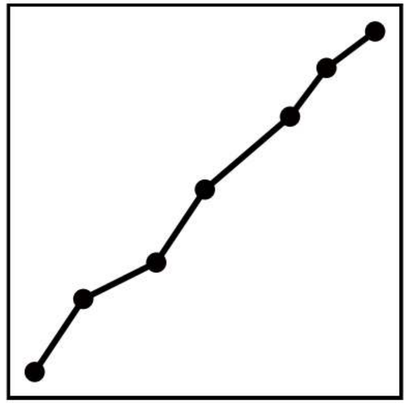


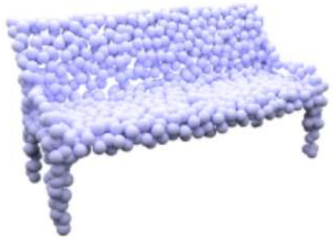


Explicit vs Implicit





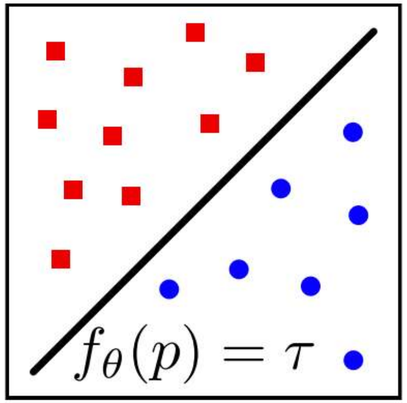




Voxels

Point Cloud

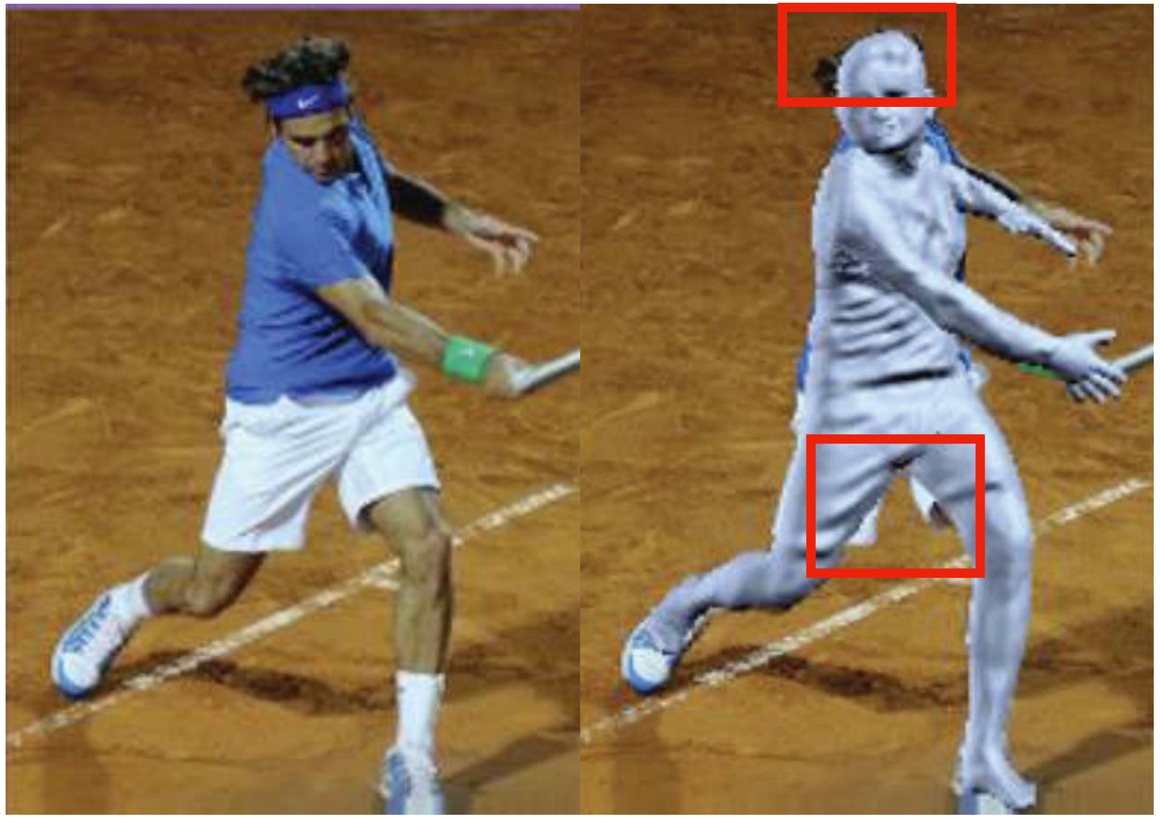
Mesh



Implicit

Explicit-based Methods

Tex2Shape (ICCV 2019)
Alldieck et al.



HMD (CVPR 2019)
Zhu *et al.*

Implicit-based Methods

PIFu (ICCV 2019)
Saito *et al.*

PIFuHD (CVPR 2020)
Saito *et al.*

Arbitrary Topologies for **PIFuHD**

Unseen Poses for **PIFuHD**

Implicit w/ 3D Prior

PaMIR (TPAMI 2021)
Zheng et al.

ICON (CVPR 2022)
Xiu *et al.*

Loose Clothing for **PaMIR/ICON**

PaMIR (TPAMI 2021)
Zheng et al.

ICON (CVPR 2022)
Xiu *et al.*

Loose Clothing for PaMIR/ICON

PaMIR (TPAMI 2021)
Zheng et al.

ICON (CVPR 2022)
Xiu *et al.*

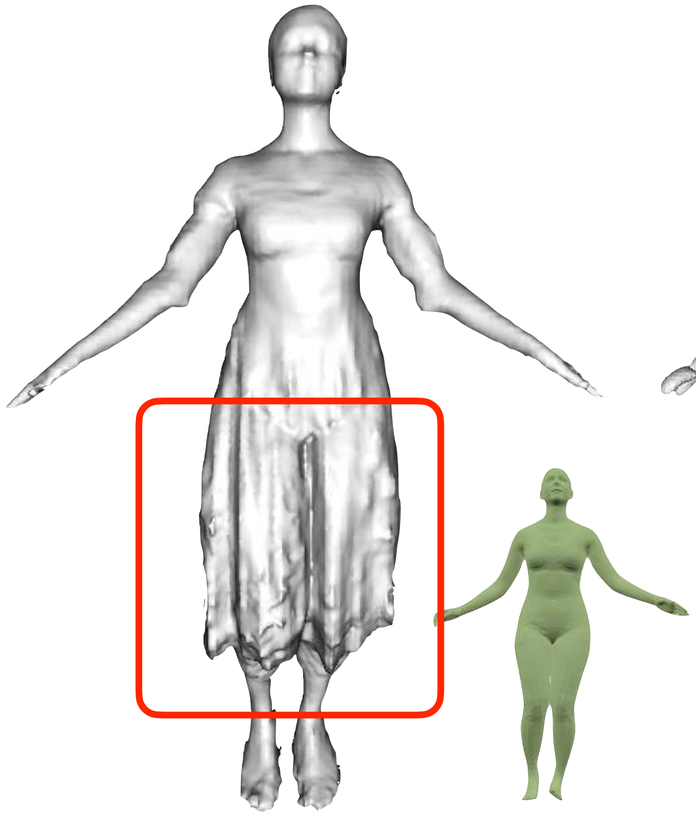
Loose Clothing

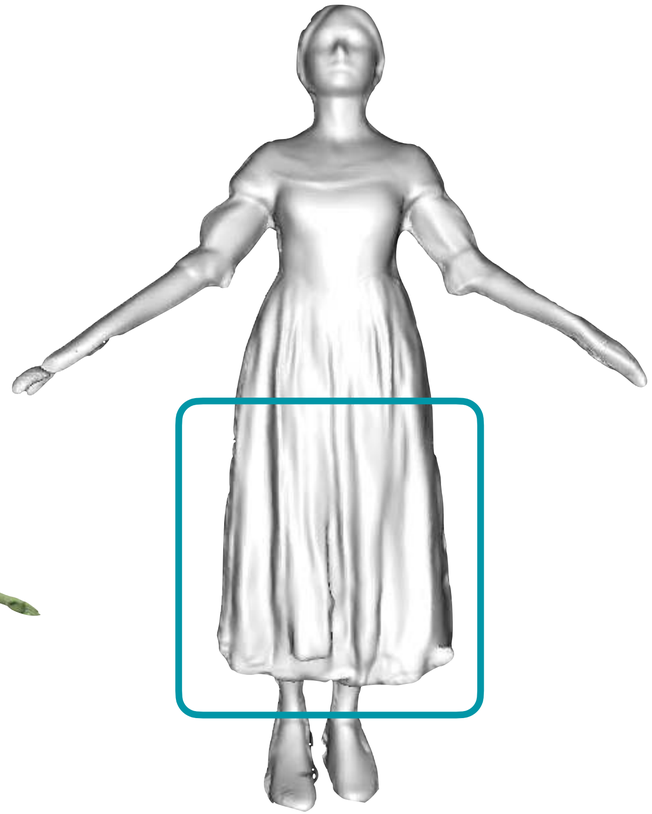
Challenging Pose

ECON is the best of both worlds

Good Normal
Bad Mesh

Good Normal
Good Mesh

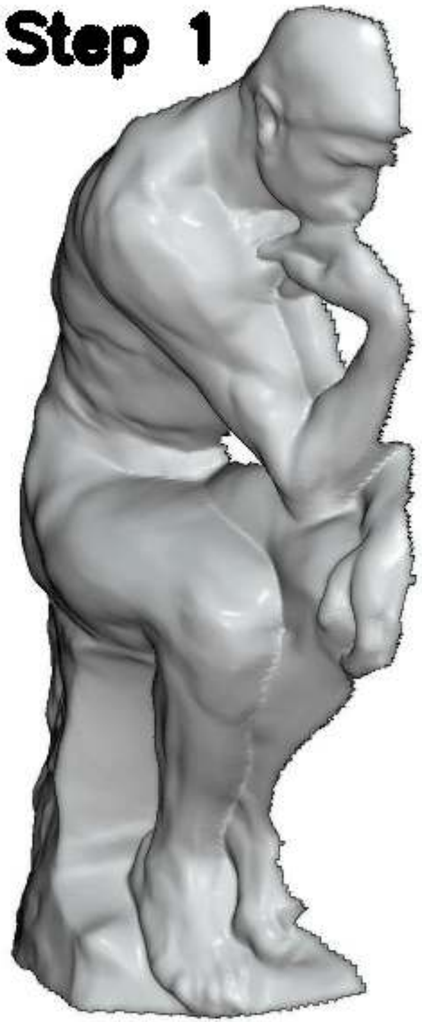


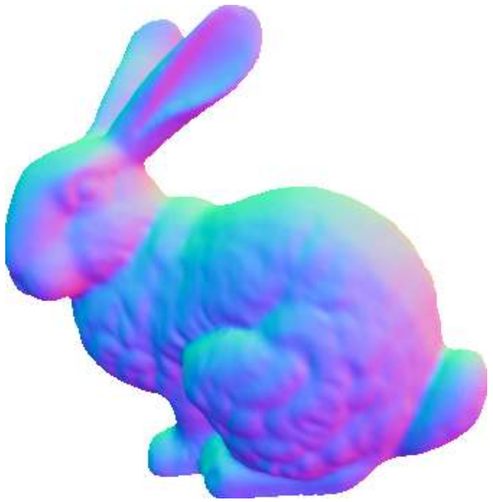


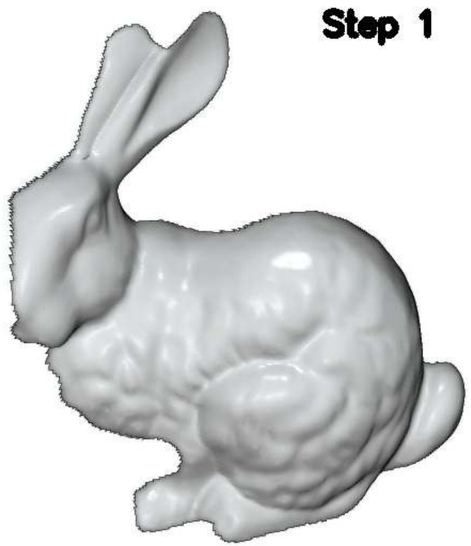
ICON (CVPR'22)
Xiu *et al.*

ECON (CVPR'23)
Xiu *et al.*

Depth-ambiguity (one-to-many) of **Normal Integration**

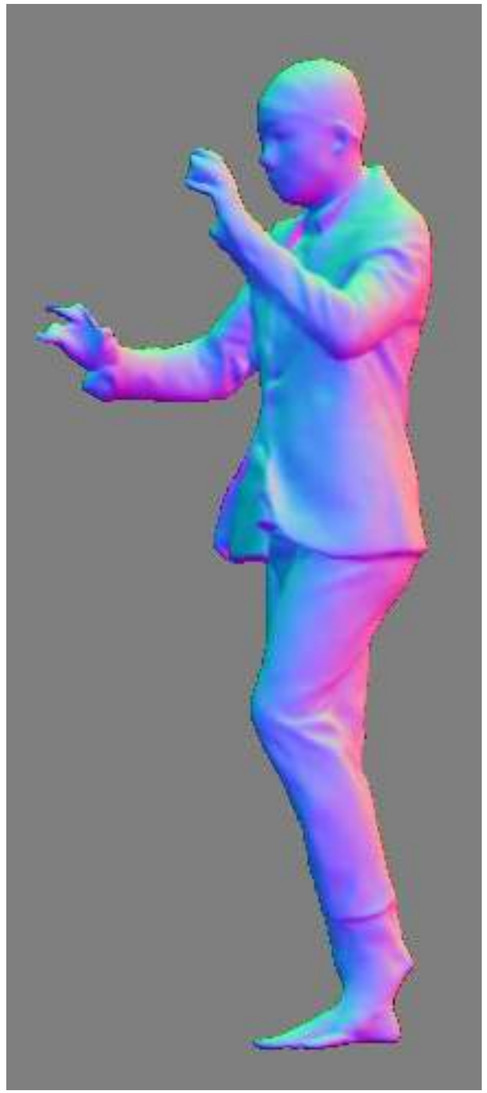




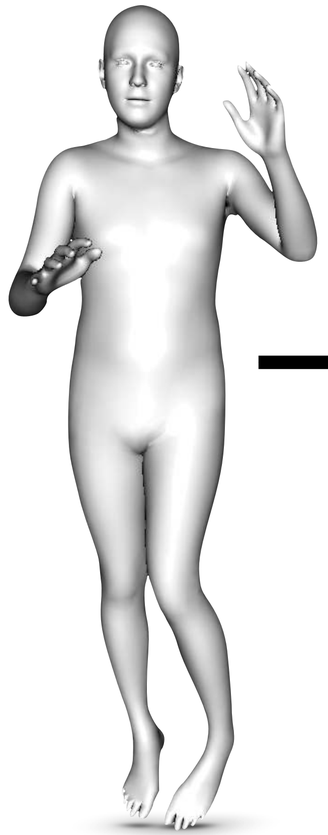


BiNI: Bilateral Normal Integration (ECCV 2022)

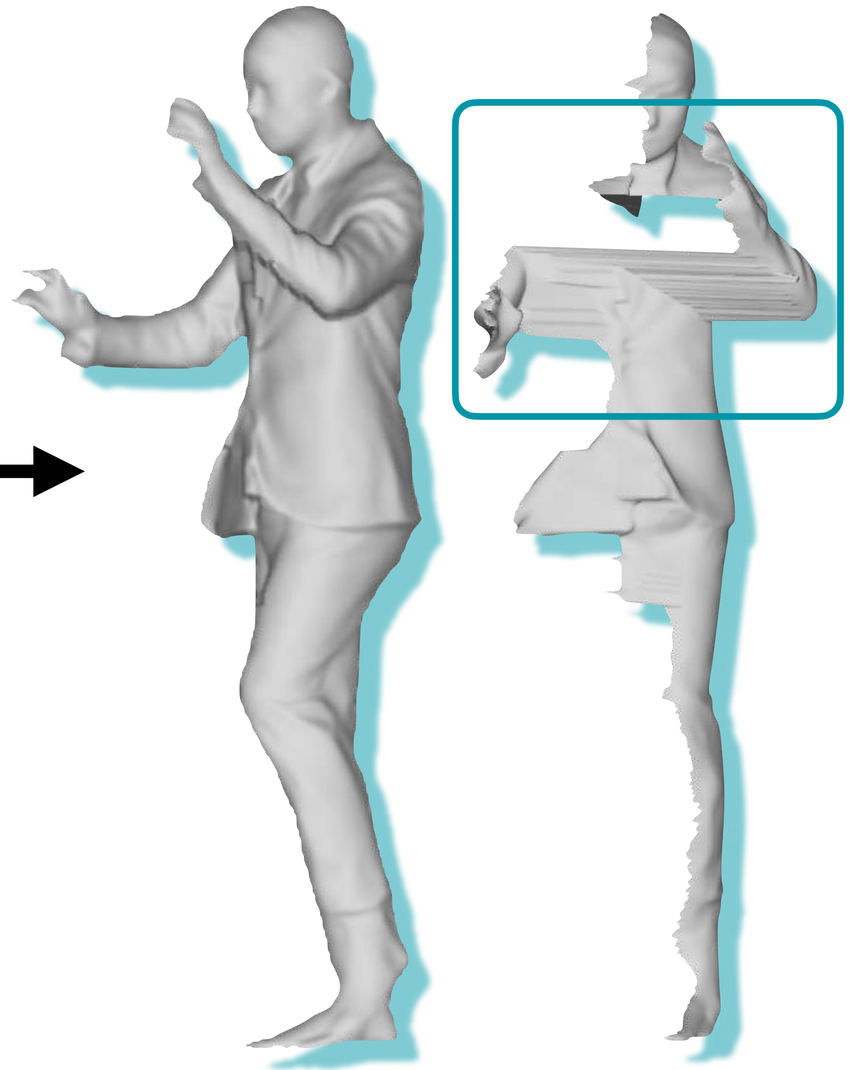
BiNI -vs- d-BiNI



BiNI



SMPL-X depth



d-BiNI

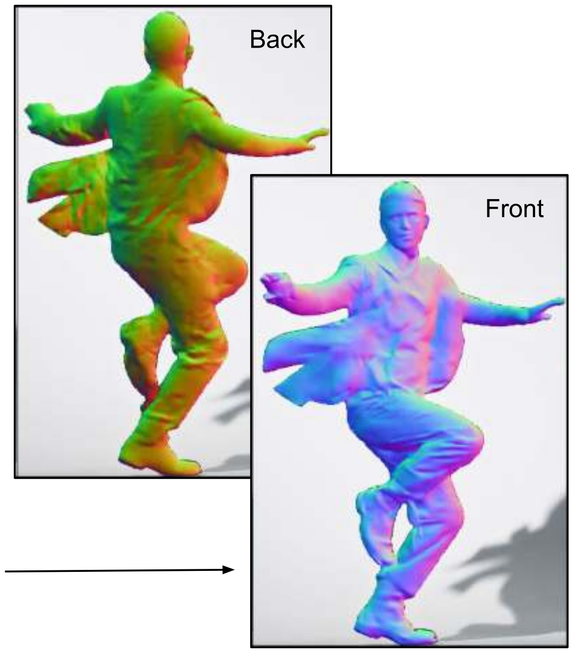
Overview of ECON

Input Image

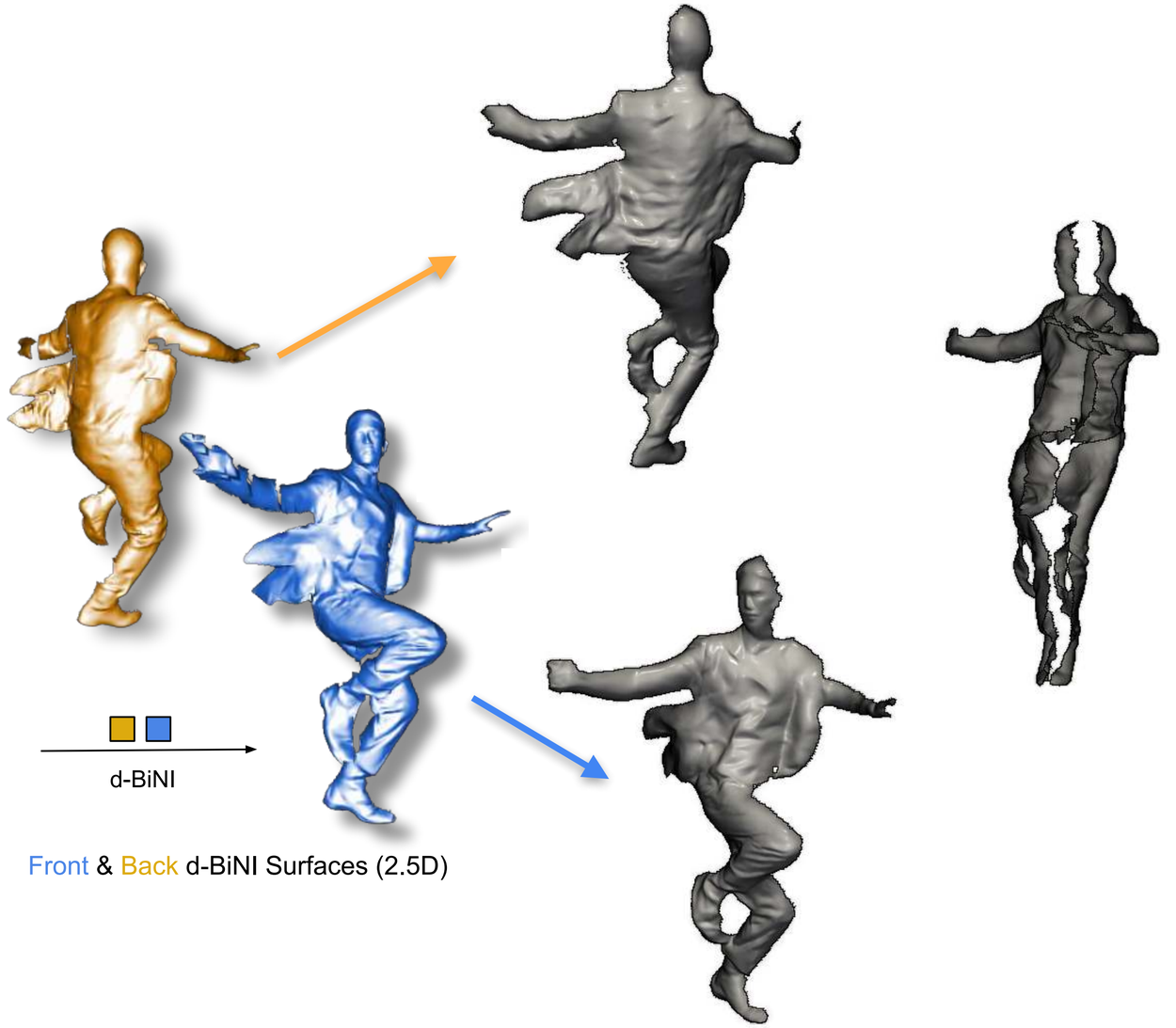
Front & Back Normal Estimation (2D)

Overview of ECON

Input Image

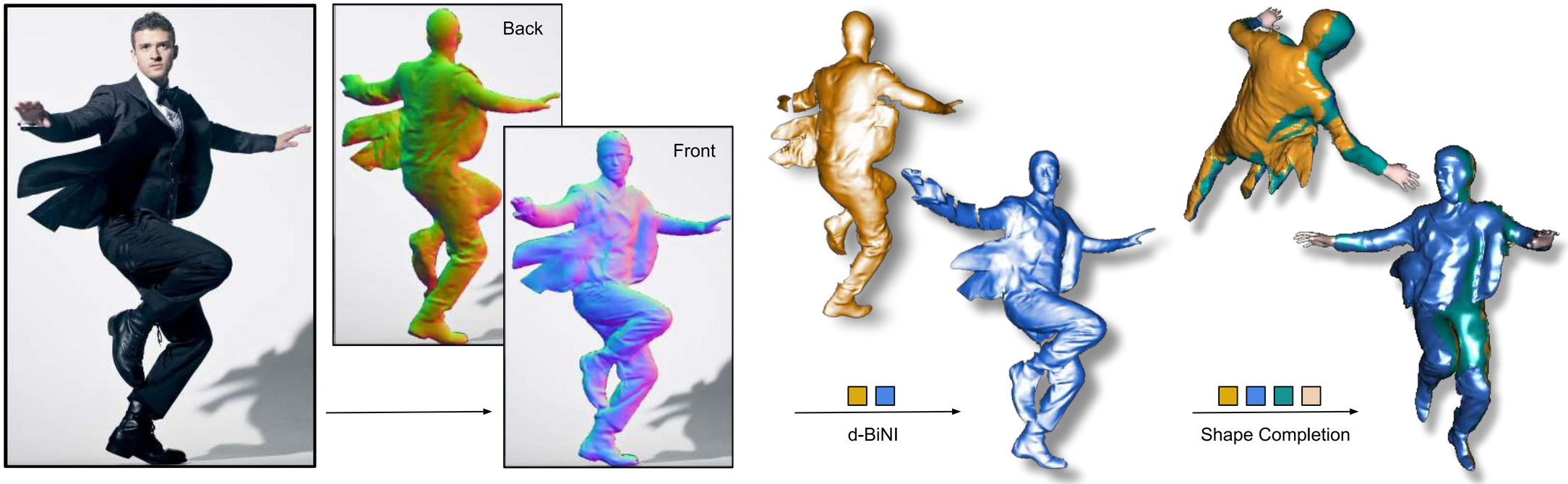


Front & Back Normal Estimation (2D)



Front & Back d-BiNI Surfaces (2.5D)

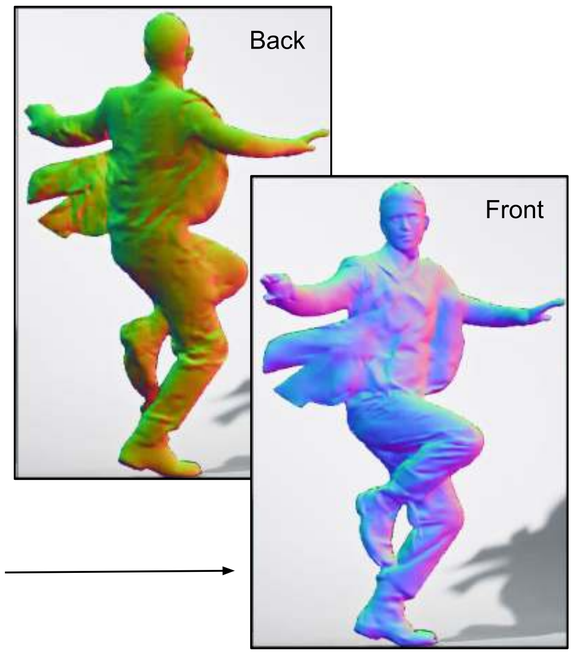
Overview of ECON



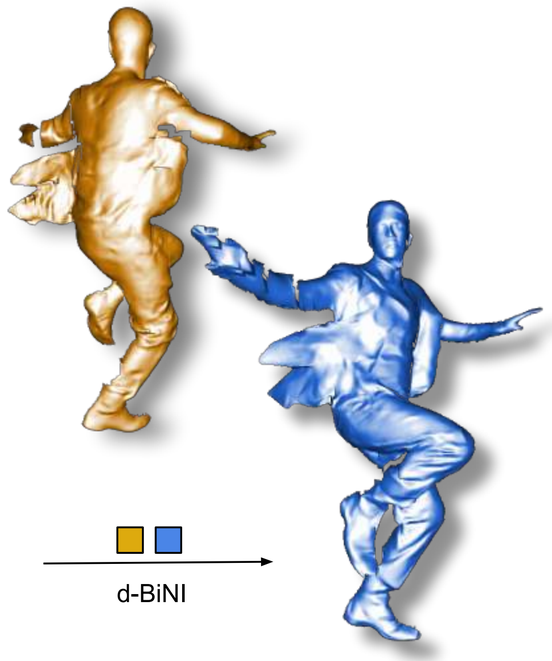
Input Image | Front & Back Normal Estimation (2D) | Front & Back d-BiNI Surfaces (2.5D) | Full Clothed Human Mesh (3D)

Overview of ECON

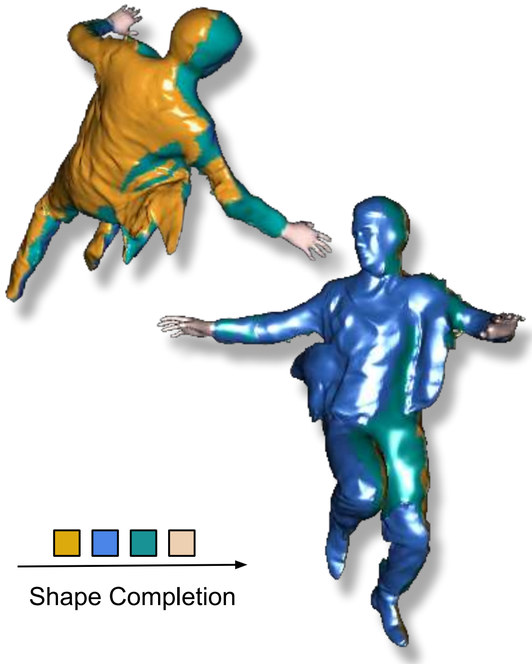
Input Image



Front & Back Normal Estimation (2D)

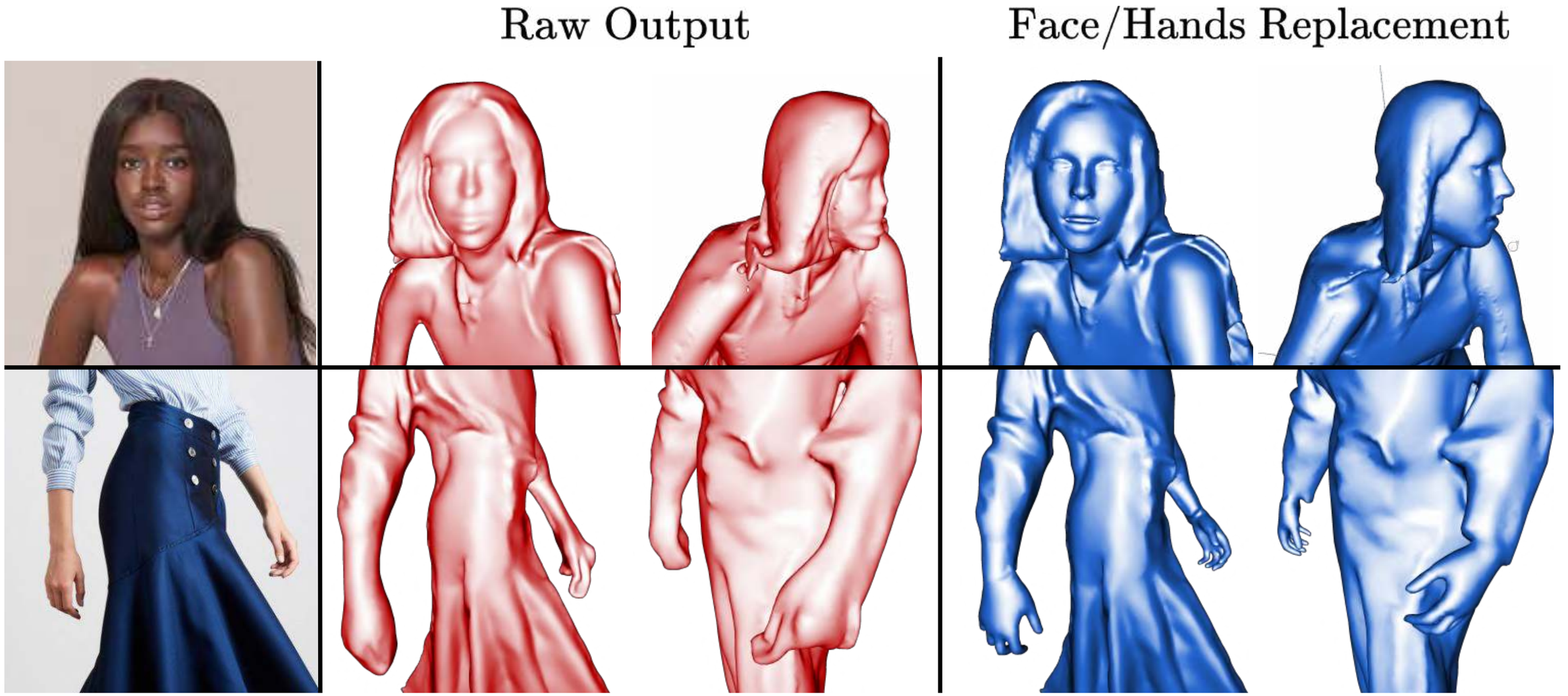


Front & Back d-BiNI Surfaces (2.5D)



Full Clothed Human Mesh (3D)

Face/Hands Replacement of **ECON**



PIFuHD vs ECON on Loose Clothes

PIFuHD vs ECON on Loose Clothes

PIFuHD vs ECON on Loose Clothes

PIFuHD vs ECON on Unseen Poses

PIFuHD vs ECON on Unseen Poses

PaMIR vs ECON on Unseen Poses

ICON vs ECON on Unseen Poses

PaMIR vs ECON on Loose Clothes

ICON vs ECON on **Loose Clothes**

ECON vs Others

Unseen Poses + Loose Clothes

PIFuHD

PaMIR

ICON

ECON vs Others

Unseen Poses + Loose Clothes

PIFuHD

ICON

PaMIR

ECON vs Others

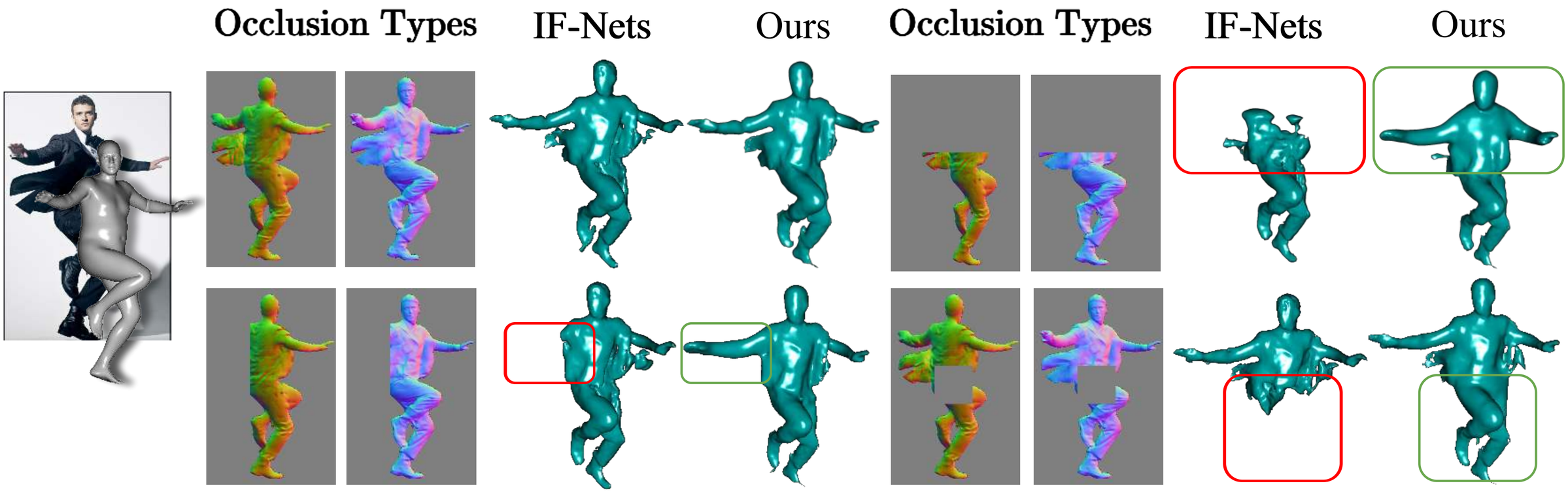
Unseen Poses + Loose Clothes

PIFuHD

ICON

PaMIR

ECON on Partial Images

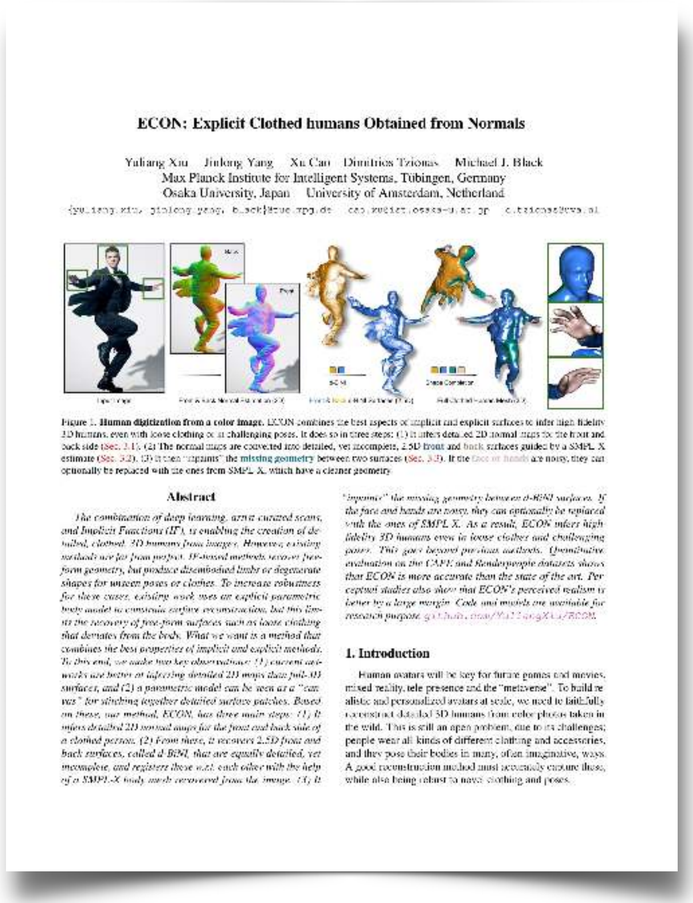


Summary (Pros, Cons)

Loose Clothing

Challenging Pose

econ.is.tue.mpg.de



ECON: Explicit Clothed humans Obtained from Normals

Yuliang Xiu Jinlong Yang Xu Cao Dimitrios Tzionas Michael J. Black
Max Planck Institute for Intelligent Systems, Tübingen, Germany
Osaka University, Japan University of Amsterdam, Netherland

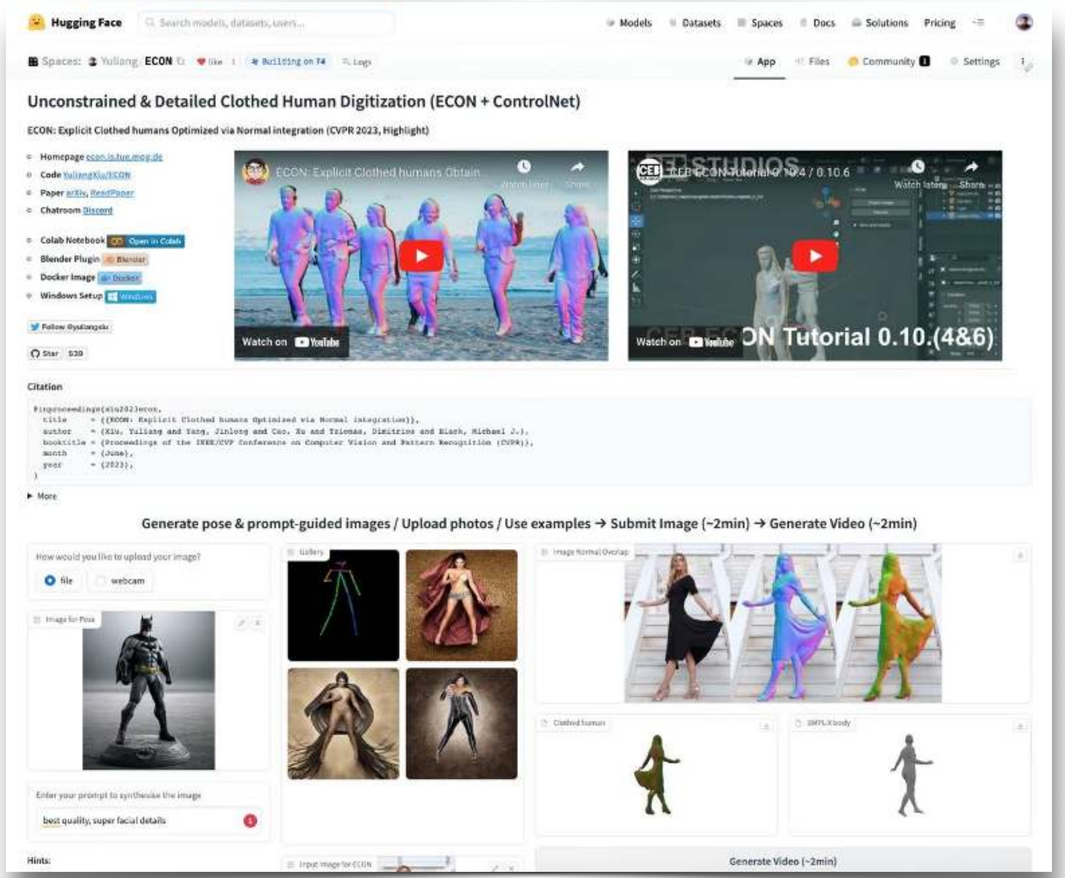
Figure 1. Human digitization from a color image. ECON combines the best aspects of implicit and explicit surfaces to infer high fidelity 3D humans, even with loose clothing or in challenging poses. It does so in three steps: (1) it infers detailed 2D normal maps for the front and back side (Sec. 3.1), (2) the normal maps are converted into detailed, yet incomplete, 2.5D front and back surfaces guided by a SMPL-X estimate (Sec. 3.2), (3) it then "inpaints" the missing geometry between two surfaces (Sec. 3.3). If the face or hands are noisy, they can optionally be replaced with the ones from SMPL-X, which have a cleaner geometry.

Abstract

The combination of deep learning, artist curated scans, and Implicit Functions (IF), is enabling the creation of detailed, clothed, 3D humans from images. However, existing methods are far from perfect. IF-based methods recover free-form geometry, but produce disembodied limbs or degenerate shapes for unseen poses or clothes. To increase robustness for these cases, existing work uses an explicit parametric body model to constrain surface reconstruction, but this limits the recovery of free-form surfaces such as loose clothing that deviates from the body. What we want is a method that combines the best properties of implicit and explicit methods. To this end, we make two key observations: (1) current architectures can better exploit explicit 2D maps than 3D voxels or points, and (2) a parametric model can be seen as a "canvas" for stitching together detailed surface patches. Based on these, our method, ECON, has three main steps: (1) It infers detailed 2D normal maps for the front and back side of a clothed person. (2) From these, it recovers 2.5D front and back surfaces, called d-BiNI, that are equally detailed, yet mutually inconsistent, and registers these w.r.t. each other with the help of a SMPL-X body mesh recovered from the image. (3) It "inpaints" the missing geometry between d-BiNI surfaces. If the face and hands are noisy, they can optionally be replaced with the ones of SMPL-X. As a result, ECON infers high fidelity 3D humans even in loose clothes and challenging poses. This goes beyond previous methods. Quantitative evaluation on the CAPE and Renderpeople datasets shows that ECON is more accurate than the state of the art. Perceptual studies also show that ECON's perceived realism is better by a large margin. Code and models are available for research purpose. github.com/YuliangXiu/ECON

1. Introduction

Human avatars will be key for future games and movies, mixed reality, tele-presence and the "metaverse". To build realistic and personalized avatars at scale, we need to faithfully reconstruct detailed 3D humans from color photos taken in the wild. This is still an open problem, due to its challenges; people wear all kinds of different clothing and accessories, and they pose their bodies in many, often imaginative, ways. A good reconstruction method must accurately capture those, while also being robust to novel clothing and poses.



GitHub YuliangXiu/ECON

Acknowledgments: *This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No.860768 (**CLIPE project**).*

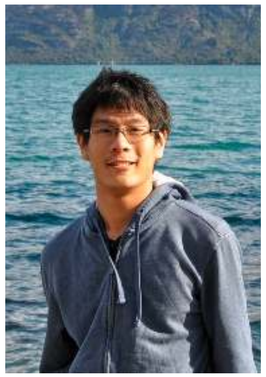
Disclosure: *MJB has received research gift funds from Adobe, Intel, Nvidia, Meta/Facebook, and Amazon. MJB has financial interests in Amazon, Datagen Technologies, and Meshcapade GmbH. While MJB is a consultant for Meshcapade, his research in this project was performed solely at, and funded solely by, the Max Planck Society.*

Yuliang/ECON

Acknowledgments

Yao Feng

Weiyang Liu



Haven Feng

Radek Daněček

Lea Hering

Tsvetelina Alexiadis

Taylor McConnell

Benjamin Pellkofer