DoubleFusion:
Real-Time Capture of Human Performances with Inner Body Shapes from a Single Depth Sensor

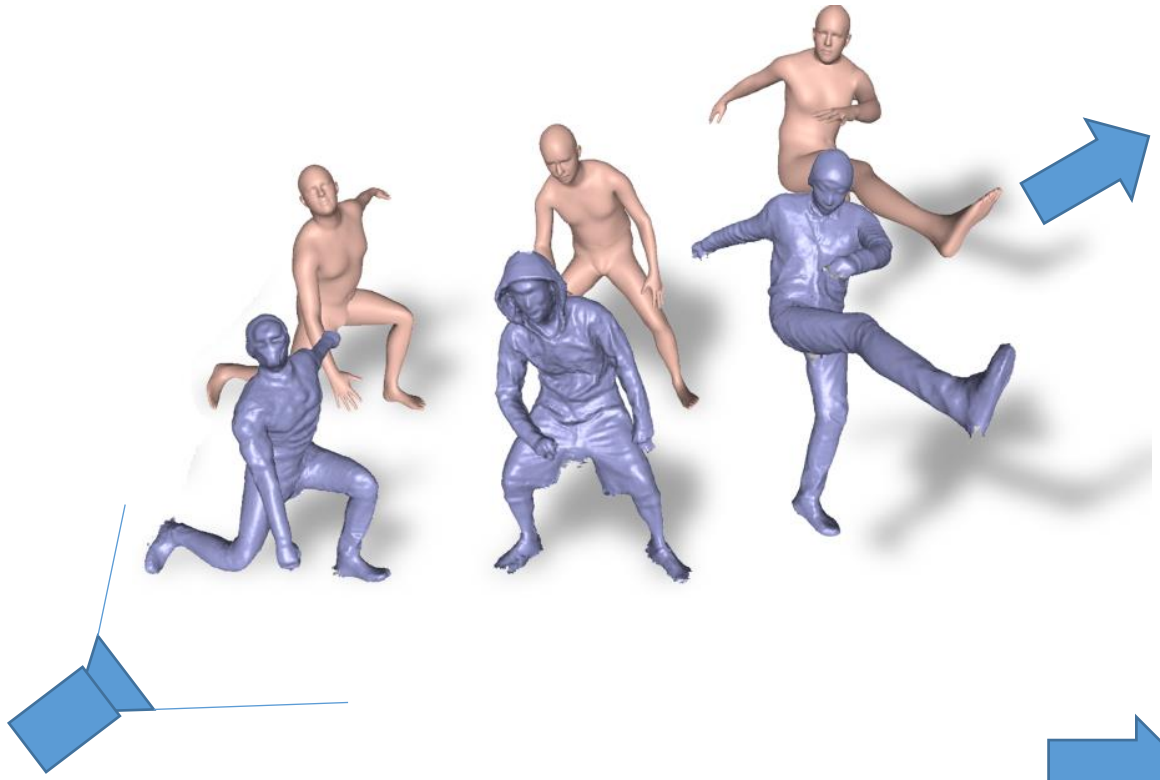
Tao Yu[1,2] , Zerong Zheng[1], Kaiwen Guo[1,3], Jianhui Zhao[2], Qionghai Dai[1],
Hao Li[4], Gerard Pons-Moll[5], Yebin Liu[1]

Google

USC University of Southern California

max planck institut informatik

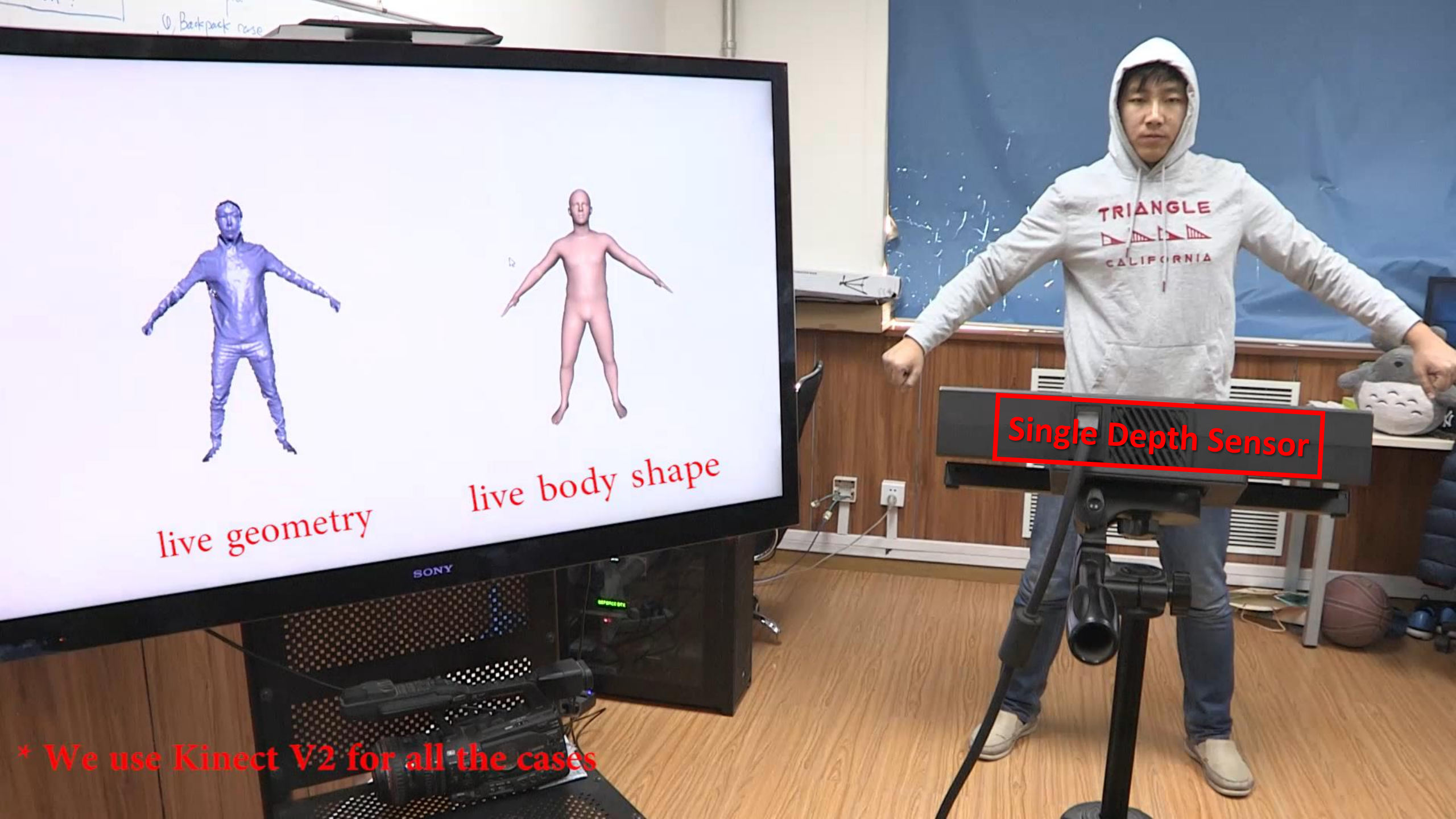
Motivation



Better Single-View Performance Capture Algorithm

Multi-View Human Performance Capture

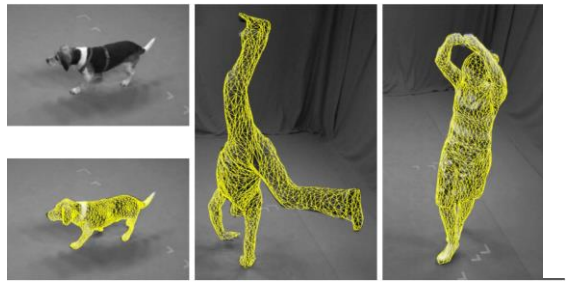
Convenient Human Performance Capture

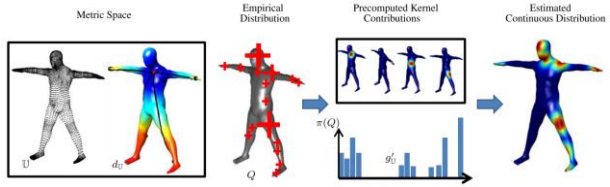

live geometry
live body shape
Single Depth Sensor
* We use Kinect V2 for all the cases

Related Work

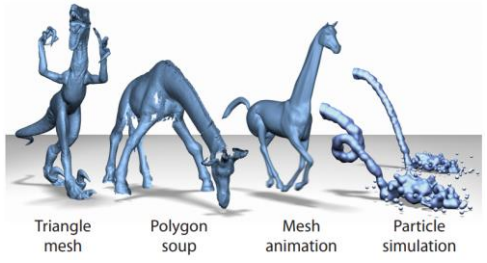
Static Template Based Dynamic Reconstruction Accurate

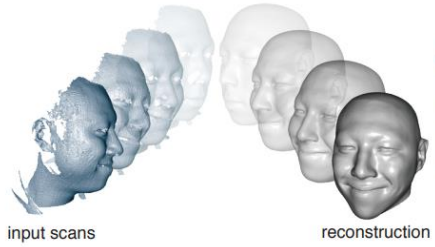
[Static Template with Embedded Skeleton]
Vlasic et al. 2008
Gall et al. 2009
Liu et al. 2011
Taylor et al. 2012
Pons-Moll et al. 2015
Ye et al. 2012

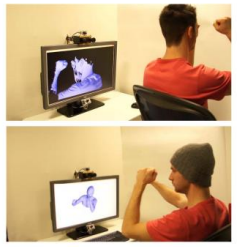
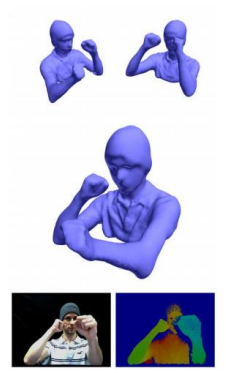





[Static Template with Non-rigid Deformation]
Sumner et al. 2007
Li et al. 2009
Guo et al. 2015
Zollhofer et al. 2014





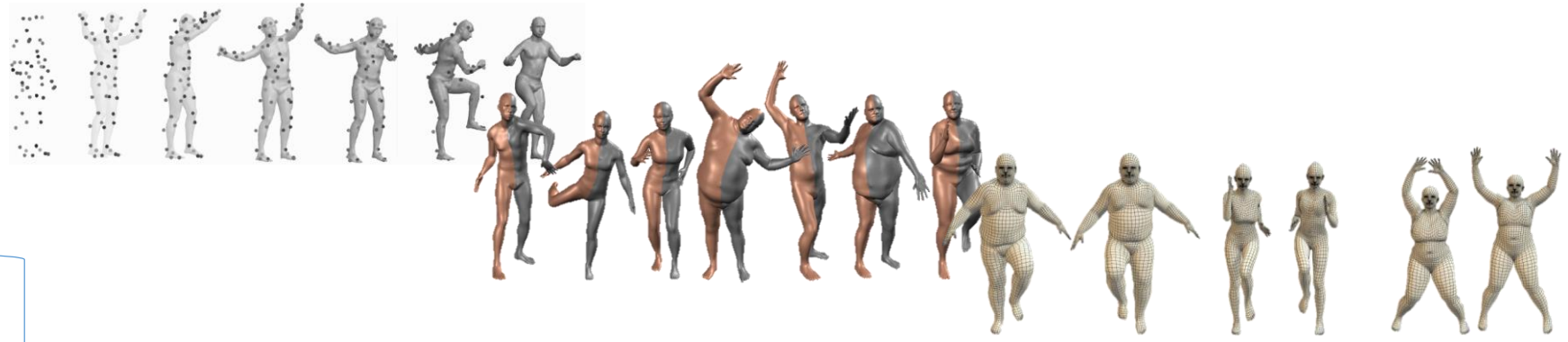




Need to pre-scan a static template, and/or embed skeleton into the template manually

Related Work

Model Based Dynamic Reconstruction

Semantic

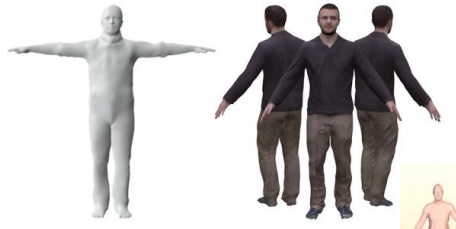


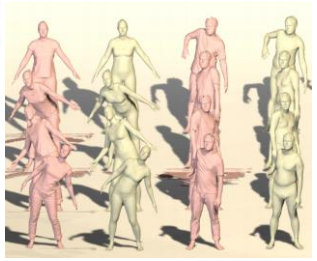
Anguelov et al. [SCAPE] 2005
Loper et al. [SMPL] 2015
Pons-Moll et al. [DYNA] 2015

Ye et al. 2014
Chen et al. 2016
Bogo et al. 2015
Bogo et al. 2016
Alldieck et al. 2018
Zhang et al. 2017
Pons-Moll et al. 2017

Cannot reconstruct detailed geometry of clothing

Need dense 4D scan as input, offline





Related Work

Free Form Dynamic Reconstruction

General

Liao et al. 2009
Li et al. 2008
Wand et al. 2009
Chang et al. 2009
Chang et al. 2010
Pekelny et al. 2008
Mitra et al. 2007
Süßmuth et al. 2008
Sharf et al. 2008
Tevs et al. 2012
Li et al. 2013
Dou et al. 2013

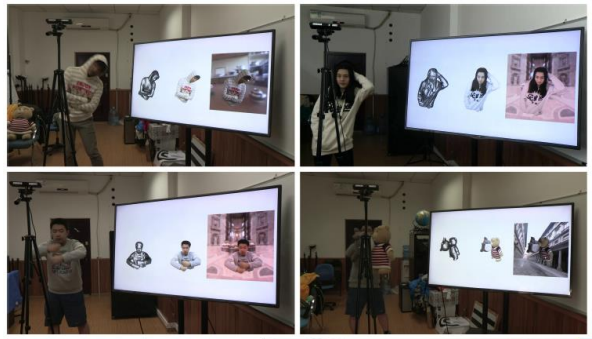
offline

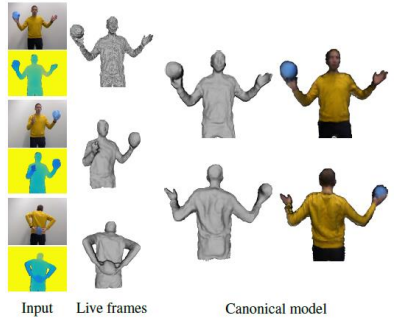
Newcombe et al. 2015
Innmann et al. 2016
Guo et al. 2017
Slavcheva et al. 2017
Dou et al. 2016
Dou et al. 2017
Yu et al. 2017
Slavcheva et al. 2018

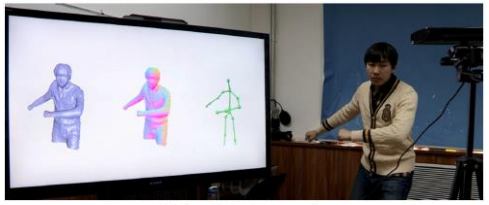
Cannot track fast motion

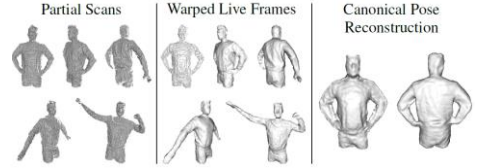
Need multi-view input

Cannot reconstruct inner body shape

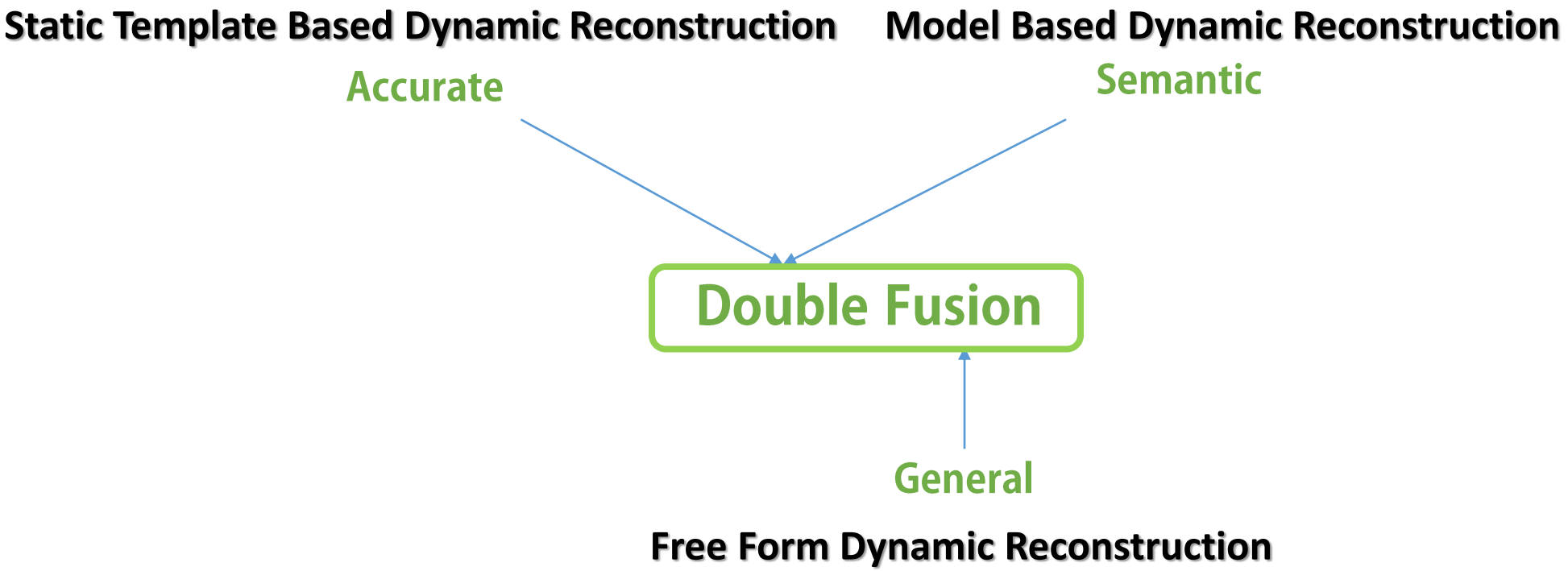




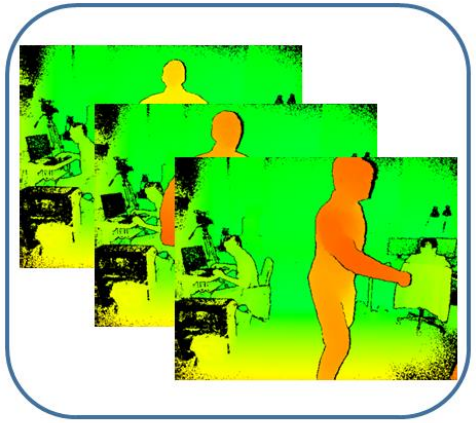




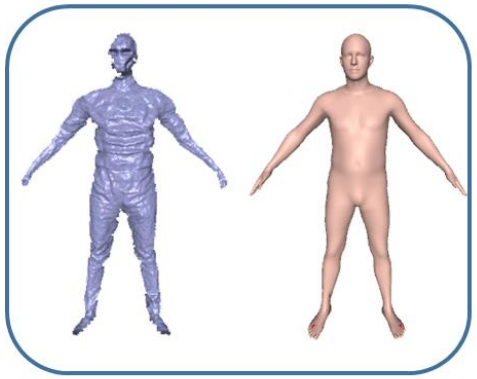
Related Work



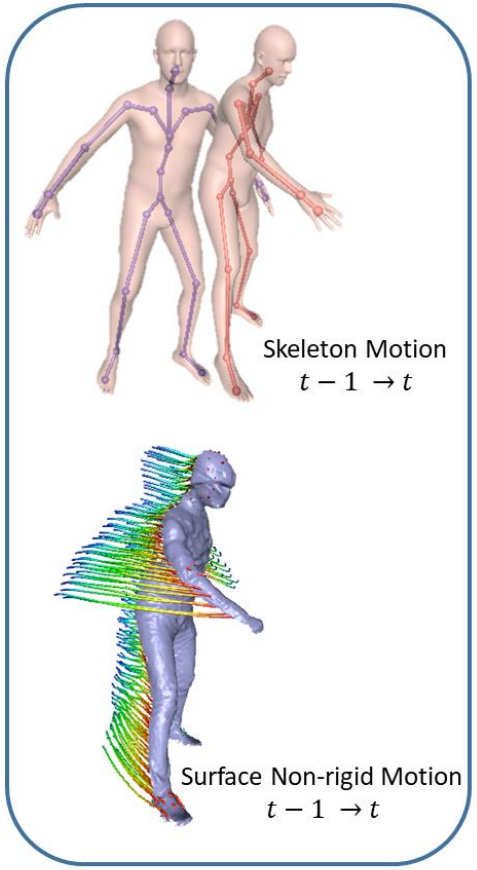
Overview



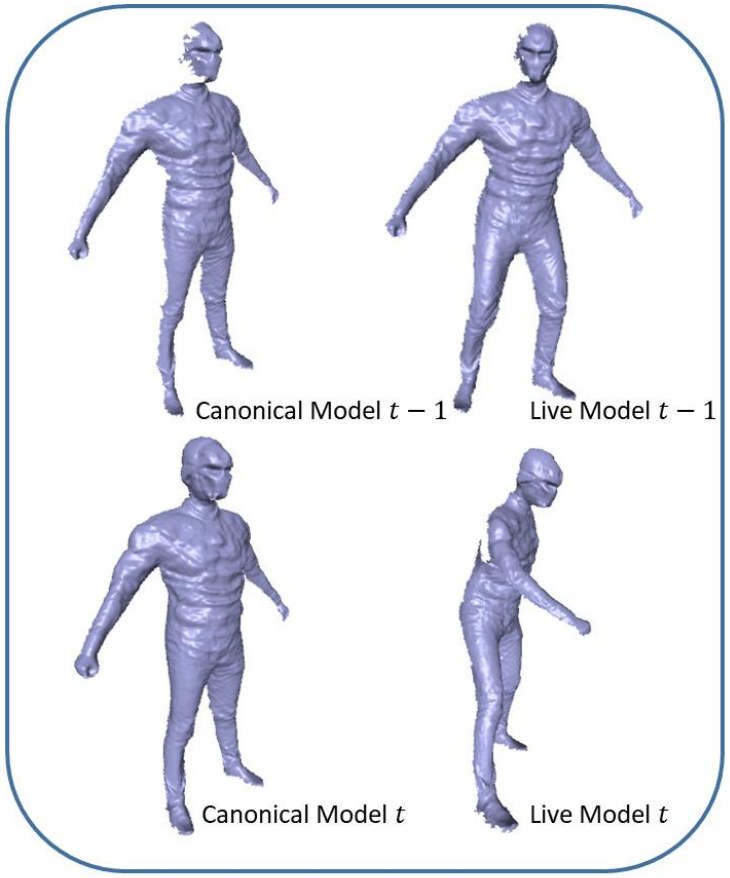
Input depth sequence



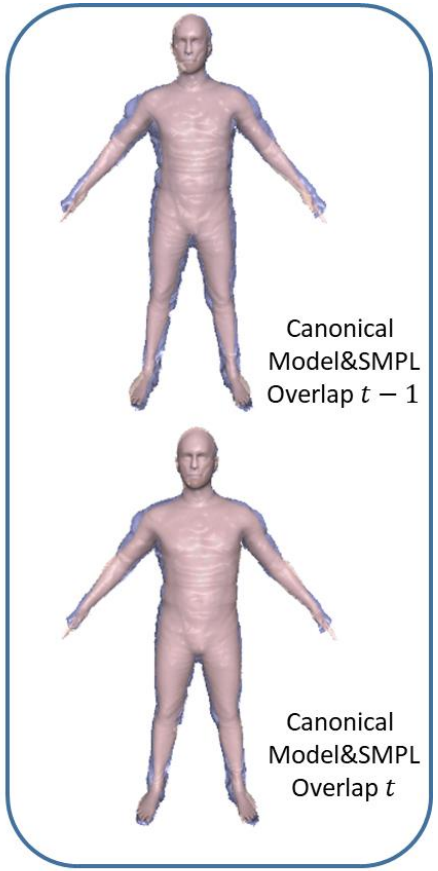
Initialization



Joint Motion Tracking



Geometry Fusion



Volumetric Shape-Pose Optimization

Double-layer Surface Representation

Double-Layer Surface

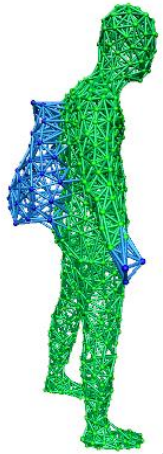
Inner Layer
[SMPL]

Outer Layer
[Fused Geometry]

Motion Description

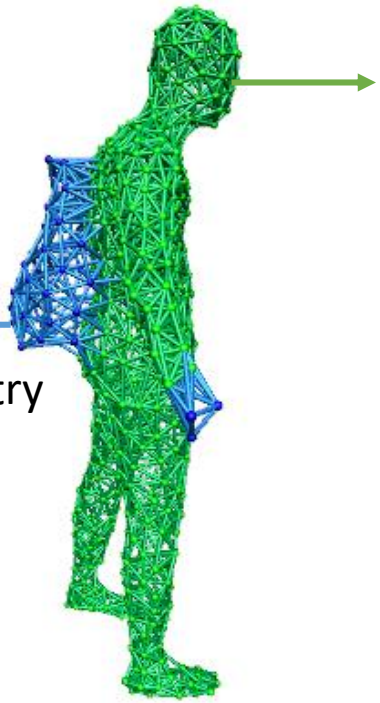
Skeleton Motion
[Kinematic Chain]

Non-rigid Motion
[Double Node Graph]



[Loper et al. 2015 SMPL: A Skinned Multi-Person Linear Model]

Double Node Graph

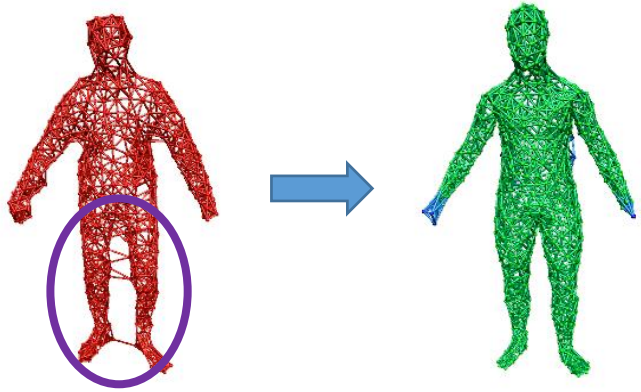


On-body node graph:

- Non-rigid deformation of near-body geometry
- Predefined on SMPL model
- Prevent erroneous connections between body parts
- Bind to skeleton for joint motion tracking

Far-body node graph:

- Non-rigid deformation of far-body geometry
- Uniformly sampled on far-body geometry
- Connected with on-body nodes

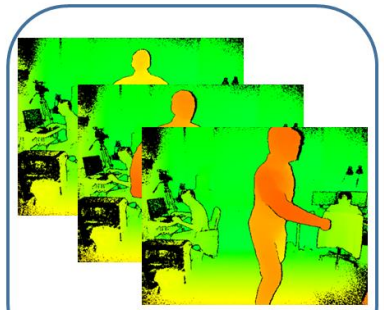


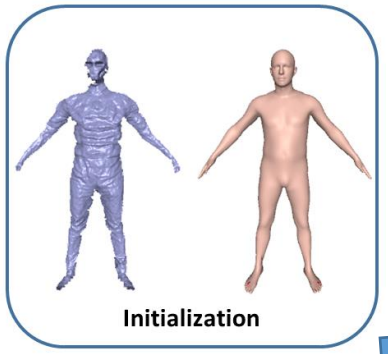
Near-body geometry: fused geometry area that near the on-body node graph.

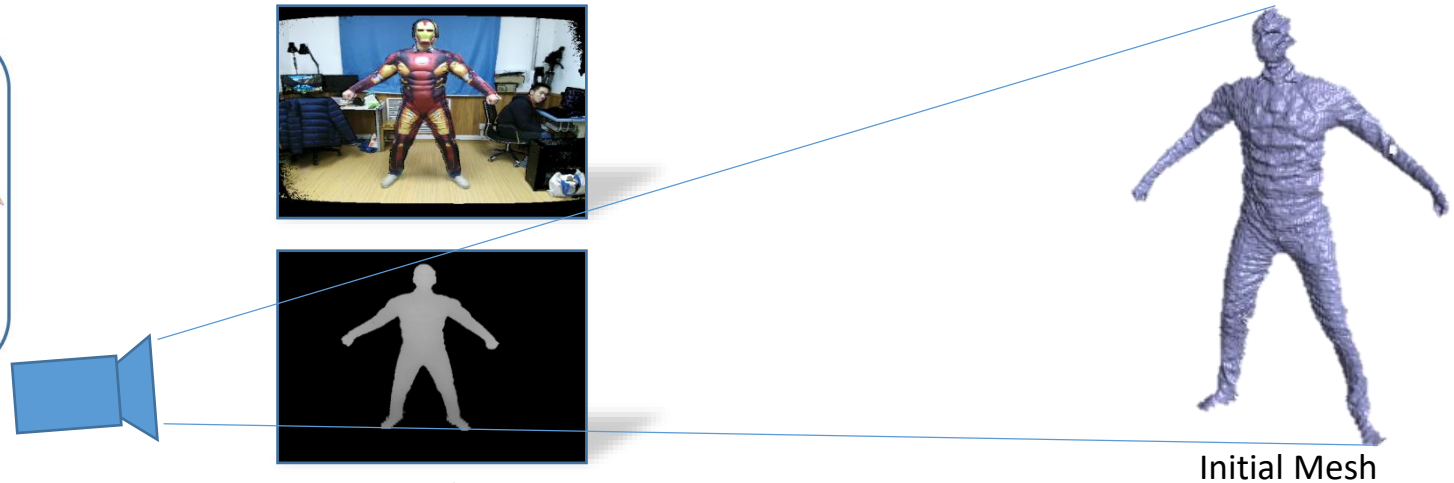
Far-body geometry: fused geometry area that far from the on-body node graph, like backpack etc.

Initialization

- Rough A Pose
- Initialize TSDF Volume & Extract Mesh

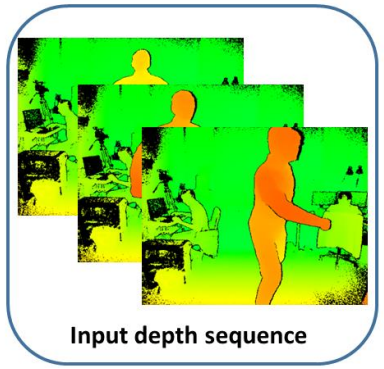


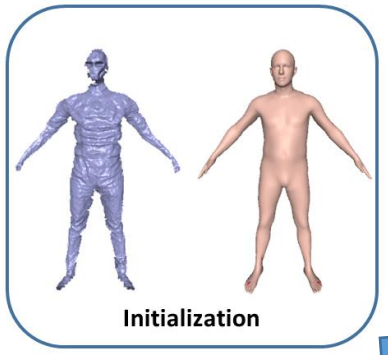


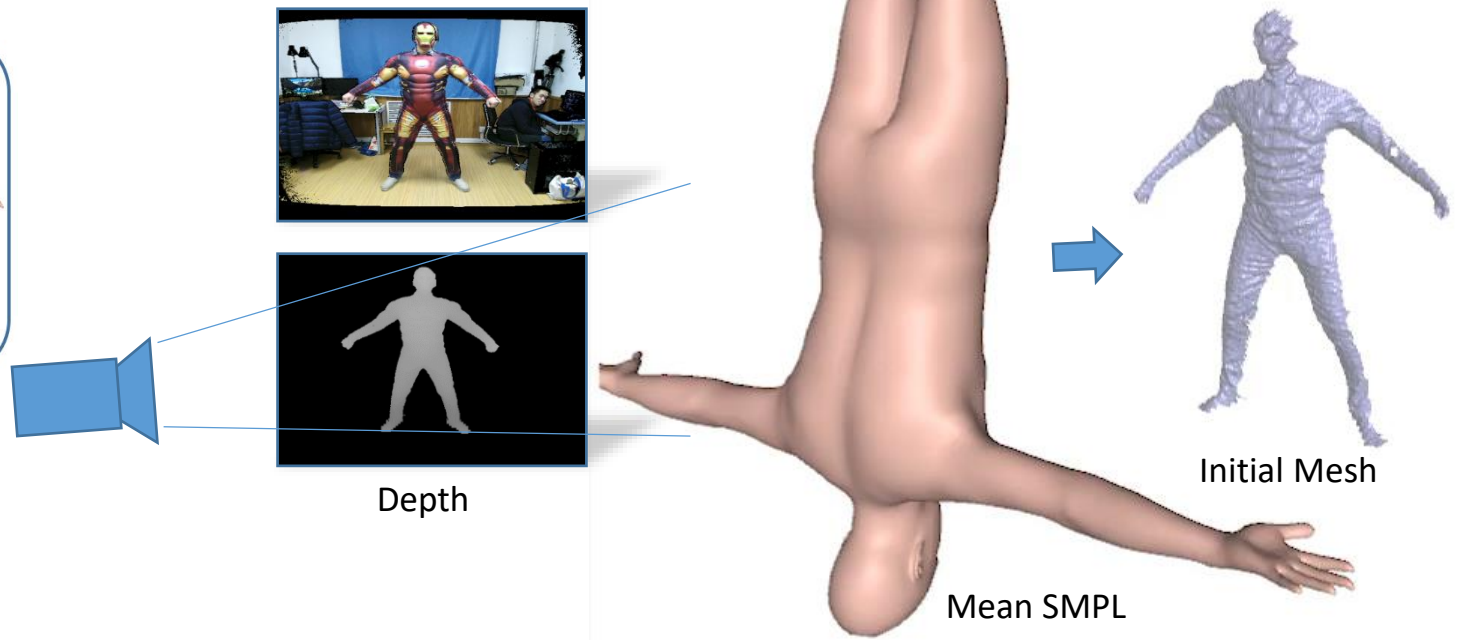


Initialization

- Rough A Pose
- Initialize TSDF Volume & Extract Mesh
- Estimate initial shape β_0 and pose θ_0

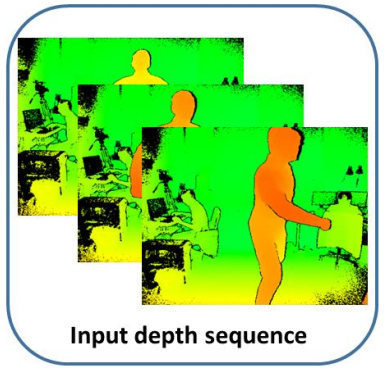


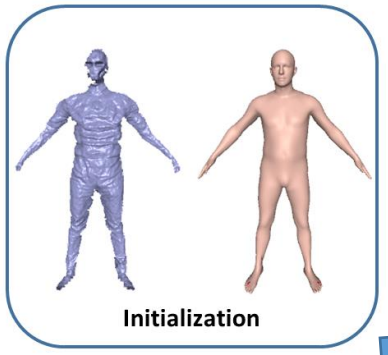


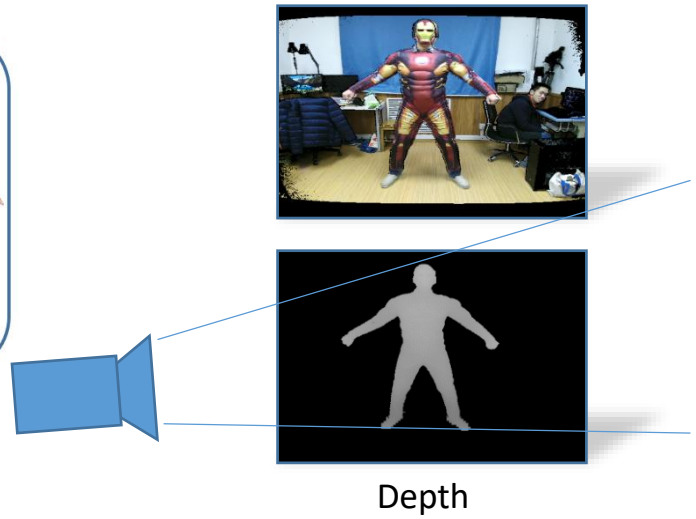


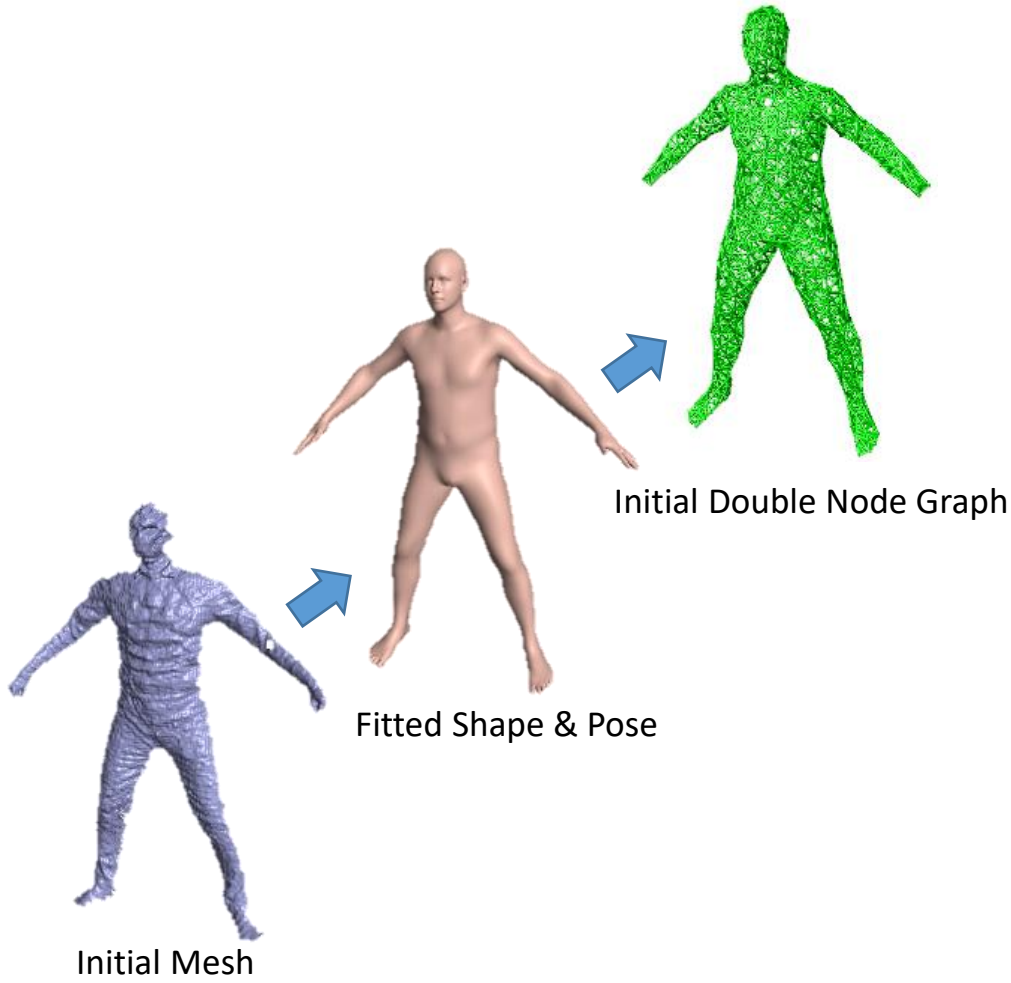
Initialization

- Rough A Pose
- Initialize TSDF Volume & Extract Mesh
- Estimate initial shape β_0 and pose θ_0
- Initialize Double Node Graph

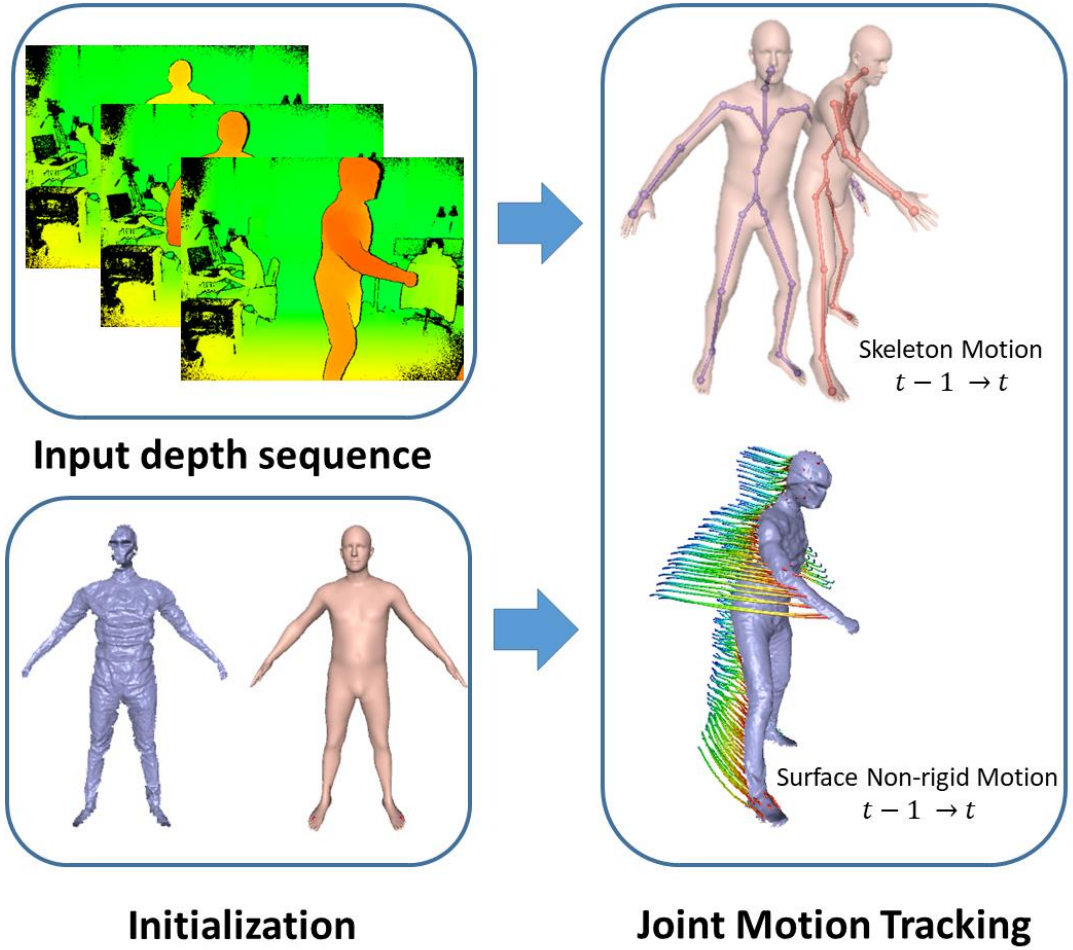




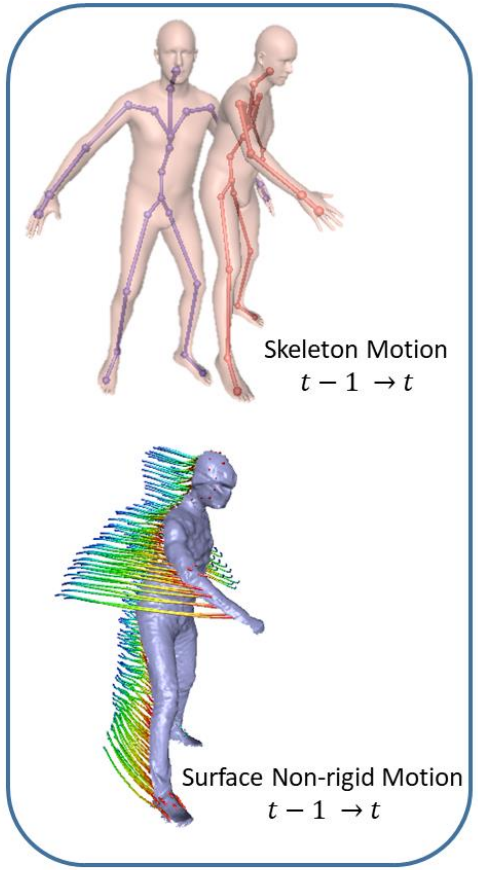




Joint Motion Tracking



Joint Motion Tracking

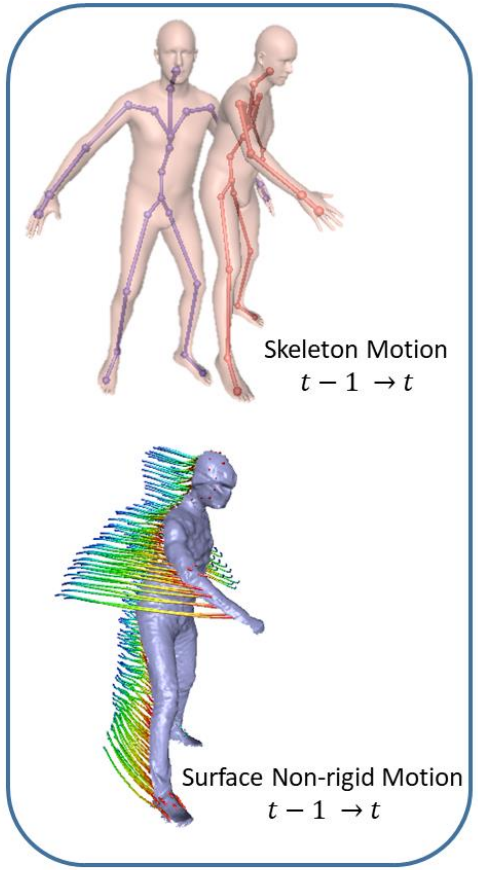


Joint Motion Tracking

$$E_{mot} = \lambda_{data} E_{data} + \lambda_{\text{bind}} \text{E}_{\text{bind}} + \lambda_{reg} E_{reg} + \lambda_{prior} \text{E}_{\text{prior}}$$

Double-Layer Surface	Skeleton Motion	Non-rigid Motion	Skeleton Motion
↓ *fit*	↕ *constraint*	↓ *constraint*	↓ *constraint*
Input Depth Map	Non-rigid Motion	Local ARAP	Natural Poses

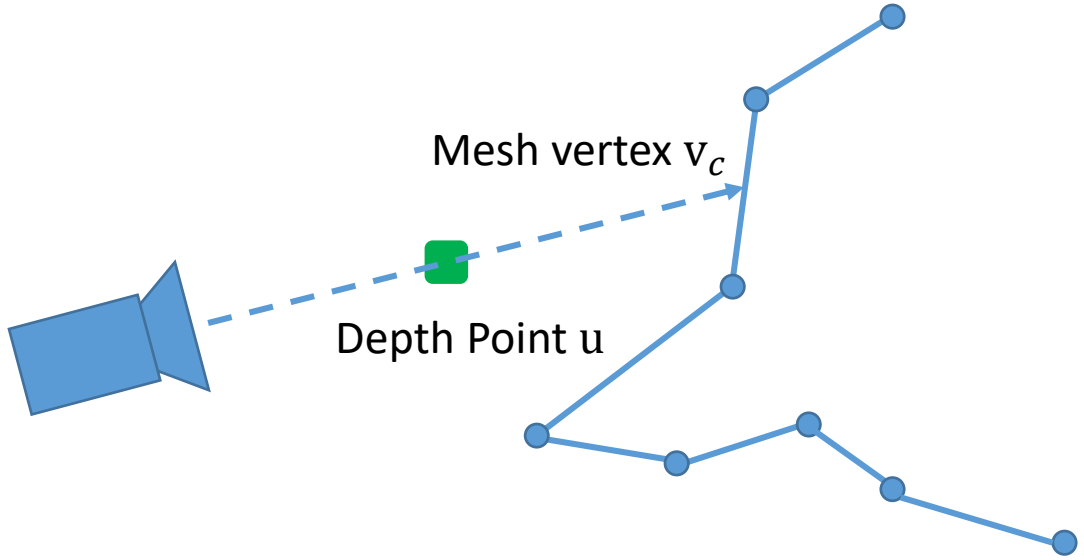
Joint Motion Tracking



Joint Motion Tracking

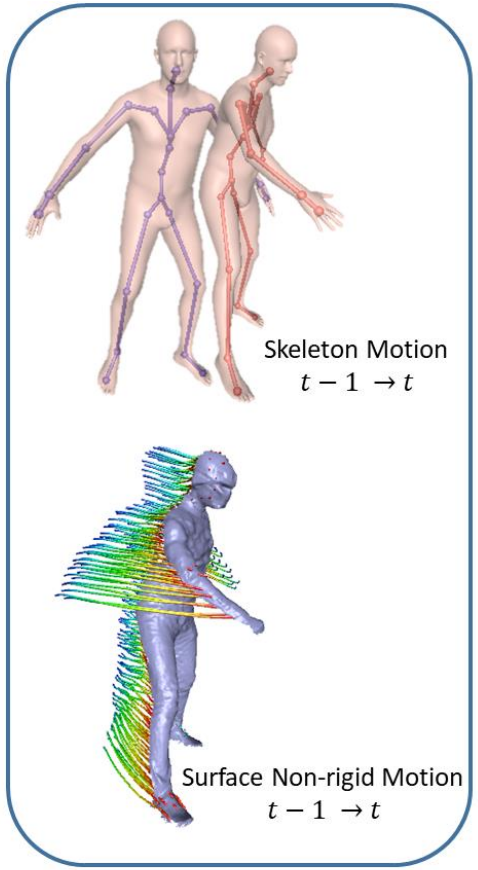
$$E_{data} = \tau_1(\mathrm{v}_c) * \psi\left(\tilde{n}_{v_c}^{\mathrm{T}}(\tilde{\mathrm{v}}_c - \mathrm{u})\right) + \left(\tau_2(\mathrm{v}_c) + \tau_3(\mathrm{v}_c)\right) * \psi\left(\hat{n}_{v_c}^{\mathrm{T}}(\hat{\mathrm{v}}_c - \mathrm{u})\right)$$

Non-rigid tracking data term



$\psi(\cdot)$: *robust Geman-McClure penalty function*
v_c: *projected mesh vertex corresponding to* u
$\tilde{n}_{v_c}$, $\tilde{\mathrm{v}}_c$: *normal & position of* v_c *after non-rigid warping*

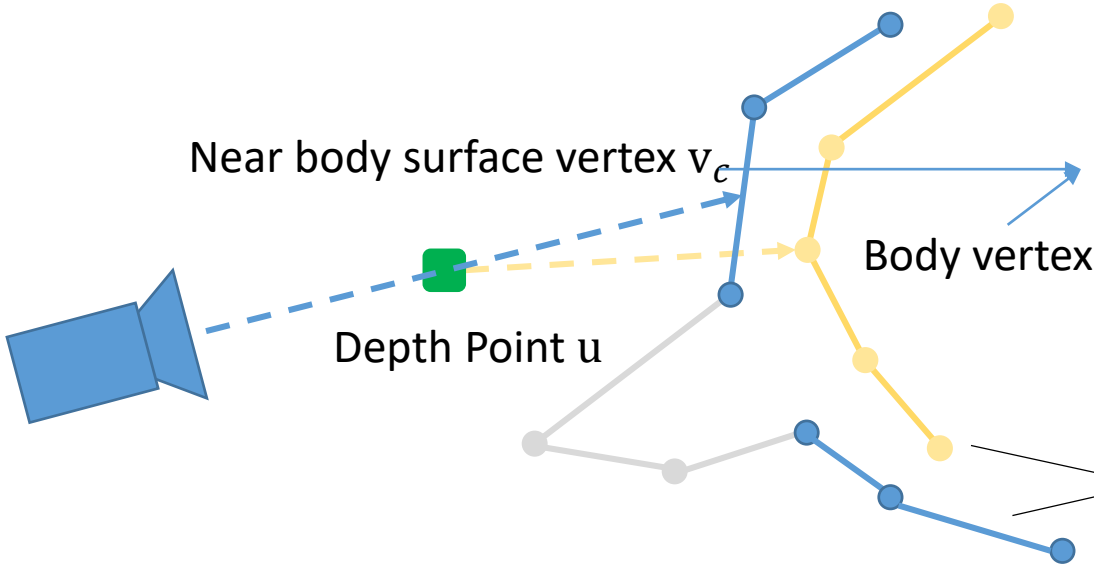
Joint Motion Tracking



Joint Motion Tracking

$$E_{data} = \tau_1(\mathrm{v}_c) * \psi\left(\tilde{n}_{v_c}^{\mathrm{T}}(\tilde{\mathrm{v}}_c - \mathrm{u})\right) + \left(\tau_2(\mathrm{v}_c) + \tau_3(\mathrm{v}_c)\right) * \psi\left(\hat{n}_{v_c}^{\mathrm{T}}(\hat{\mathrm{v}}_c - \mathrm{u})\right)$$

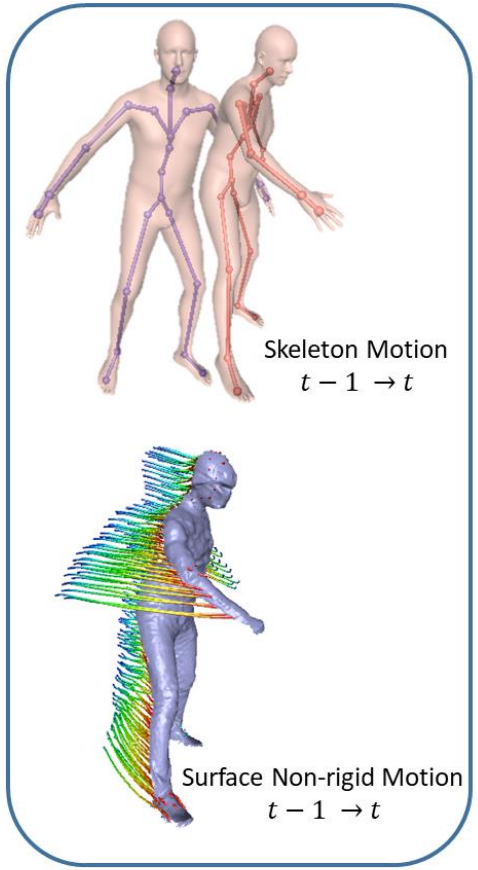
Skeleton tracking data term



$$c = \underset{i \in \{t,s\}}{\operatorname{argmax}} \left(\left(1 - \frac{\|\mathrm{v}_i - \mathrm{u}\|_2}{\delta_{\max}}\right)^2 + \mu \tilde{\mathrm{n}}_{v_i}^{\mathrm{T}} \mathrm{n}_u \right)$$

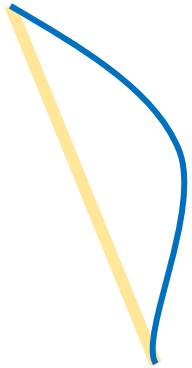
$\psi(\cdot)$: *robust Geman-McClure penalty function*
v_c: *projected mesh vertex / nearest body vertex corresponding to* u
$\hat{n}_{v_c}$, $\hat{\mathrm{v}}_c$: *normal & position of* v_c *after skinning*

Joint Motion Tracking

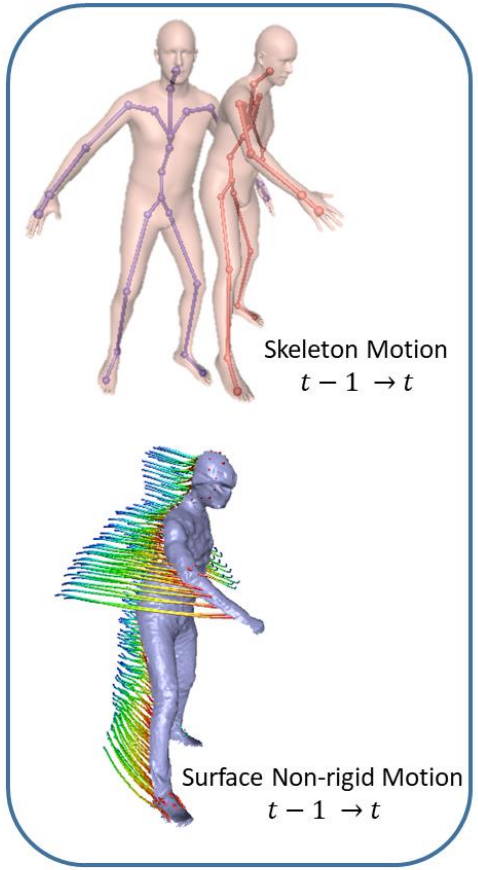


Joint Motion Tracking

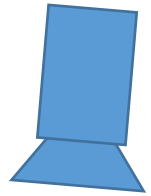
Double-Layer Surface

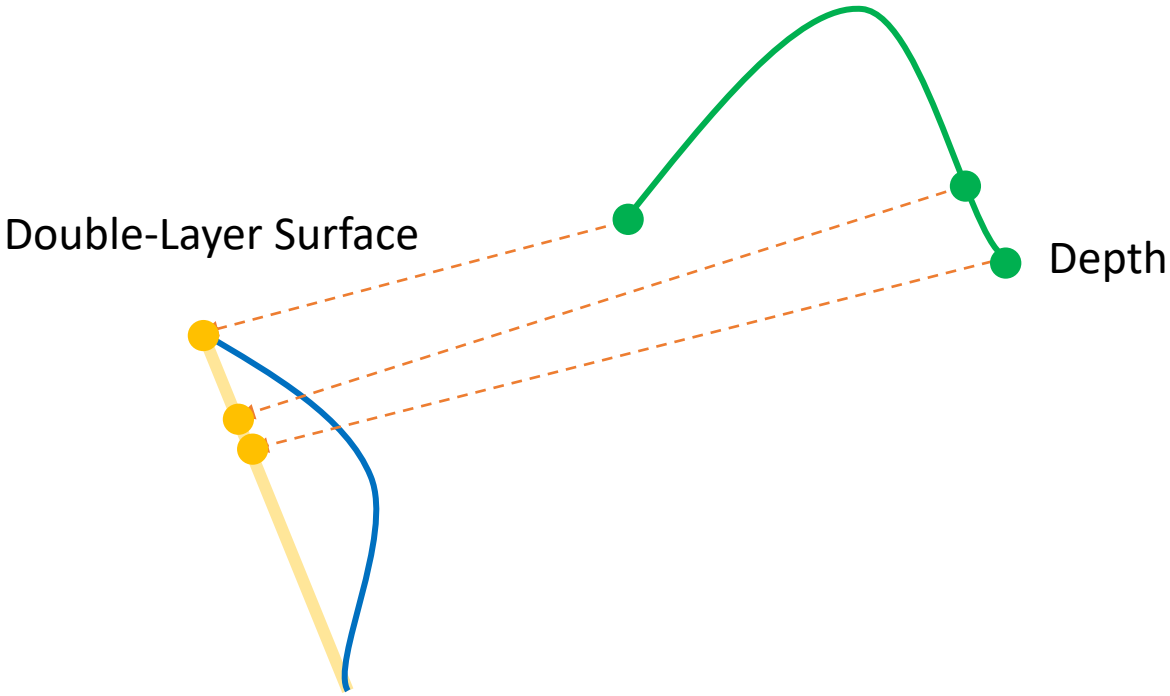


Joint Motion Tracking

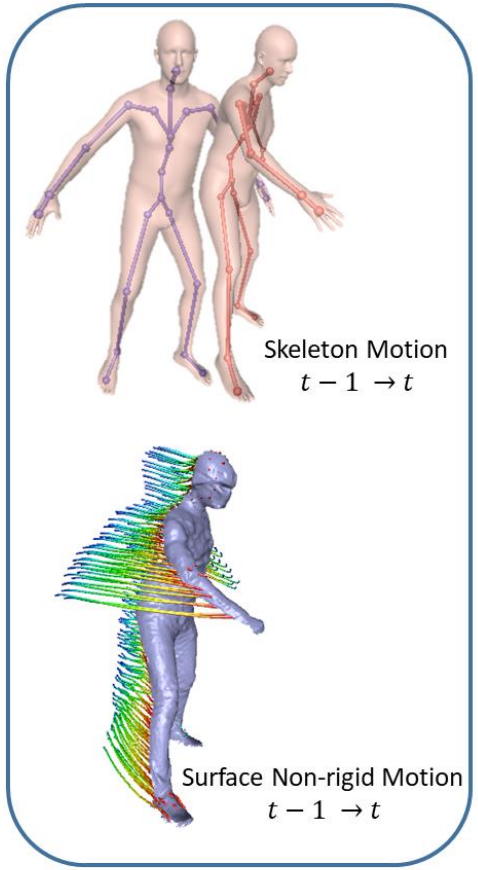


Joint Motion Tracking

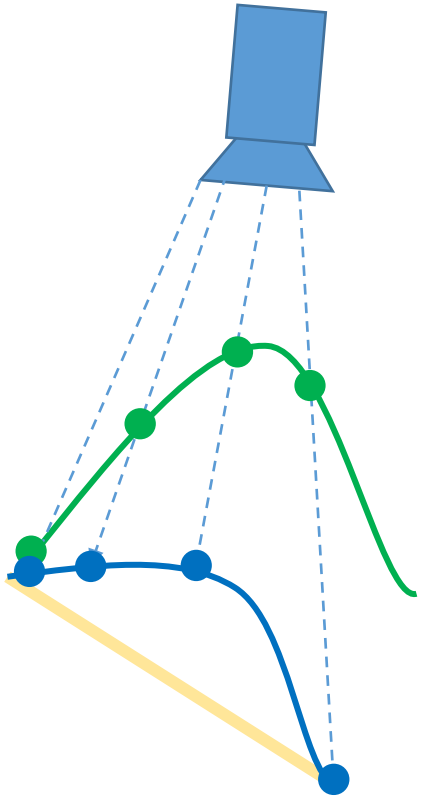




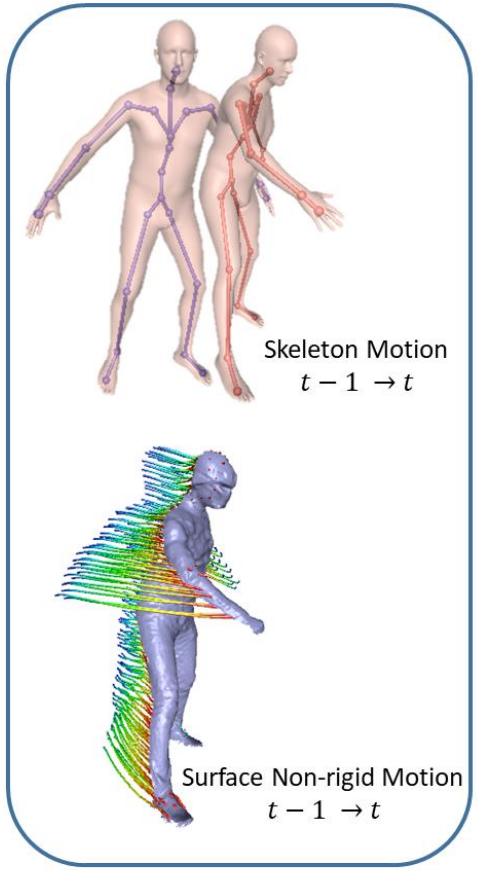
Joint Motion Tracking



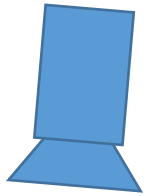
Joint Motion Tracking

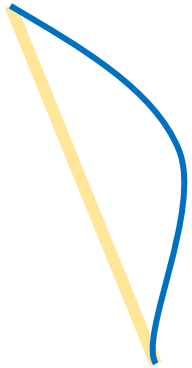


Joint Motion Tracking

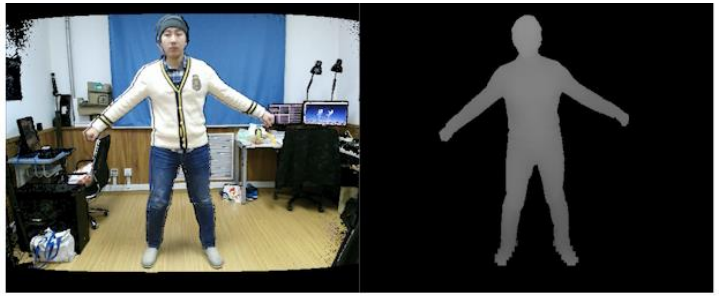


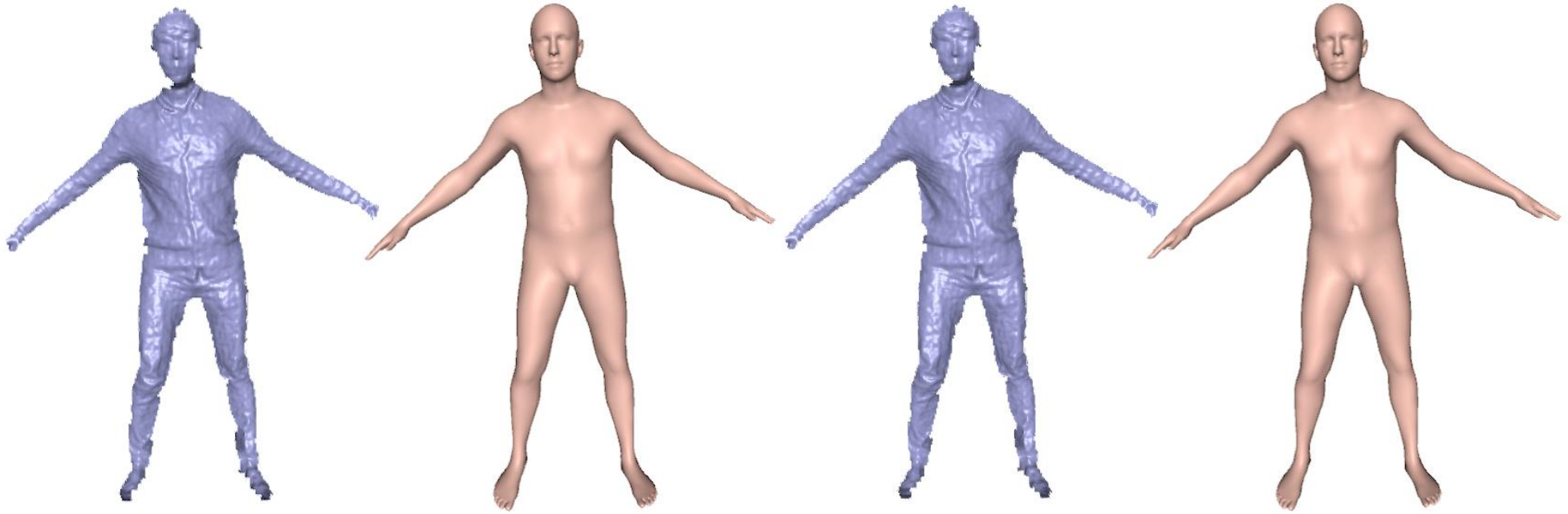
Joint Motion Tracking





Joint Motion Tracking: Evaluation

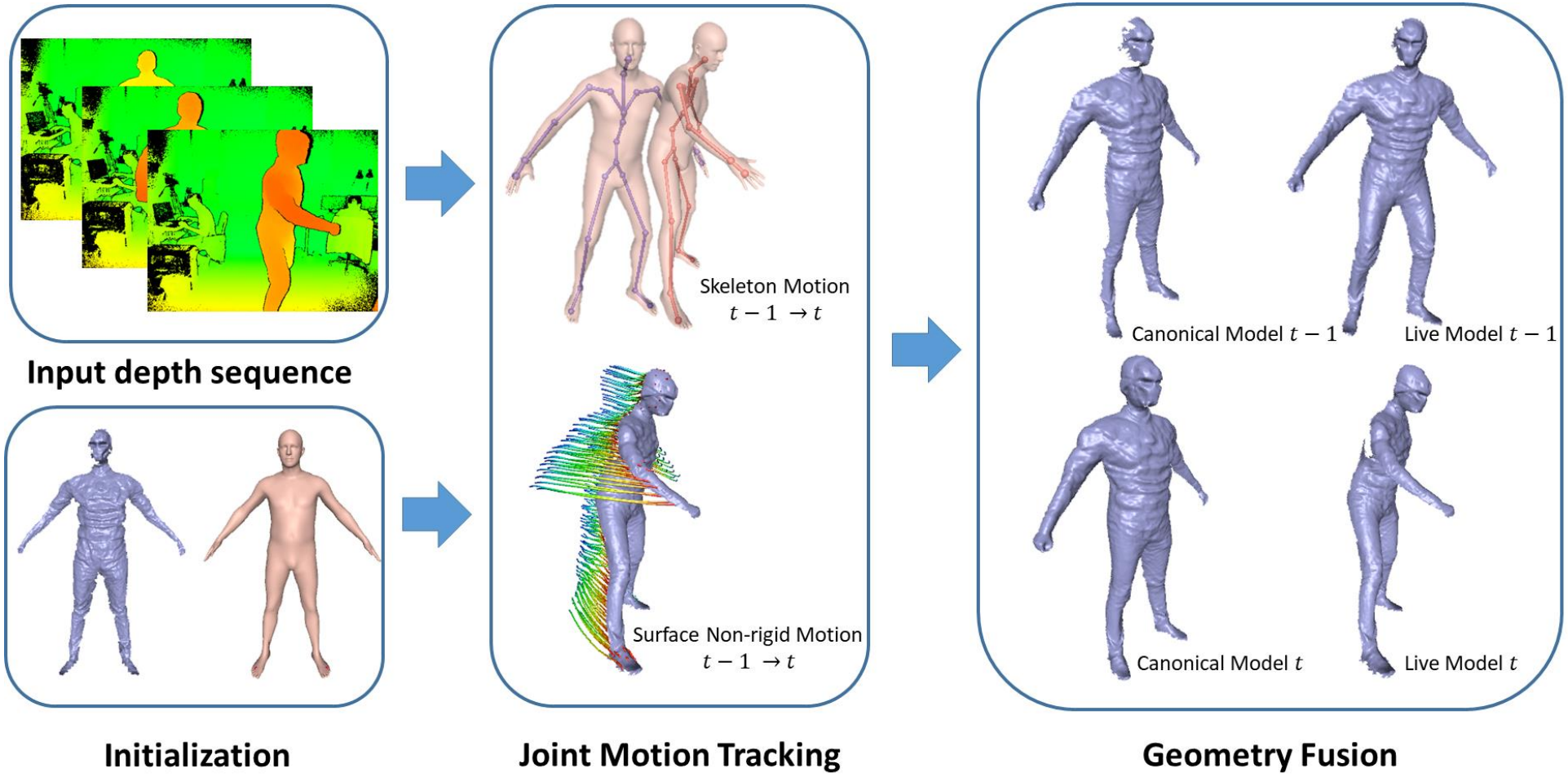




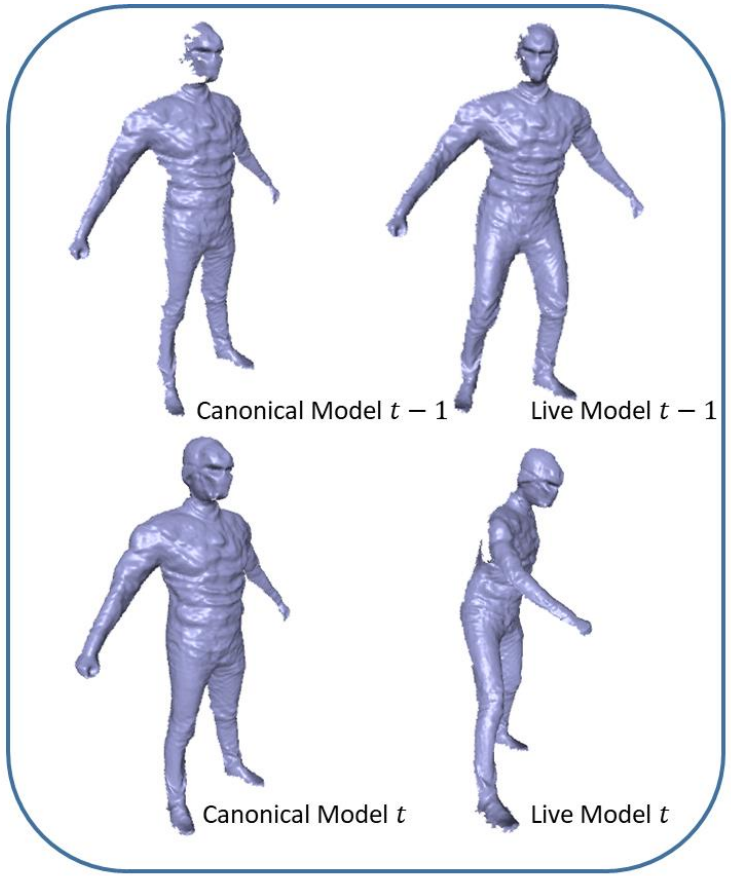
without on body correspondences

with on body correspondences

Geometry Fusion



Geometry Fusion



Geometry Fusion

Input Depth

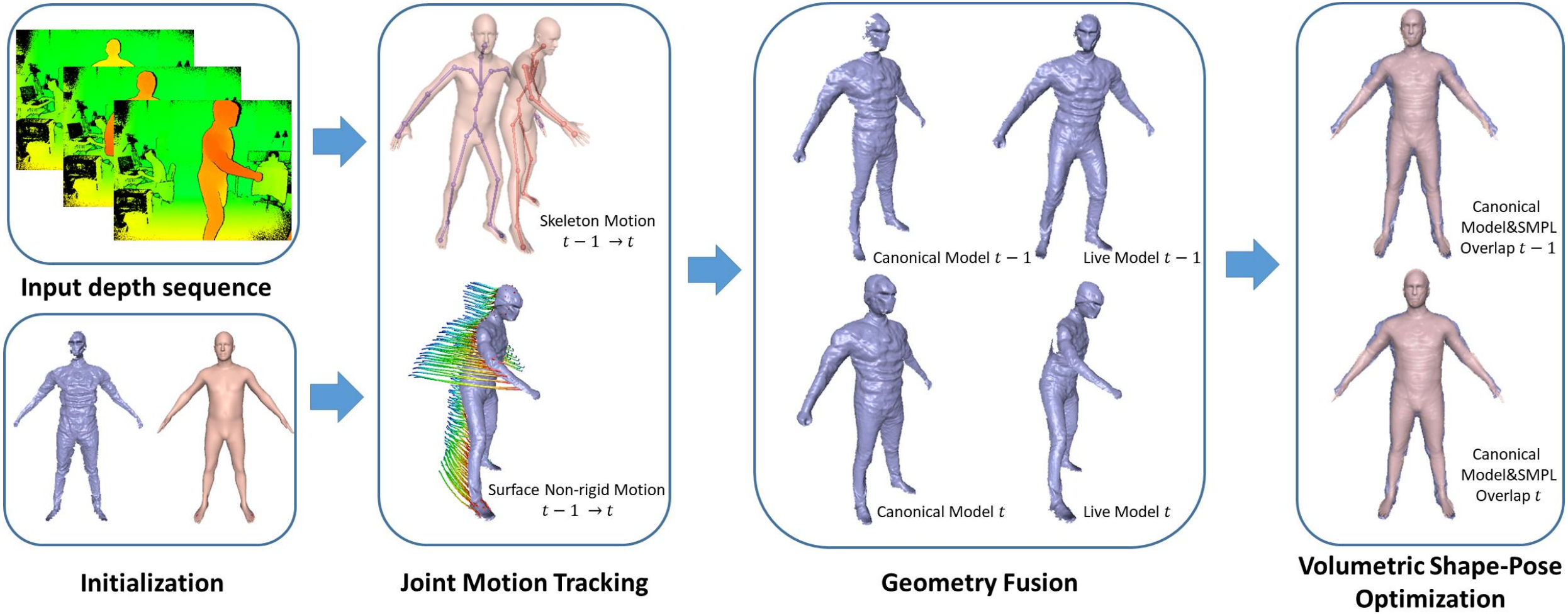
Canonical Fused Geometry

Right View

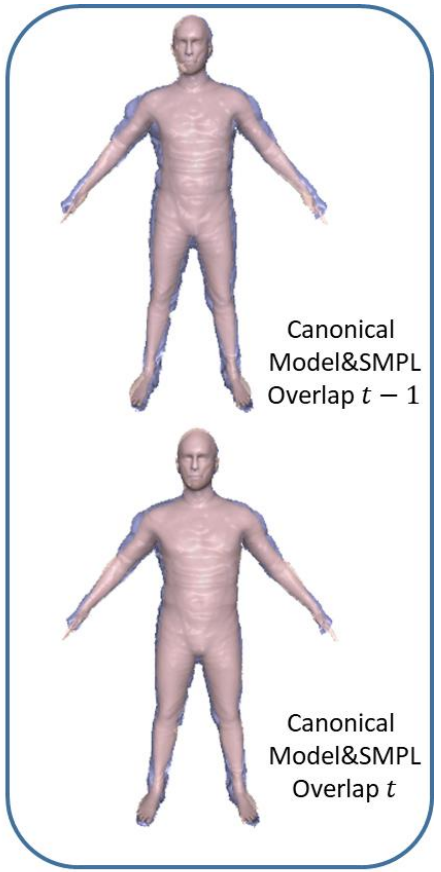
Front View

Left View

Volumetric Shape-Pose Optimization



Volumetric Shape-Pose Optimization

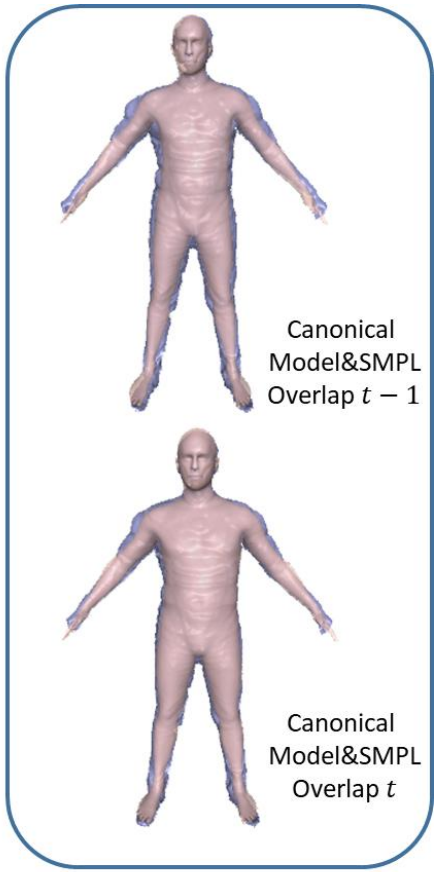


Volumetric Shape-Pose Optimization

$$E_{shape} = \lambda_{data}E_{sdata} + \lambda_{sreg}E_{sreg} + \lambda_{pri}E_{pri}$$

Inner Body Model	Solved Shape & Pose	Solved Pose
fit	*constraint*	*constraint*
TSDF	Temporal Smooth	Natural Poses

Volumetric Shape-Pose Optimization

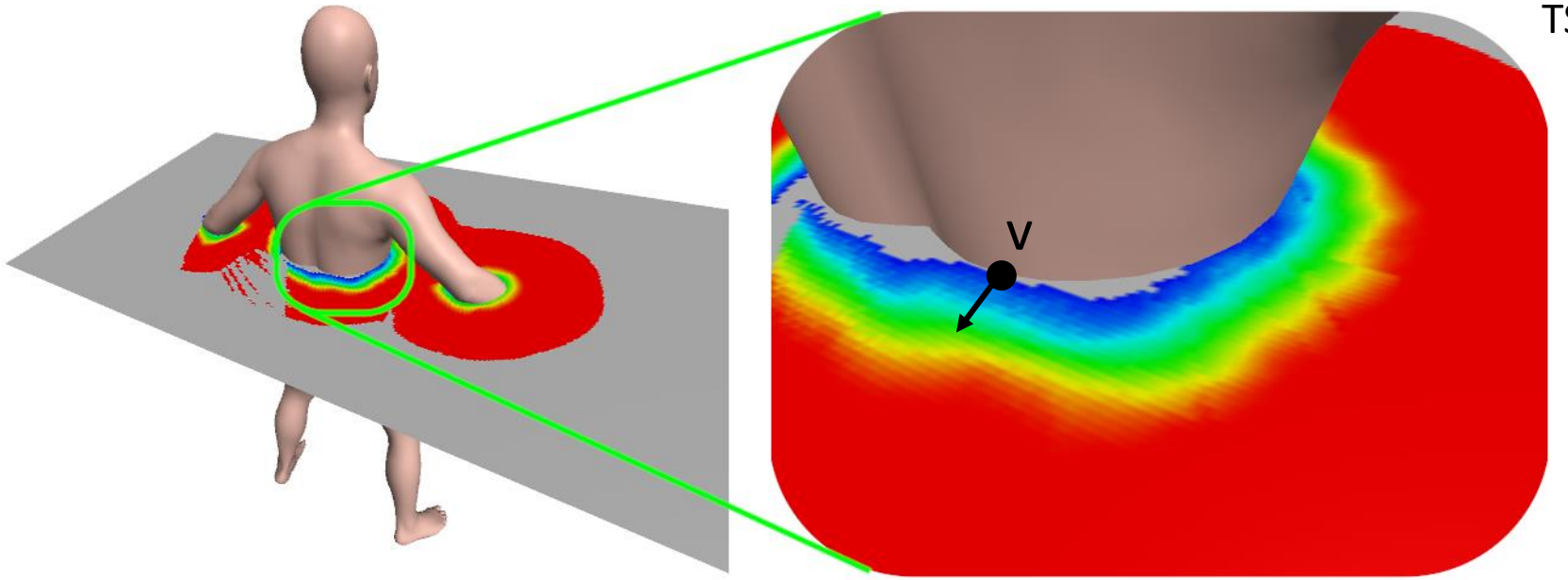


Volumetric Shape-Pose Optimization

$$E_{sdata}(\beta,\theta)=\sum_{\bar{v}\in\bar{T}}\psi\big(\mathrm{D}(\mathrm{v})\big)$$

$\mathrm{D}(\mathrm{v})$: sampled TSDF value of v

$\mathrm{v}= W(T(\bar{\mathrm{v}};\beta,\theta);J(\beta),\theta)$



Normalized TSDF Value

1.0

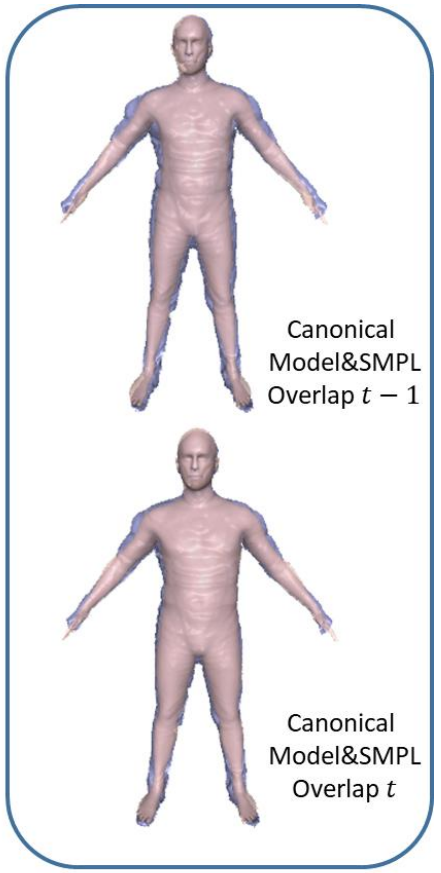
0.0 Surface

-1.0

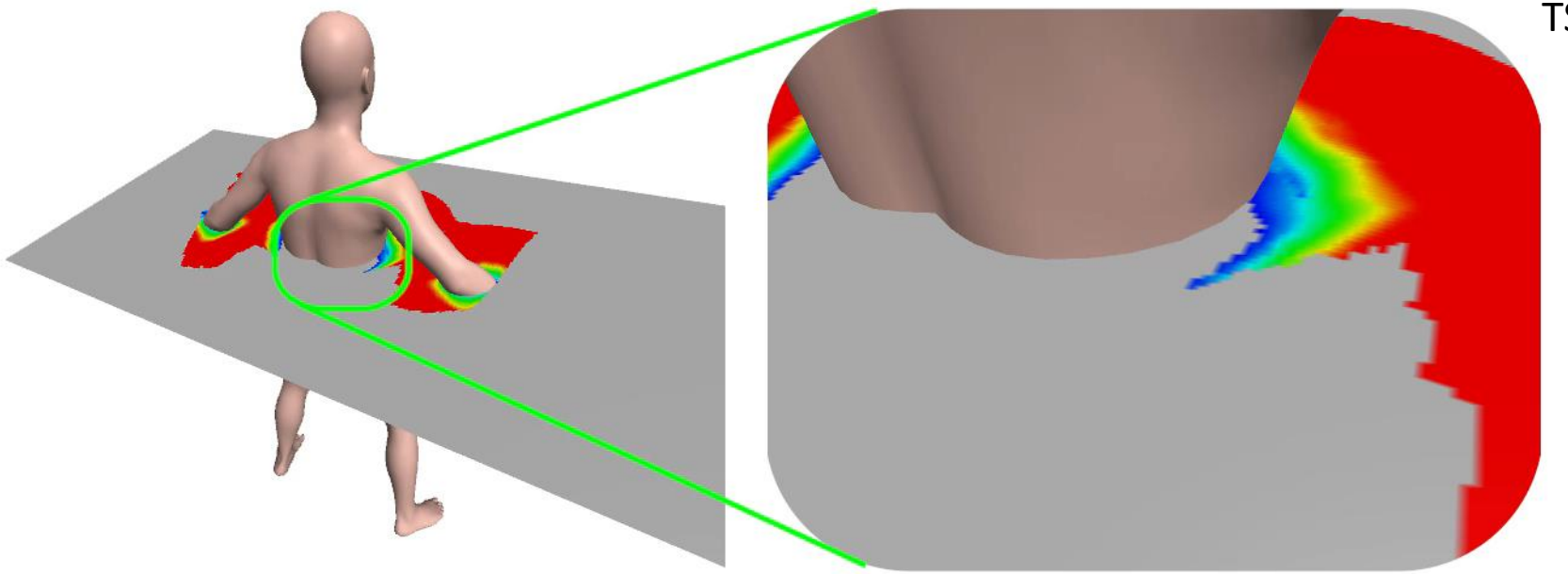
Unobserved Area

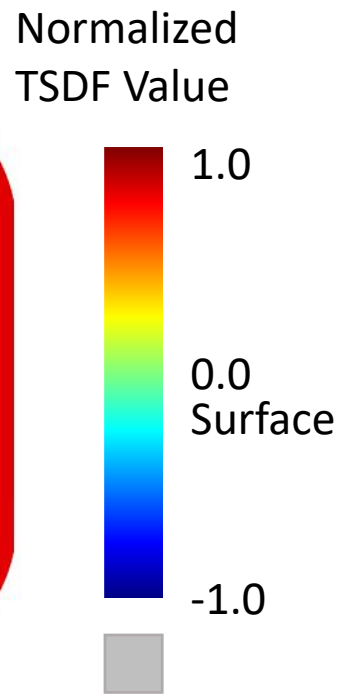
Volumetric Shape-Pose Optimization

$$E_{shape} = E_{sdata} + E_{sreg} + E_{pri}$$

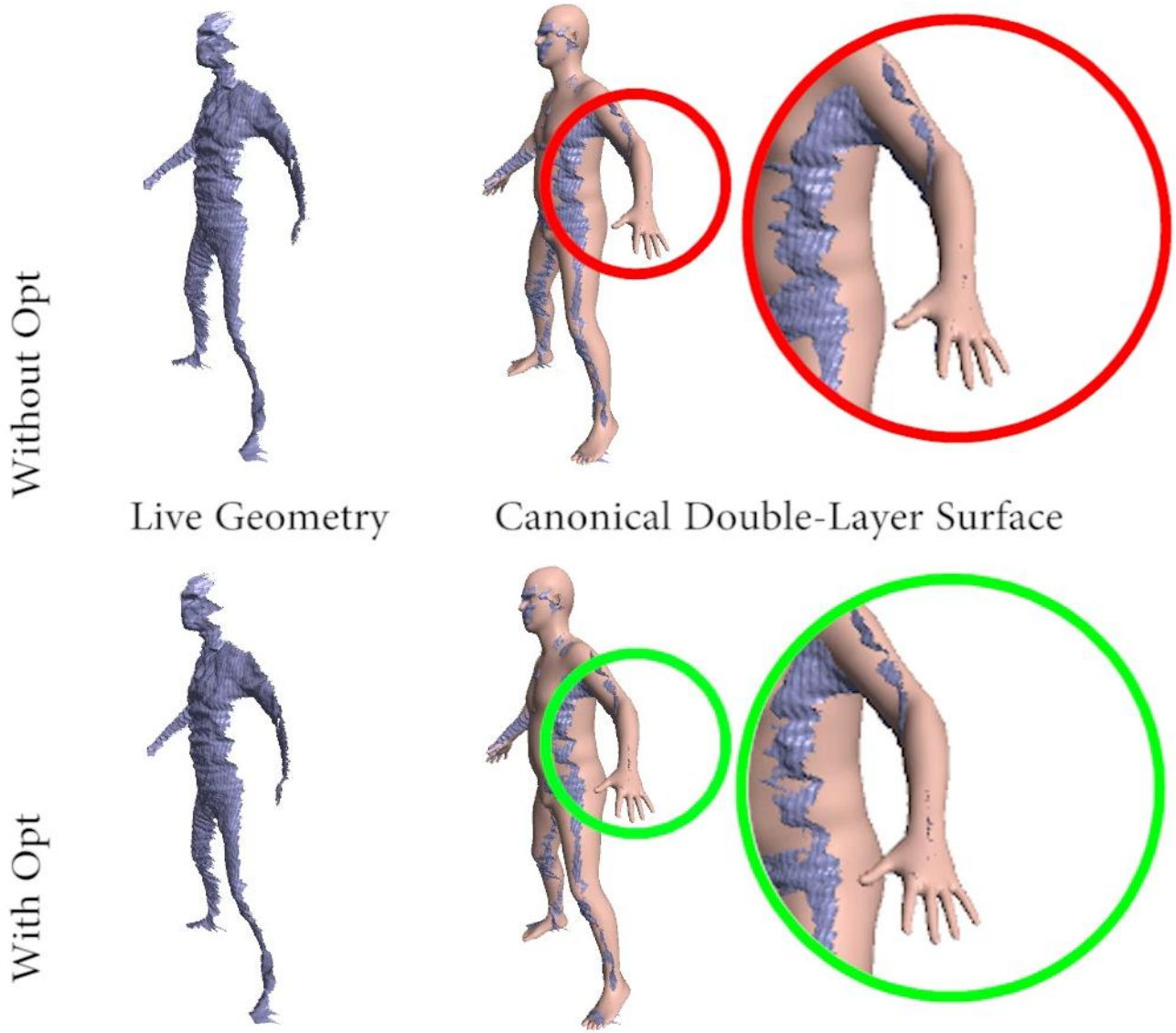


Volumetric Shape-Pose Optimization





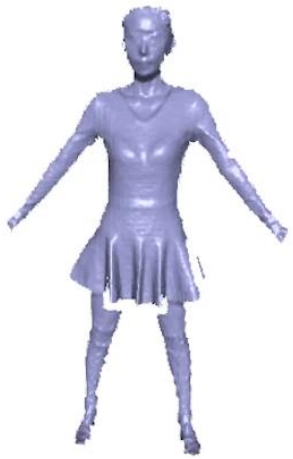
Volumetric Shape-Pose Optimization: Evaluation

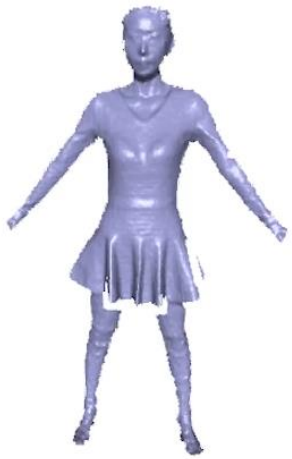


Tracking Failure

Deformation Artifacts

Double Node Graph: Evaluation





green: on body node graph
blue: far body node graph
red: free form sampled graph

free sampled node graph

on body node graph only

double node graph

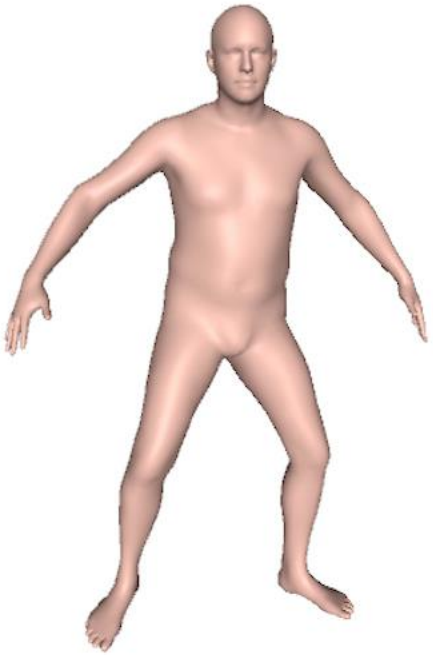
Comparison

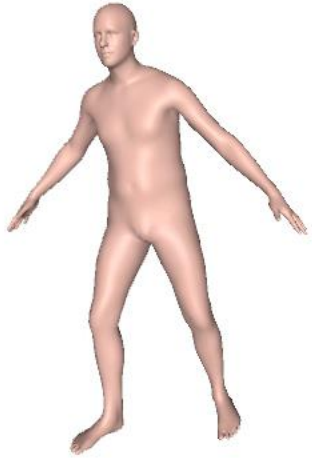
Comparison: Moonwalk

Results

* reference image, not used.
(depth image with mapped color)

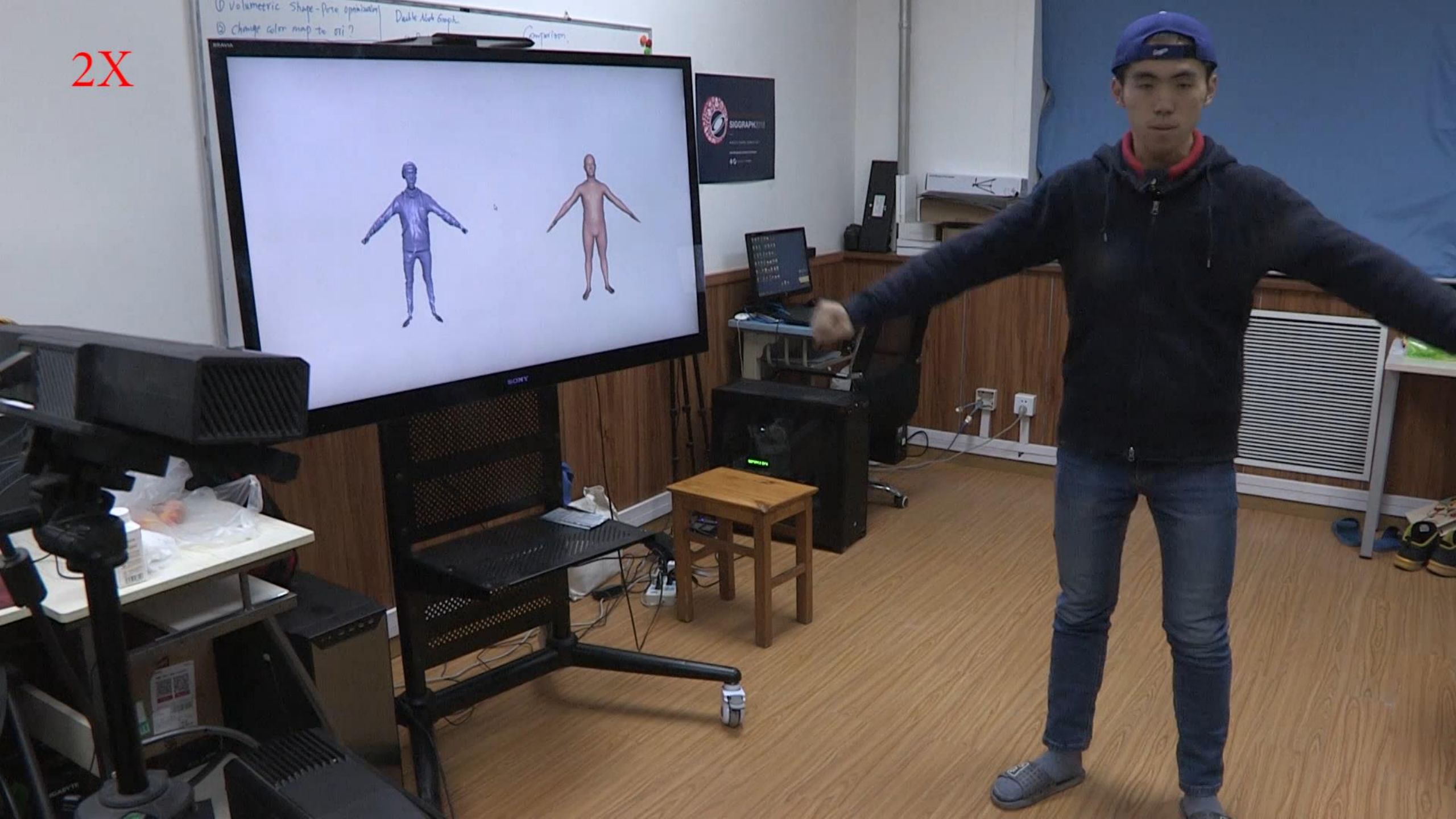
* input depth image





Iron Man

Kicking


2X

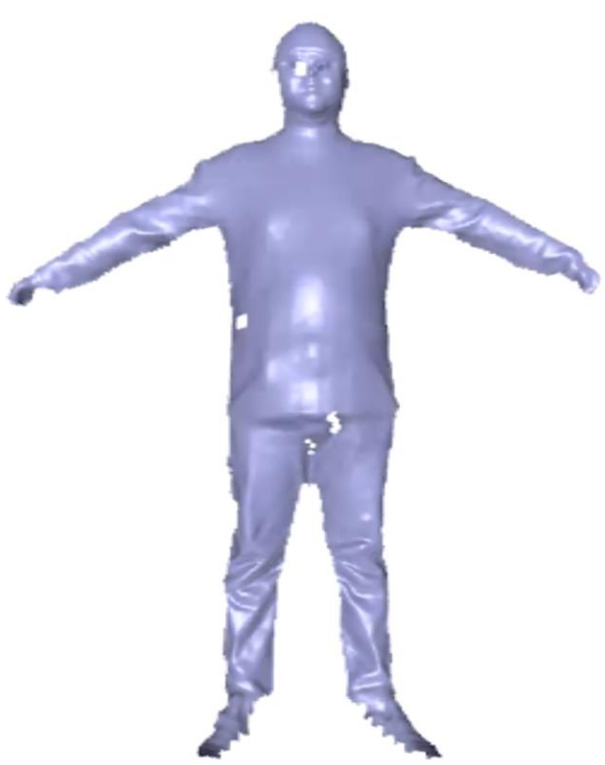
Conclusion

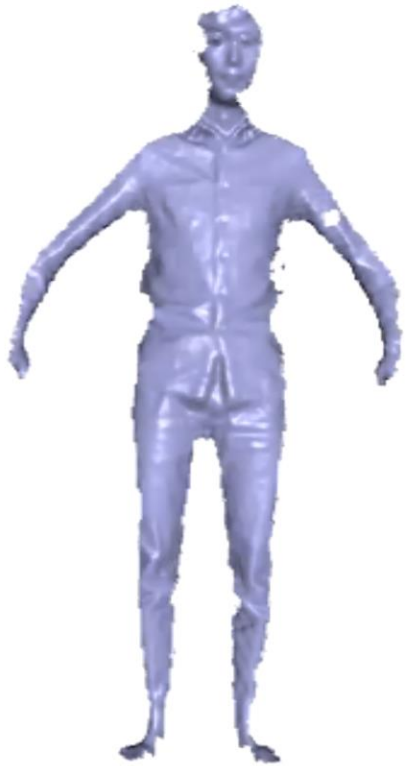
- Convenient & robust single-view performance capture in real-time
- First method that can reconstruct outer surface and inner body shape simultaneously
- Double-Layer surface representation is the key of robust tracking & plausible loop closure
- Applications include: AR/VR, gaming and virtual try-on etc

Limitations

- The reconstruction of very wide cloth.
- Geometry Separations / Topological Changes.
- Human-Object Interactions.






Thank You