ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Universität Bielefeld

Robust Articulated-ICP for Real-Time Hand Tracking

Andrea Tagliasacchi*
Sofien Bouaziz

Matthias Schröder*
Mario Botsch

Anastasia Tkach
Mark Pauly

* equal contribution

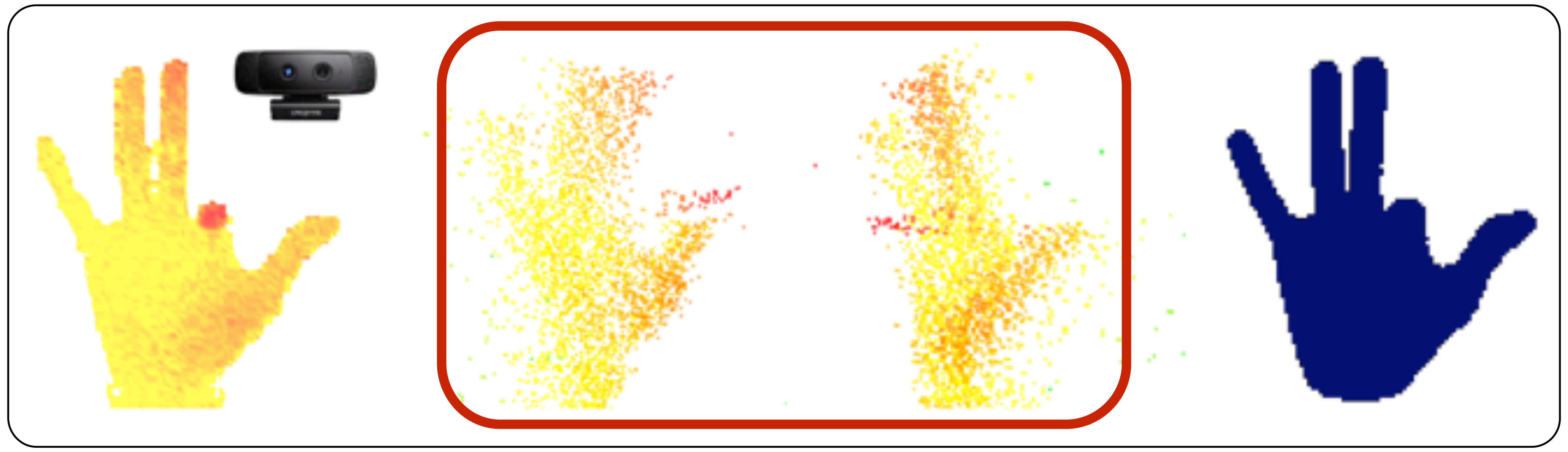
Real-Time Tracking Setup

Data from (single) RGBD Sensors

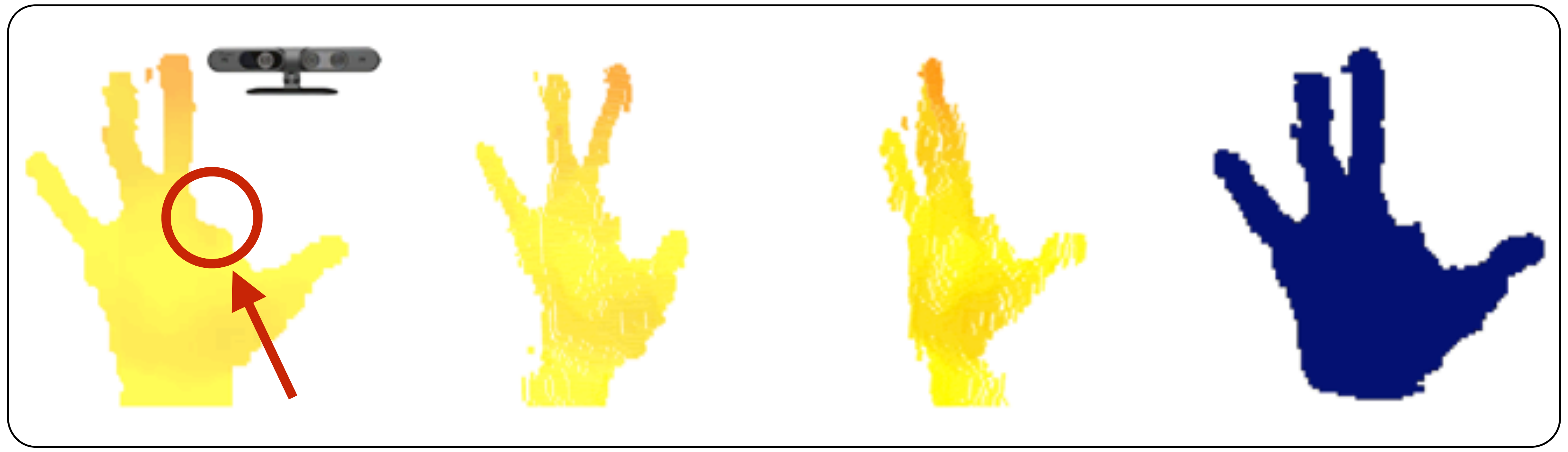
PrimeSense (Carmine)

SoftKinetic

Intel RealSense



low SNR along in depth (z-axis)



completely discards small portions of geometry

Motivation? Augmented Reality

Microsoft HoloLens - PR Video (hololens.com)

Intel Perceptual SDK

Oculus Research (VR)

MagicLeap (AR)

HoloLens (AR)

Previous Work

Appearance-based
(guess solely based on current frame)

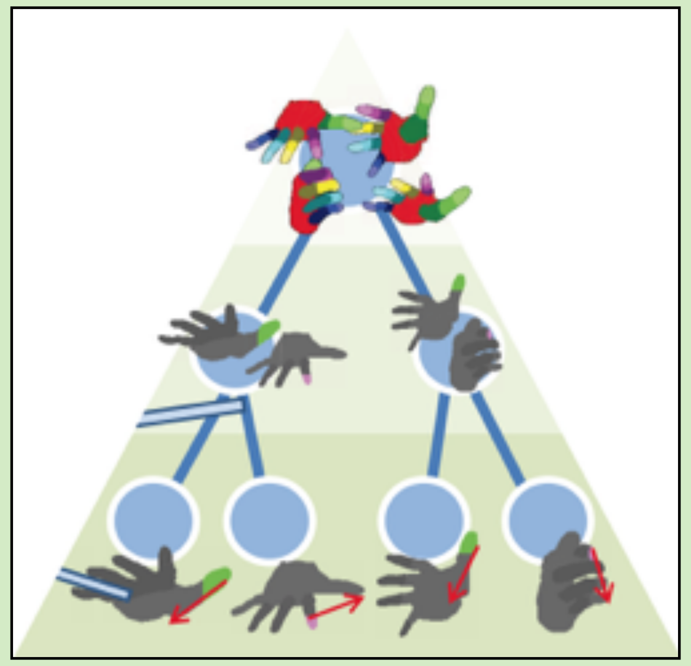
Model-based
(registers model of previous frame)

Previous Work

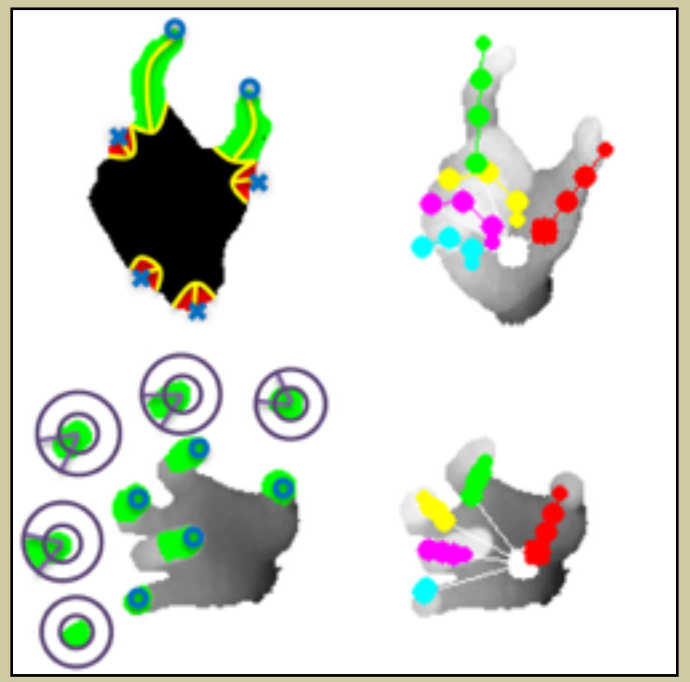
Appearance "vision"

[Tompson SIG'14]

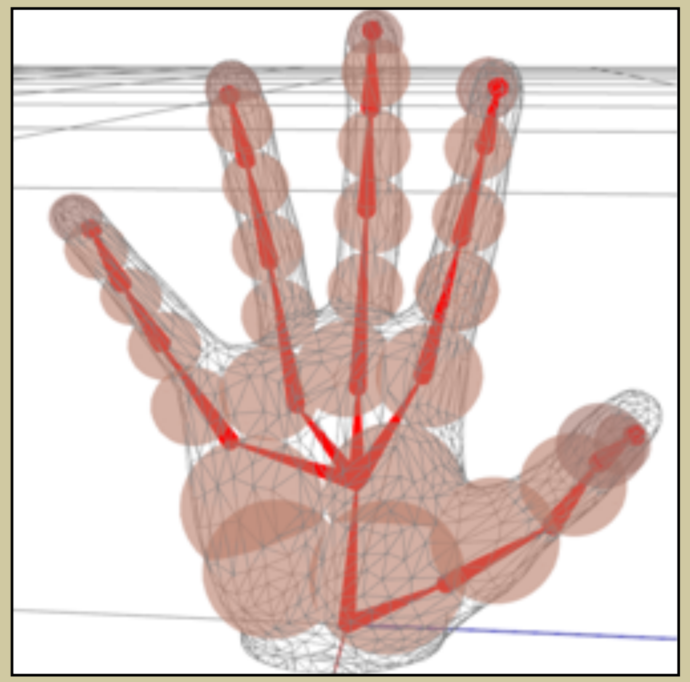
[Keskin ICCV'12]



[Tang CVPR'14]

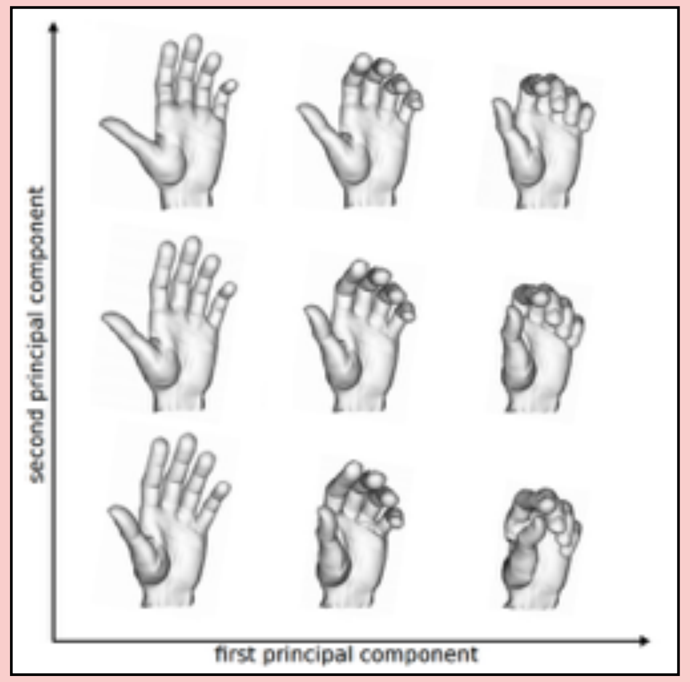


[Qian CVPR'14]



[Sridhar 3DV'14]

[Oikono. CVPR'14]



[Schroder ICRA'14]

Model "geometry"

[Melax I3D'13]

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

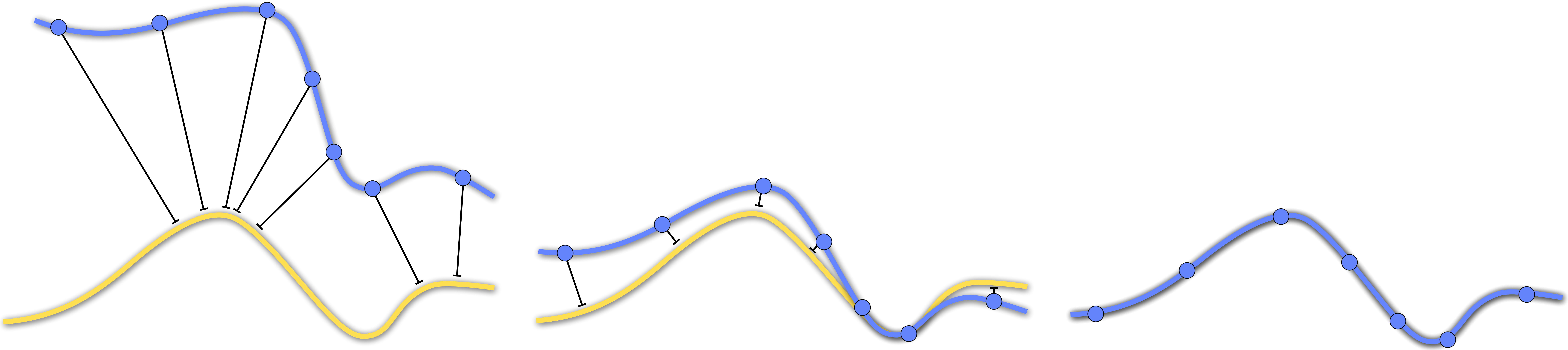
Universität Bielefeld

Contributions

- combined 2D/3D registration (within ICP)
- occlusion-aware correspondences (ICP)
- regularization with statistical pose-space prior
- extensible and unified real-time solver (>60fps)
- **revamping ICP** for articulated tracking

ICP: Iterative Closest Point

- **Step 1:** optimizing correspondences
- **Step 2:** optimizing transformations



update correspondences

update transformation
update correspondences

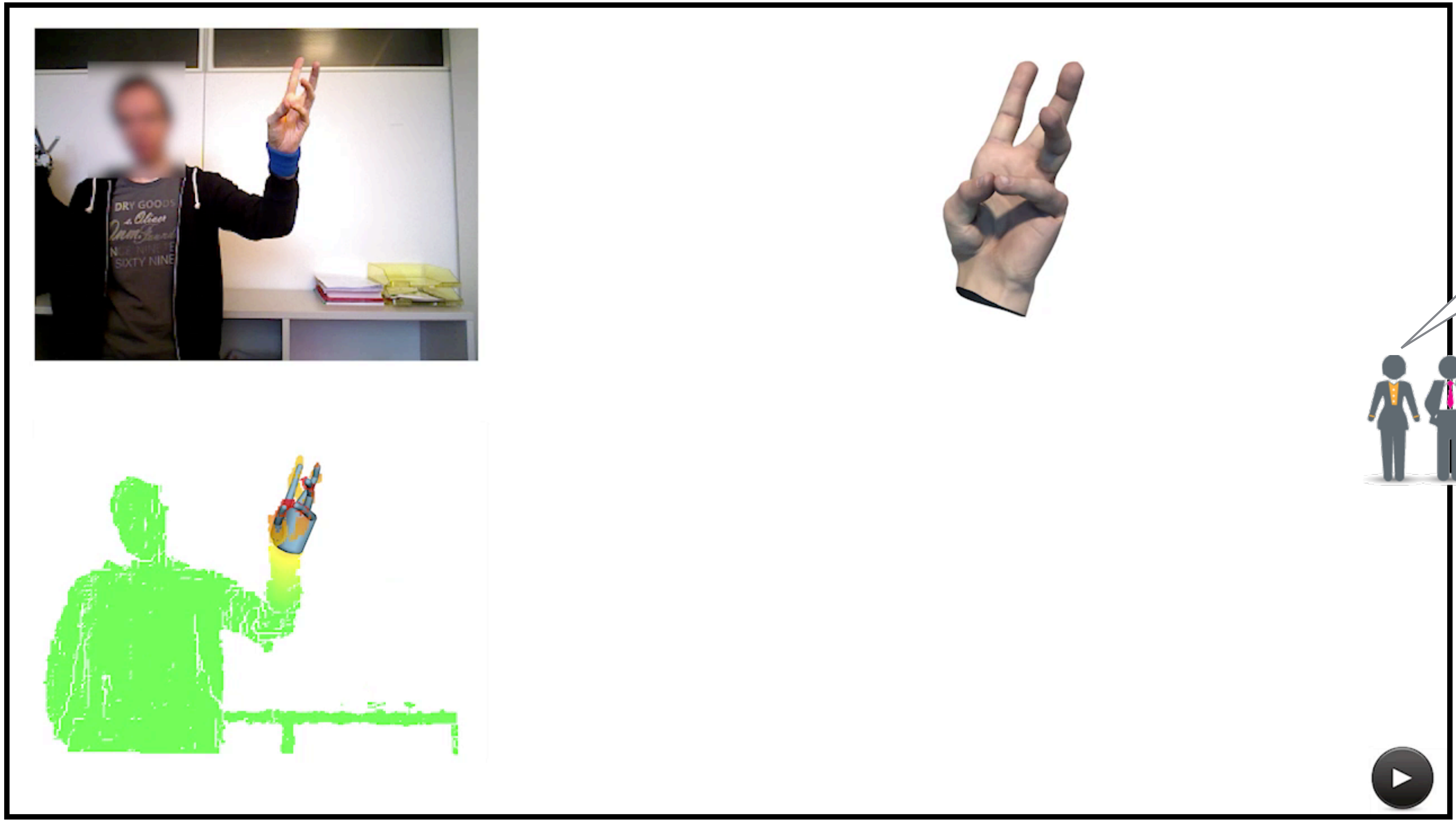
update transformation

for more details please refer to Sparse-ICP [Bouaziz, Tagliasacchi, Pauly SGP'13]

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Robust Articulated ICP

Universität Bielefeld

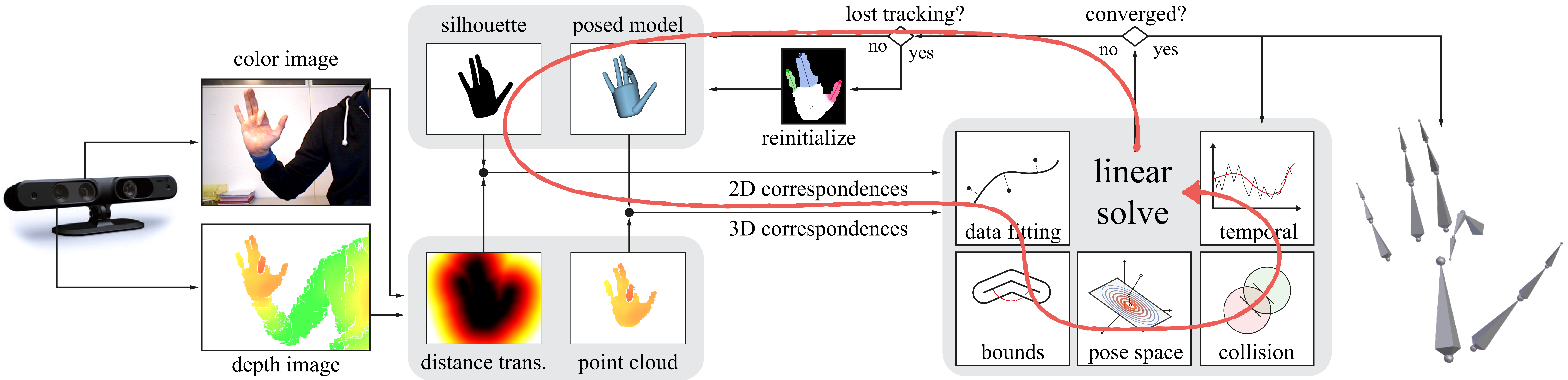


[Wei et al. SIGA'12] our tracking process successfully tracks the entire motion sequence while ICP fails to track most of frames. This is because **ICP is often sensitive to initial poses and prone to local minimum**, particularly involving tracking high-dimensional human body poses from noisy depth data.

[Zhang et al. SIGA'14] The accompanying video clearly shows that our tracking process is **much more robust than the ICP algorithm** […] our tracking process successfully tracks the entire motion sequence while **ICP fails to track most of frames.**

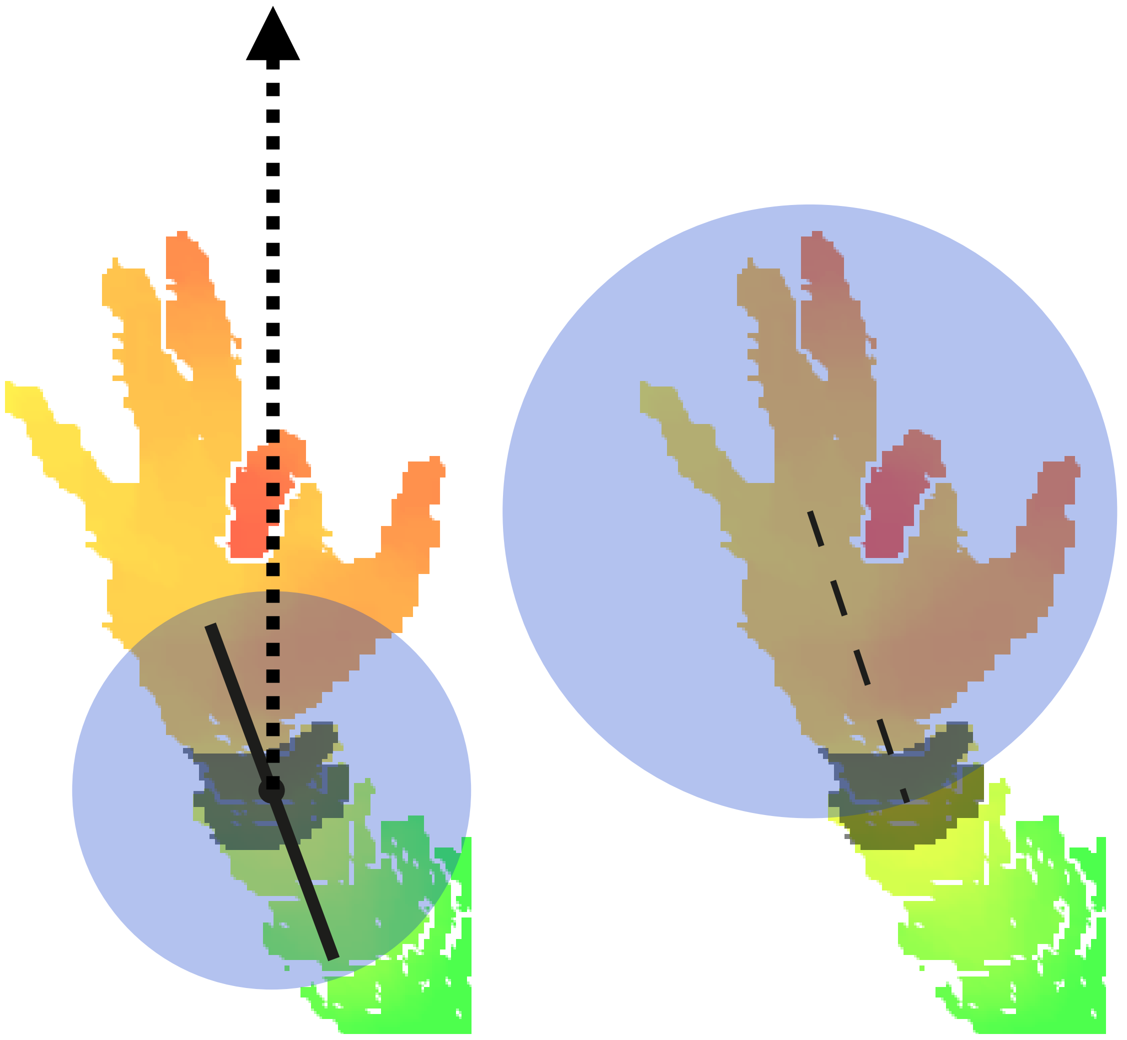
[Qian et al. CVPR'14] It uses alternate and gradient based optimization, converges fast, and is suitable for realtime applications. However, it can be **easily trapped in poor local optima** and cannot handle non-rigid objects well. Yet, it is still **insufficient for high-dimensional articulated** hands, especially under free viewpoints.

System Overview

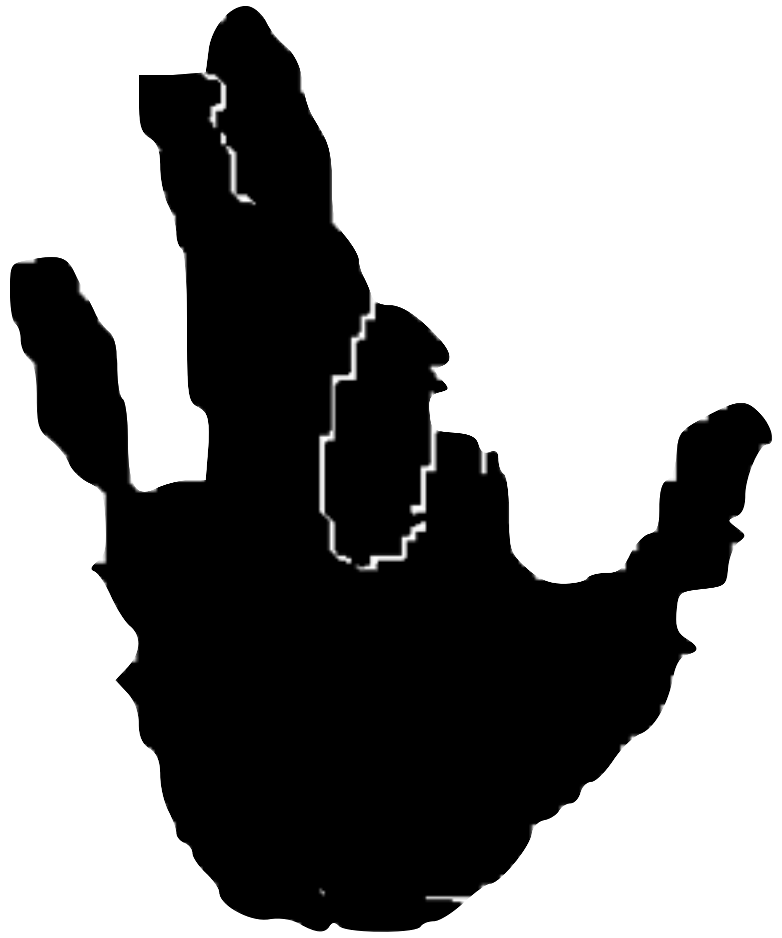


$$\min_{\boldsymbol{\theta}} \underbrace{E_{\text{points}} + E_{\text{silh.}} + E_{\text{wrist}}}_{\text{Fitting terms}} + \underbrace{E_{\text{pose}} + E_{\text{kin.}} + E_{\text{temporal}}}_{\text{Prior terms}}$$

Preprocessing



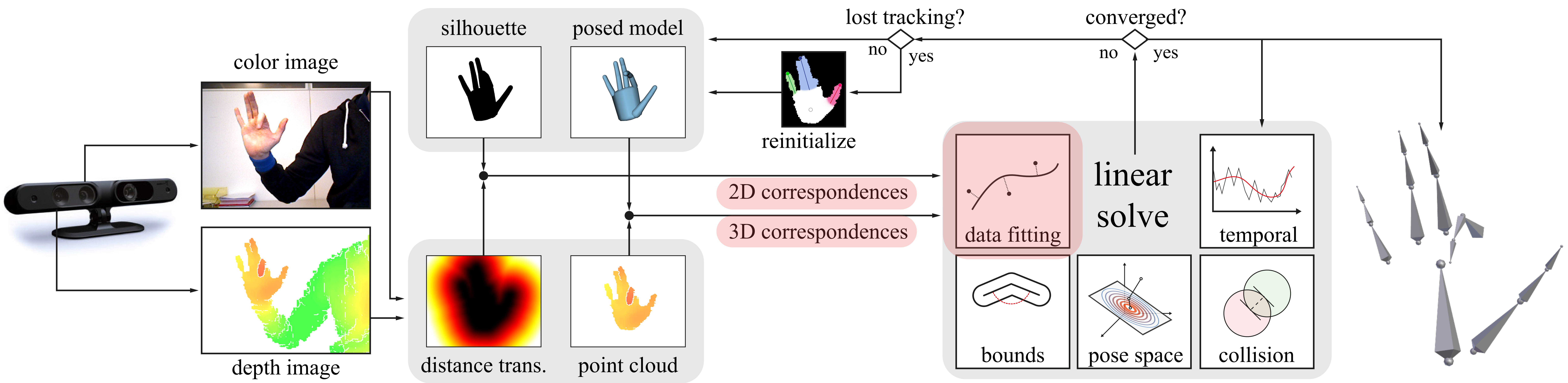
$\mathcal{S}_s$ - sensor silhouette



- color to identify the Region-of-Interest
- **demo**: assumption on picking "+y" for PCA
- … but all this could be learned!!

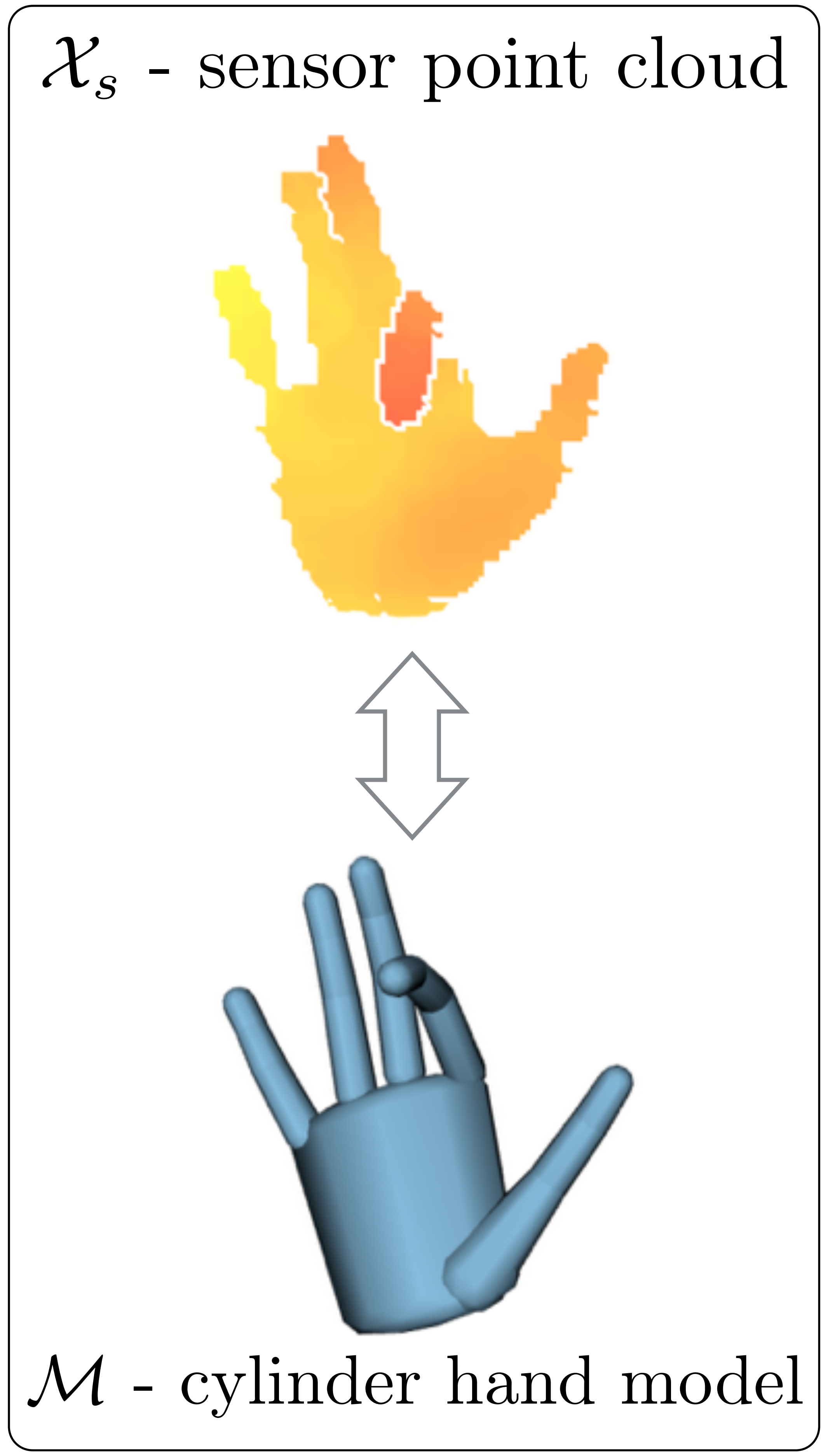
[Tompson et al. TOG'14]

Data Fitting Energies



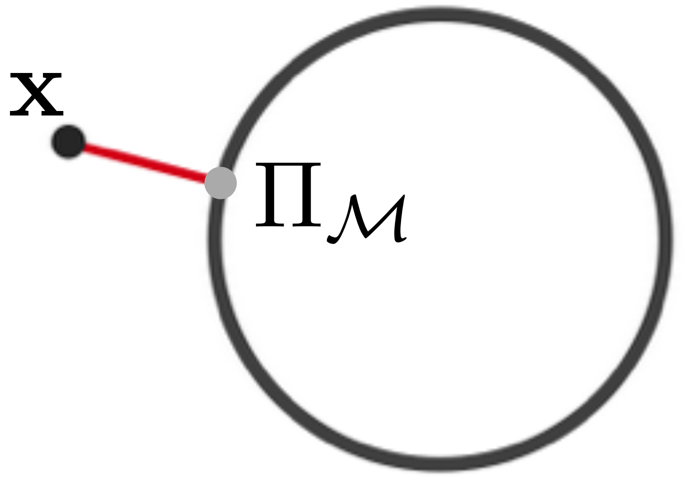
$$\min_{\boldsymbol{\theta}} \underbrace{E_{\text{points}} + E_{\text{silh.}} + E_{\text{wrist}}}_{\text{Fitting terms}} + \underbrace{E_{\text{pose}} + E_{\text{kin.}} + E_{\text{temporal}}}_{\text{Prior terms}}$$

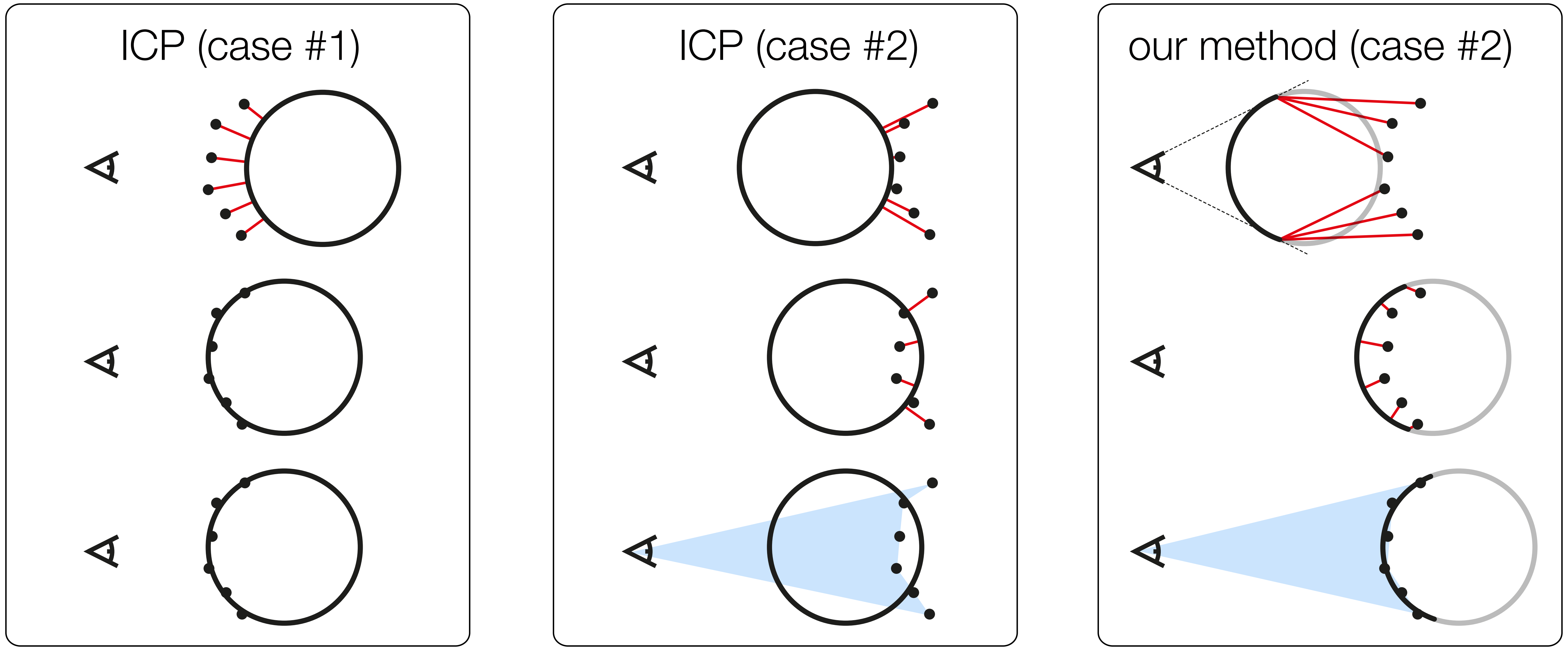
3D Registration (w/ occlusions)



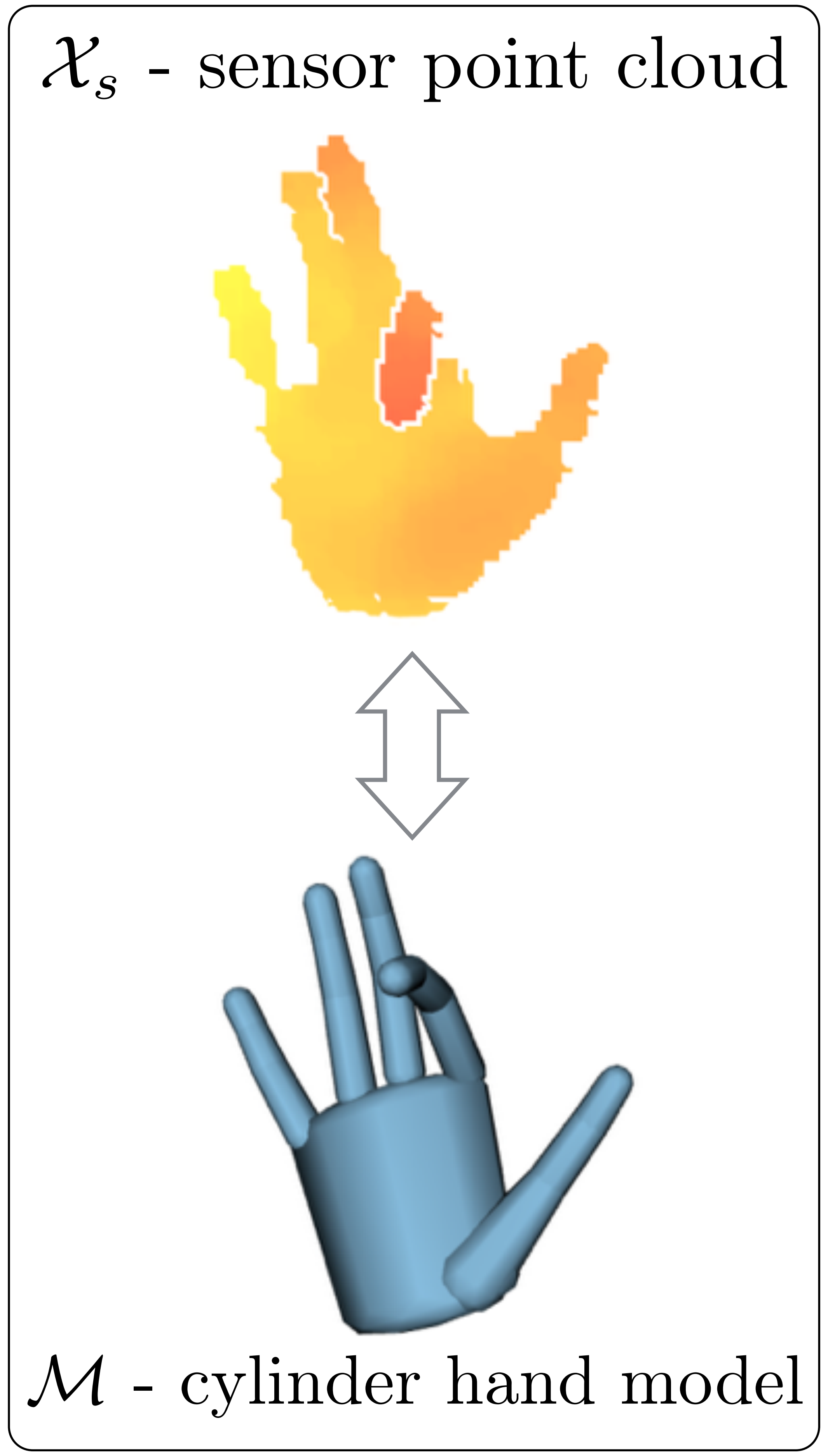
3D Registration

$$E_{\text{points}} = \omega_1 \sum_{\mathbf{x} \in \mathcal{X}_s} \| \mathbf{x} - \Pi_{\mathcal{M}}(\mathbf{x}, \boldsymbol{\theta}) \|_2^1$$

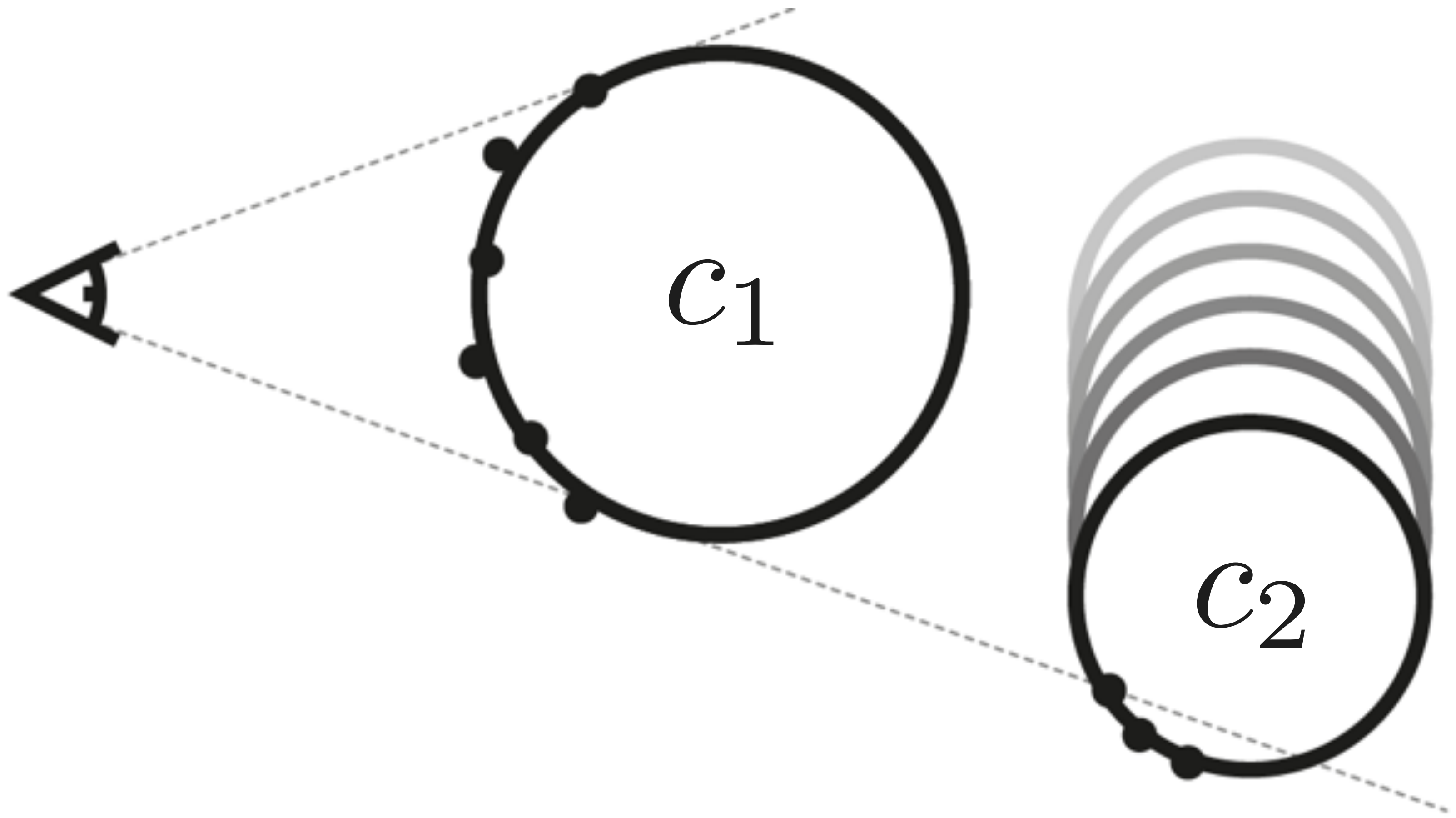




Lesson #1: insufficient to render

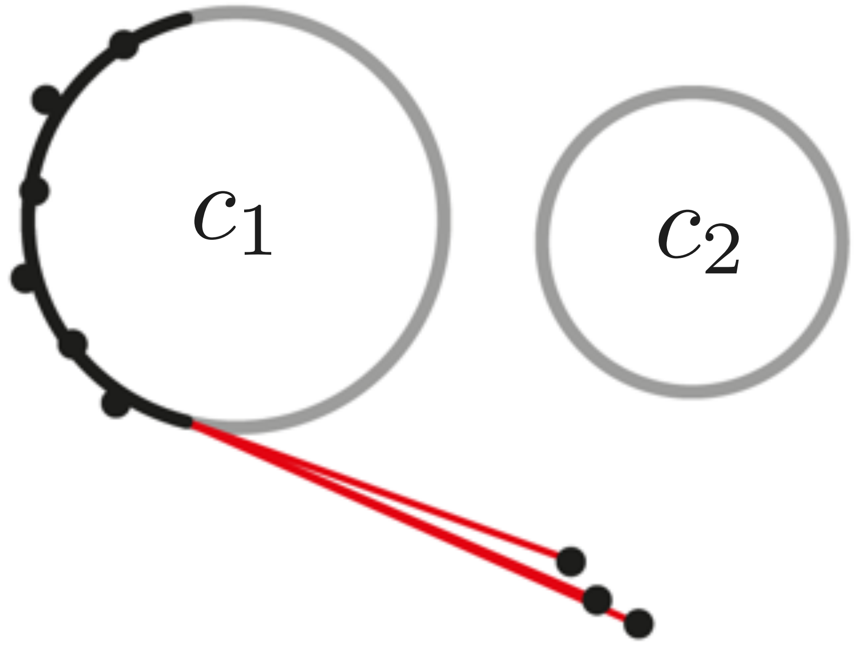


3D Registration

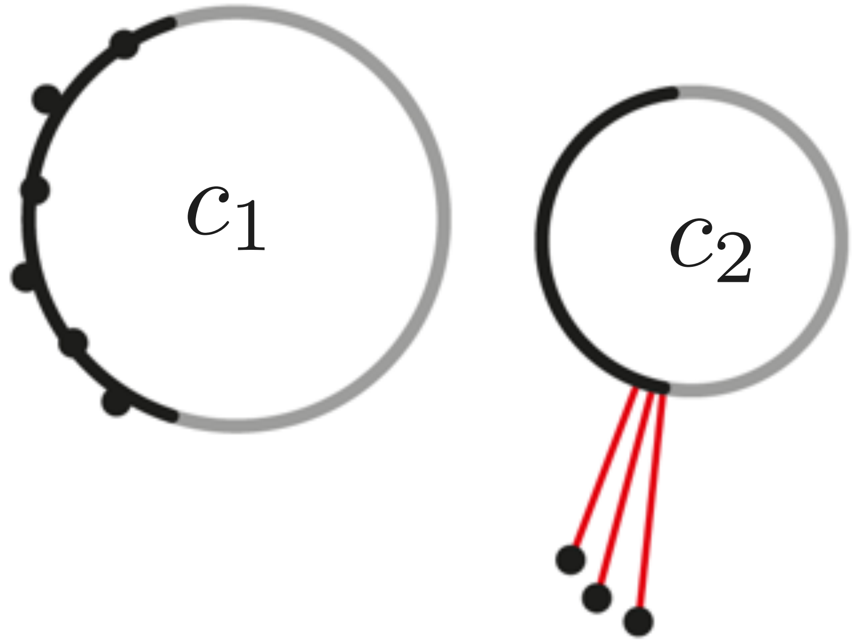


Ground Truth Motion
(finger 2 comes out of occlusion)

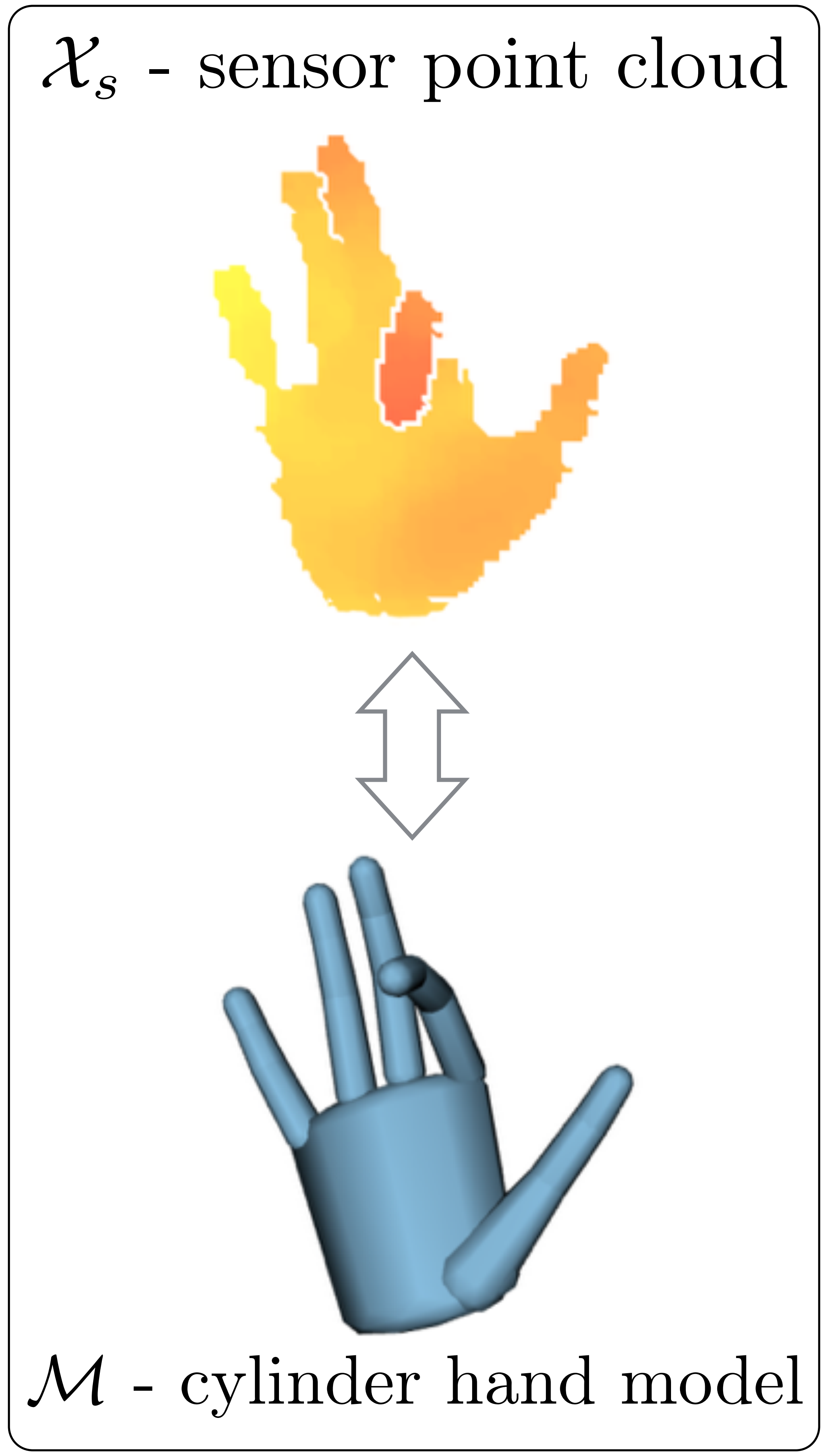
Correspondences of **[Wei et al. SIGA'12]**
(renders the hand model into a point cloud)



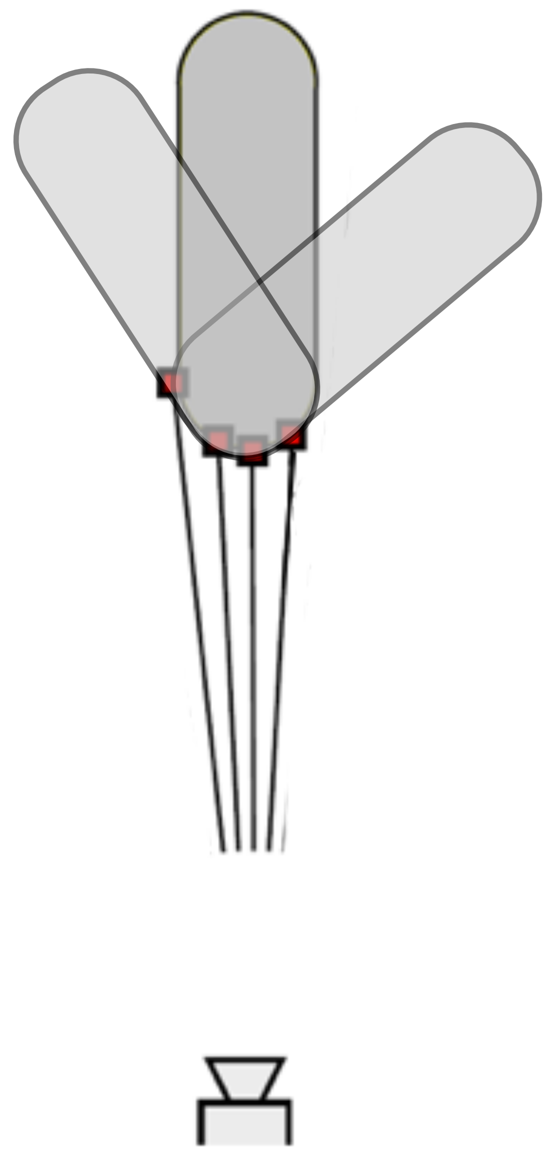
Correspondences of **[Our Method]**
(computes correspondences in close form)



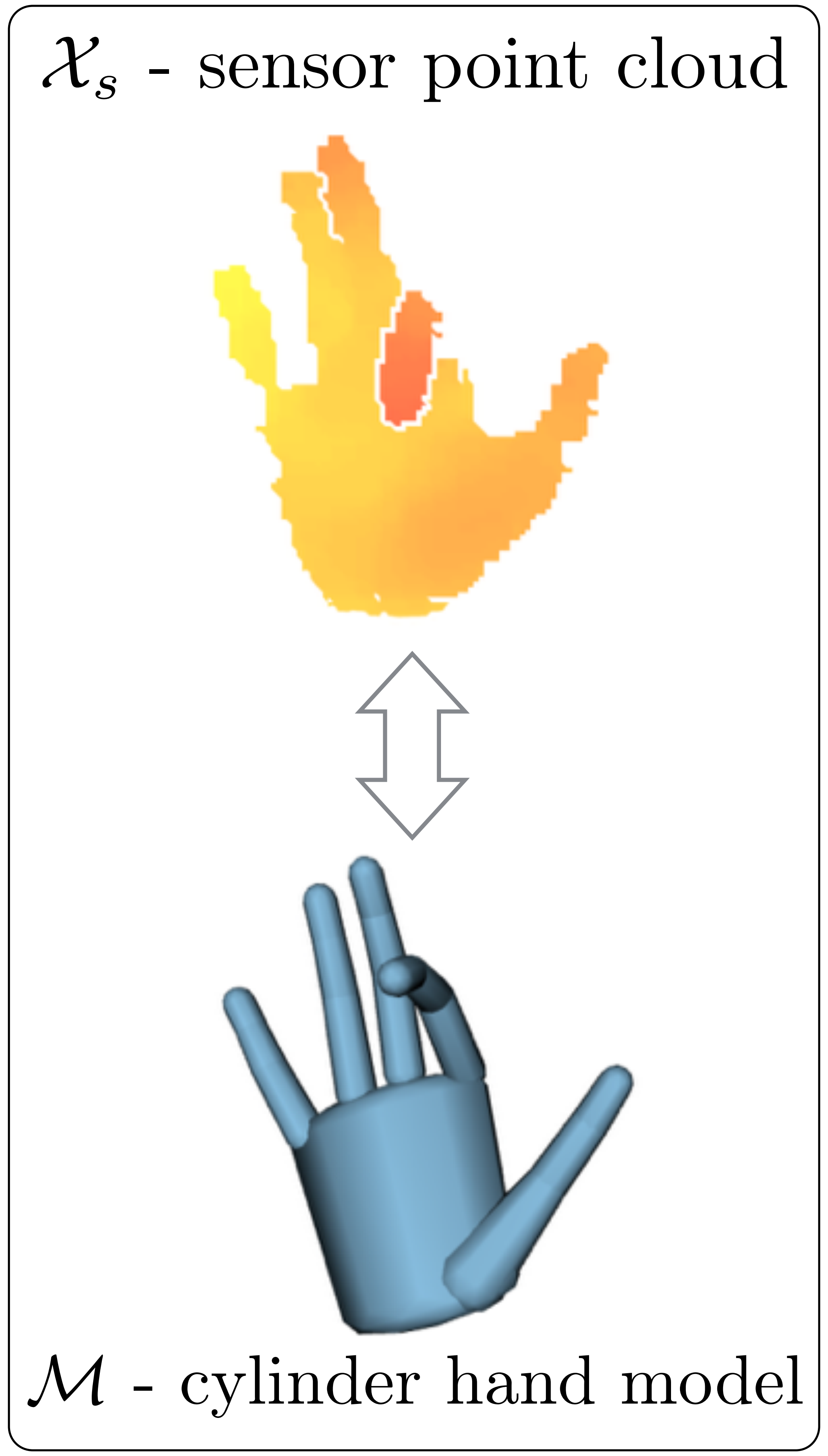
2D/3D Registration



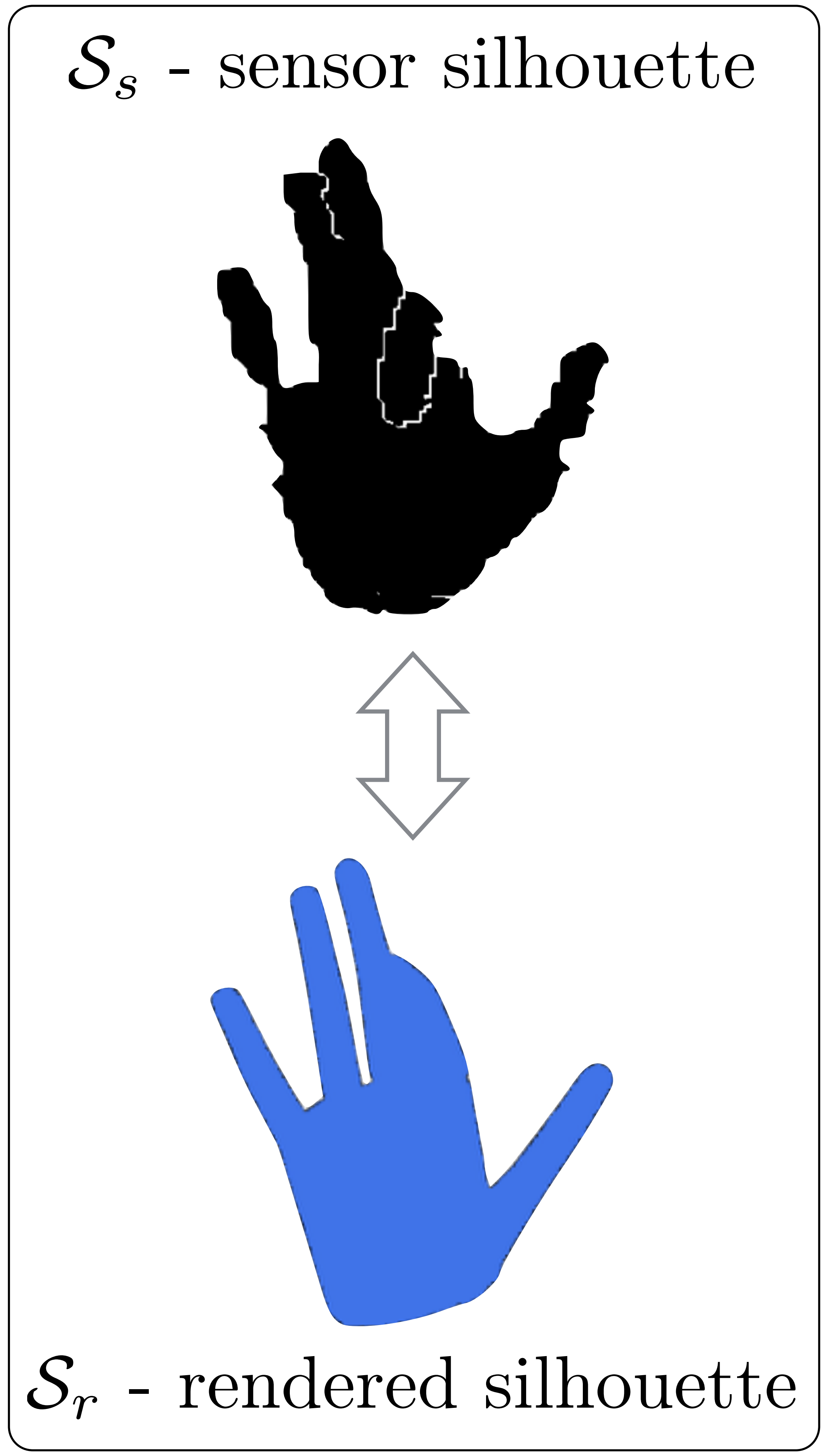
3D Registration



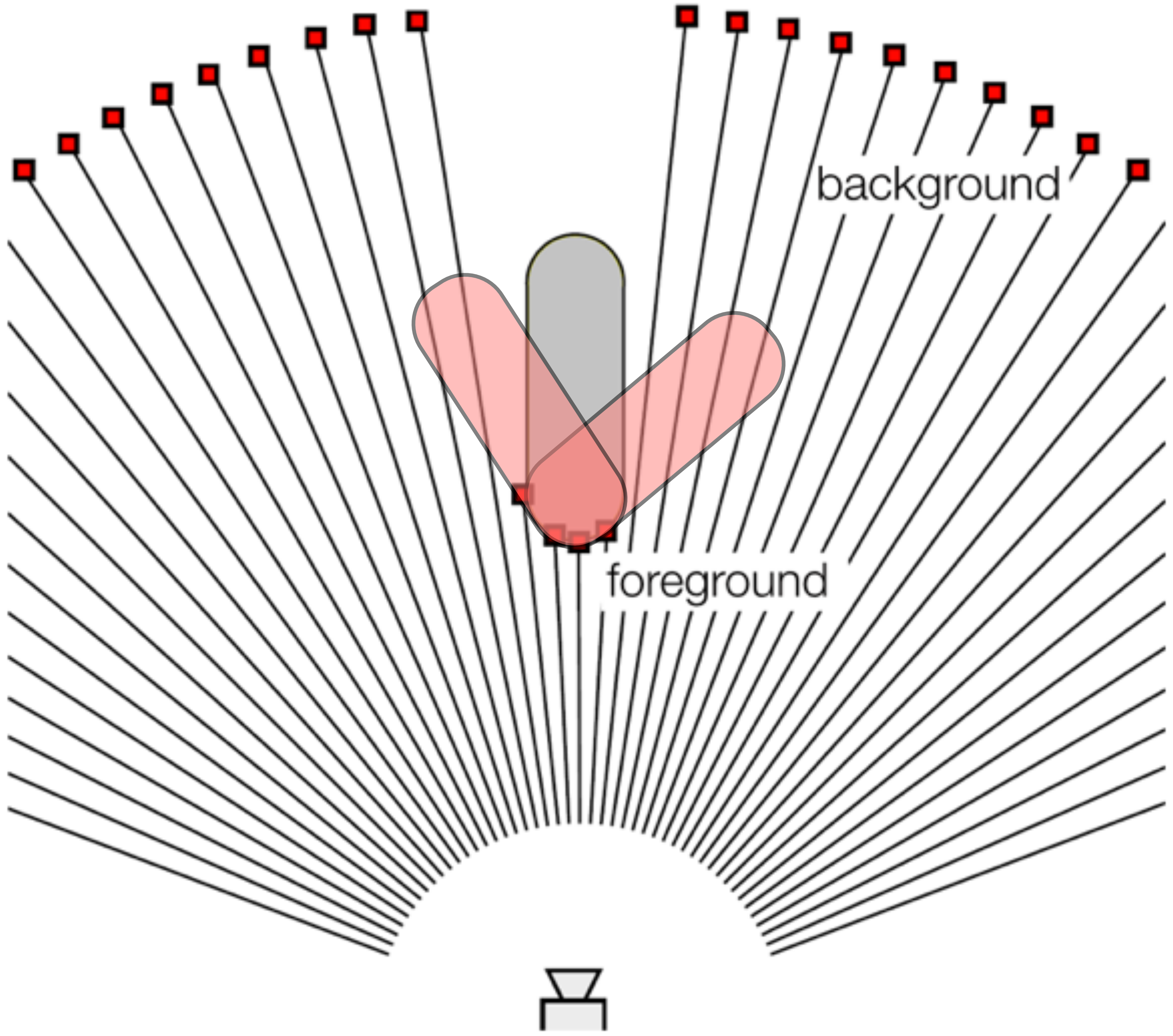
2D/3D Registration



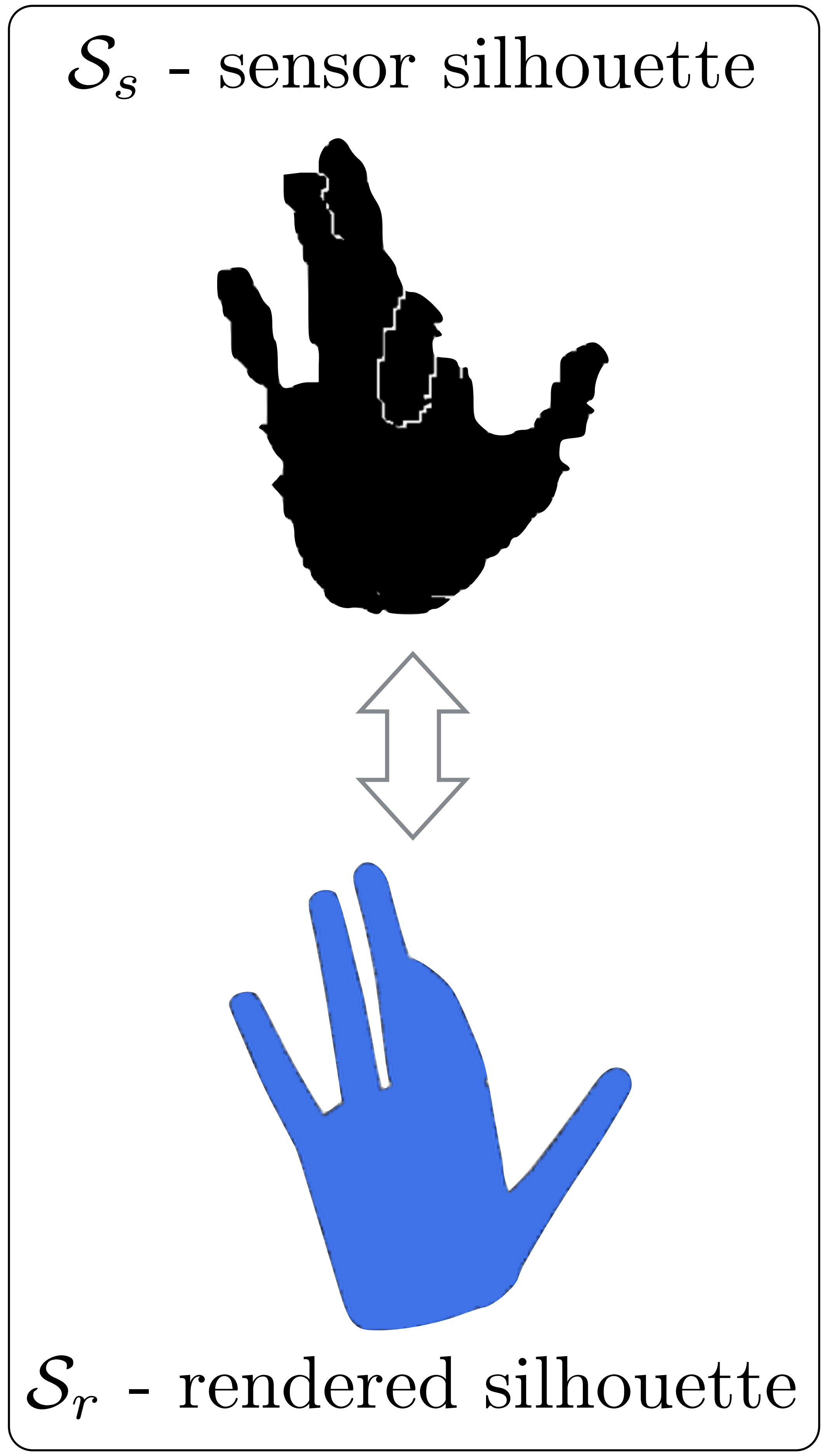
3D Registration



2D Registration

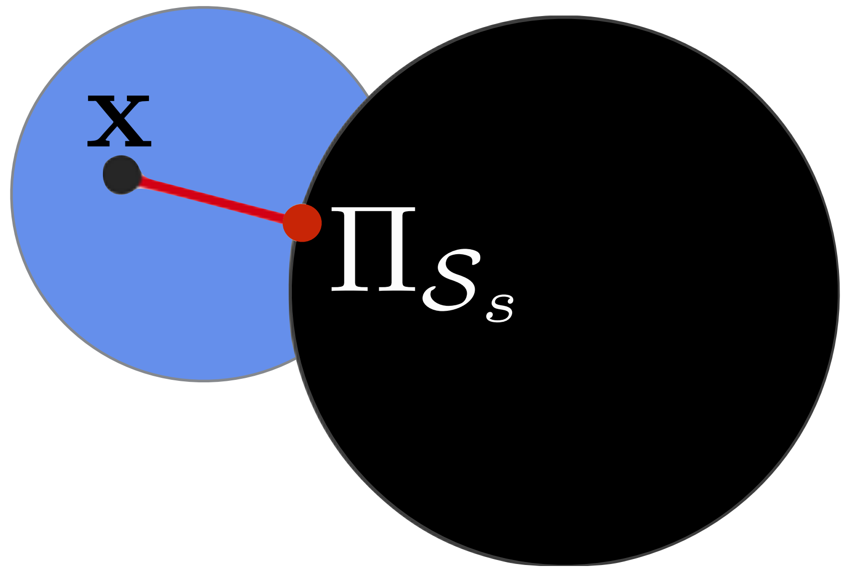


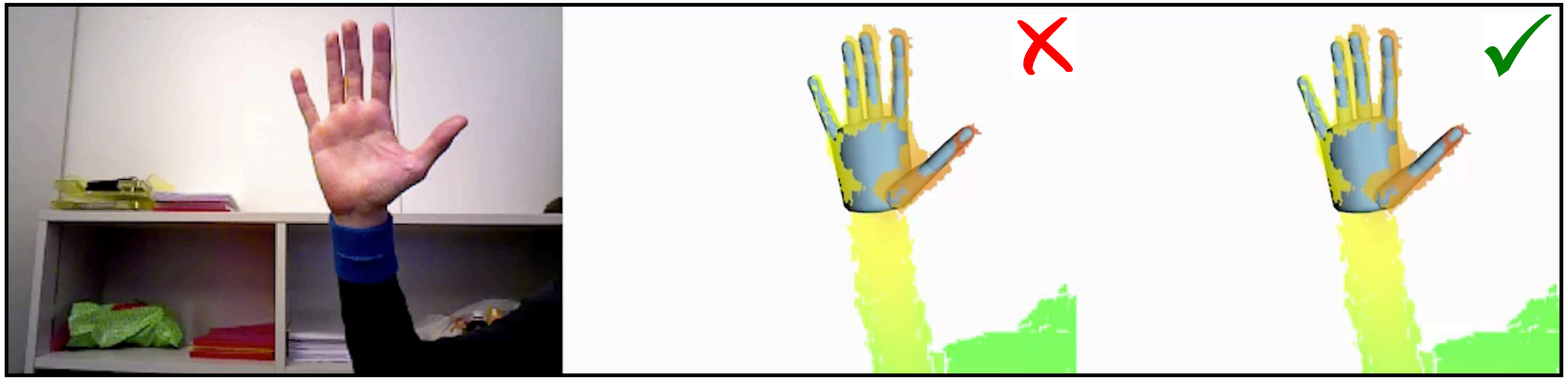
2D Registration



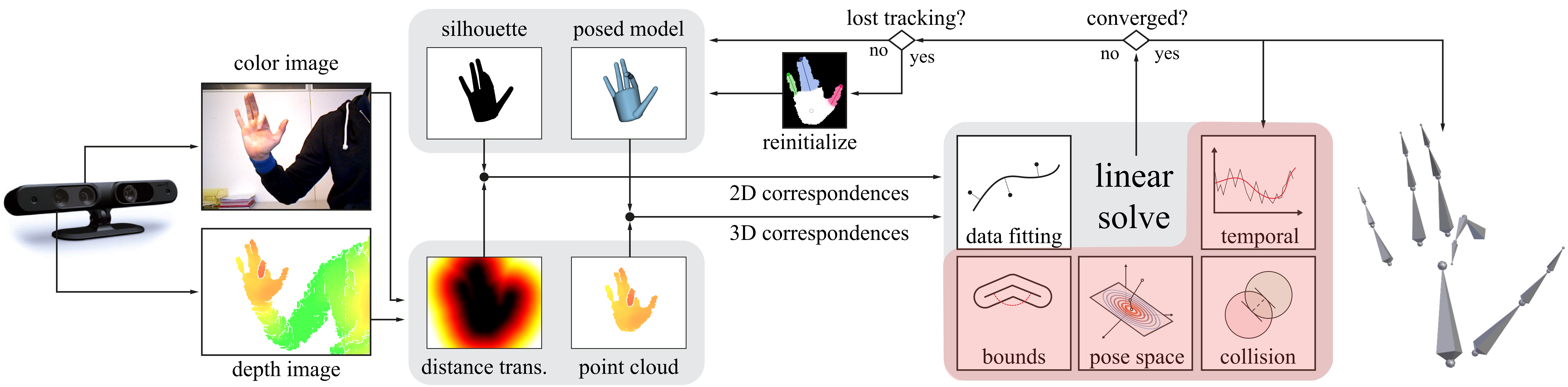
2D Registration

$$E_{\text{silhouette}} = \omega_2 \sum_{\mathbf{p} \in \mathcal{S}_r} \| \mathbf{p} - \Pi_{\mathcal{S}_s}(\mathbf{p}, \boldsymbol{\theta}) \|_2^2$$



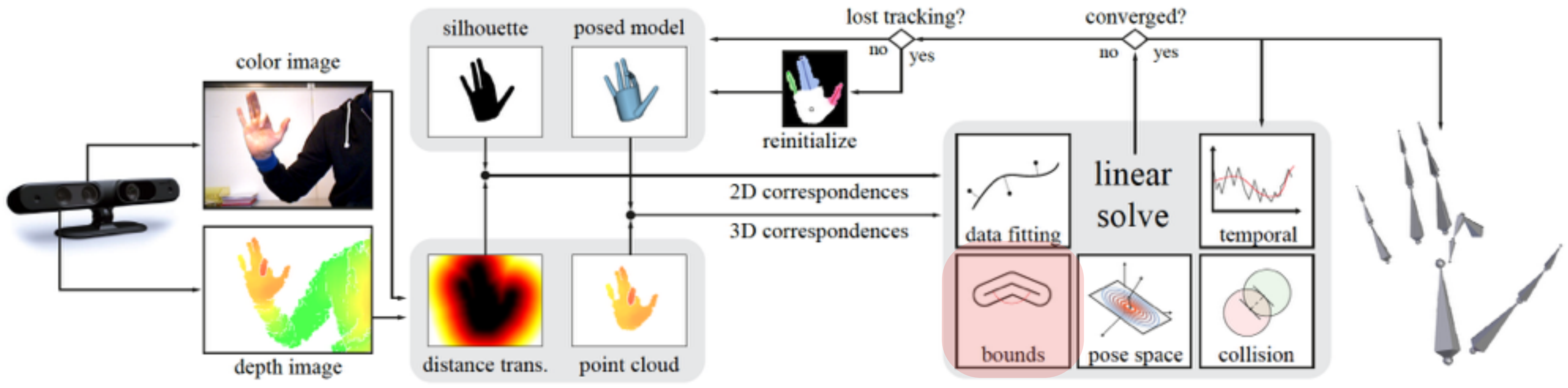


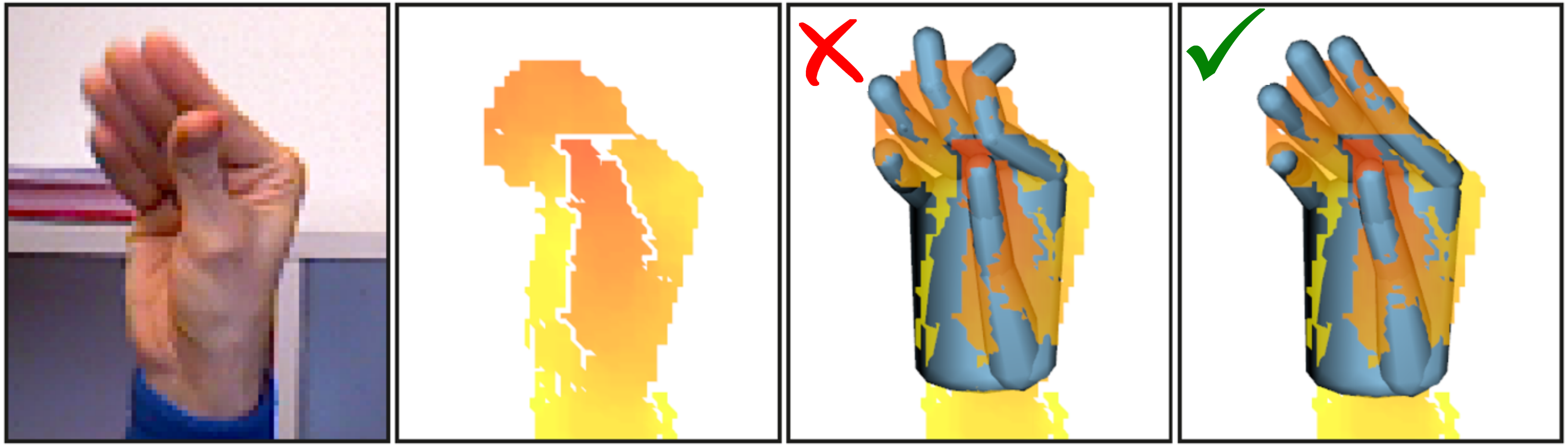
Model Prior Energies



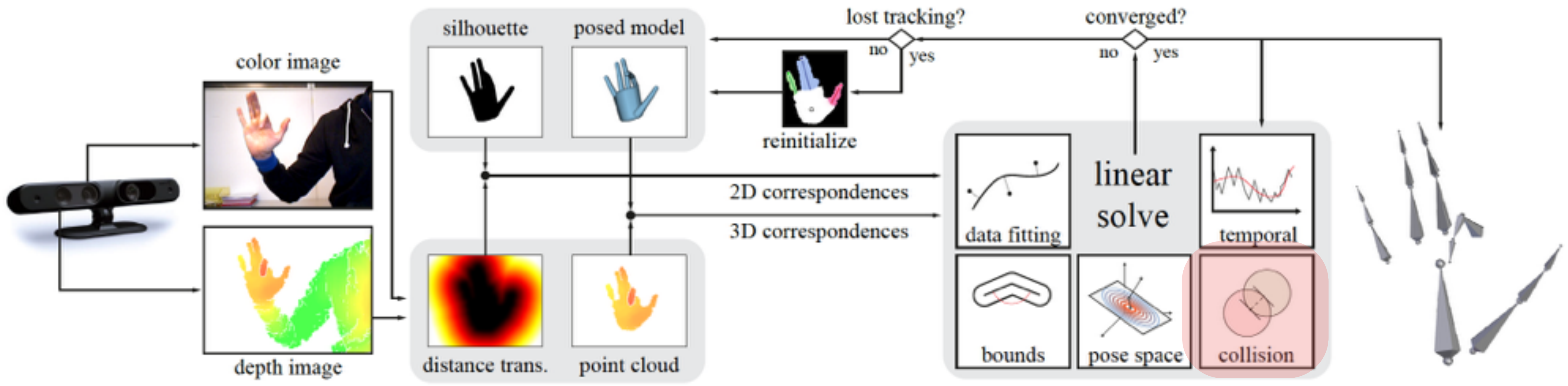
$$\min_{\boldsymbol{\theta}} \underbrace{E_{\text{points}} + E_{\text{silh.}} + E_{\text{wrist}}}_{\text{Fitting terms}} + \underbrace{E_{\text{pose}} + E_{\text{kin.}} + E_{\text{temporal}}}_{\text{Prior terms}}$$

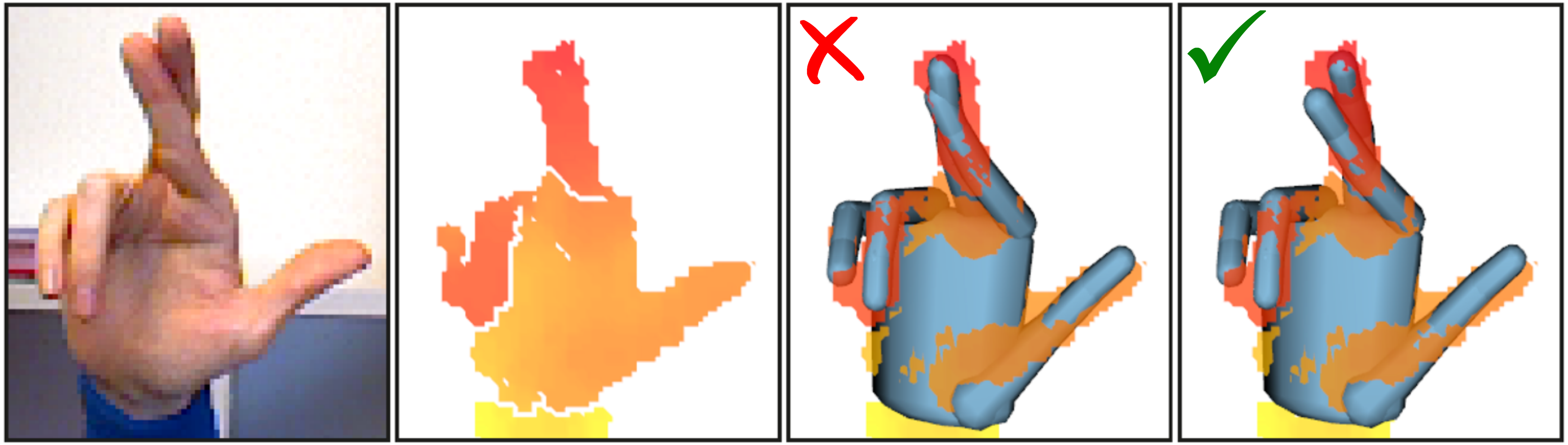
Joint Bounds Energy



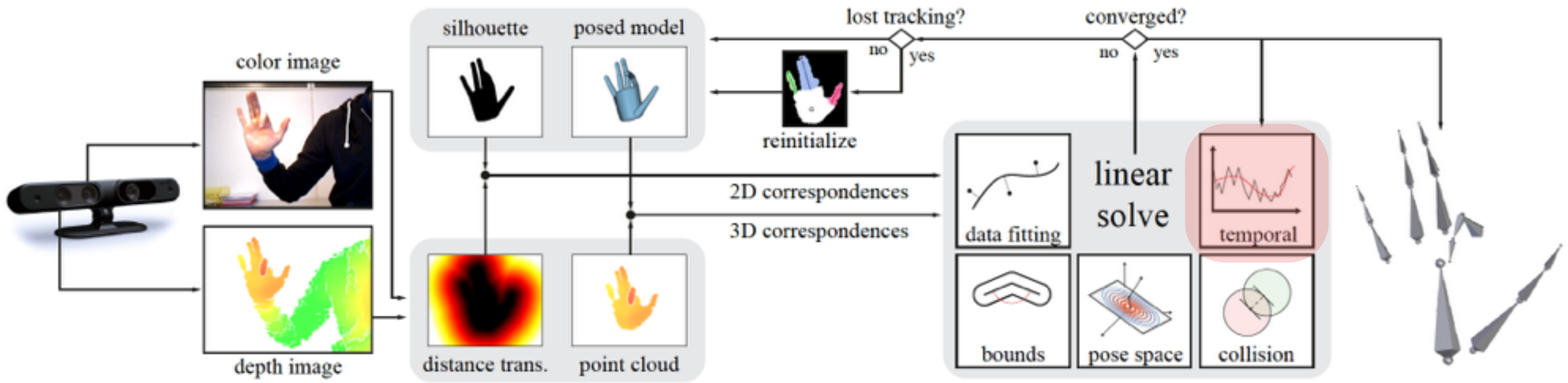


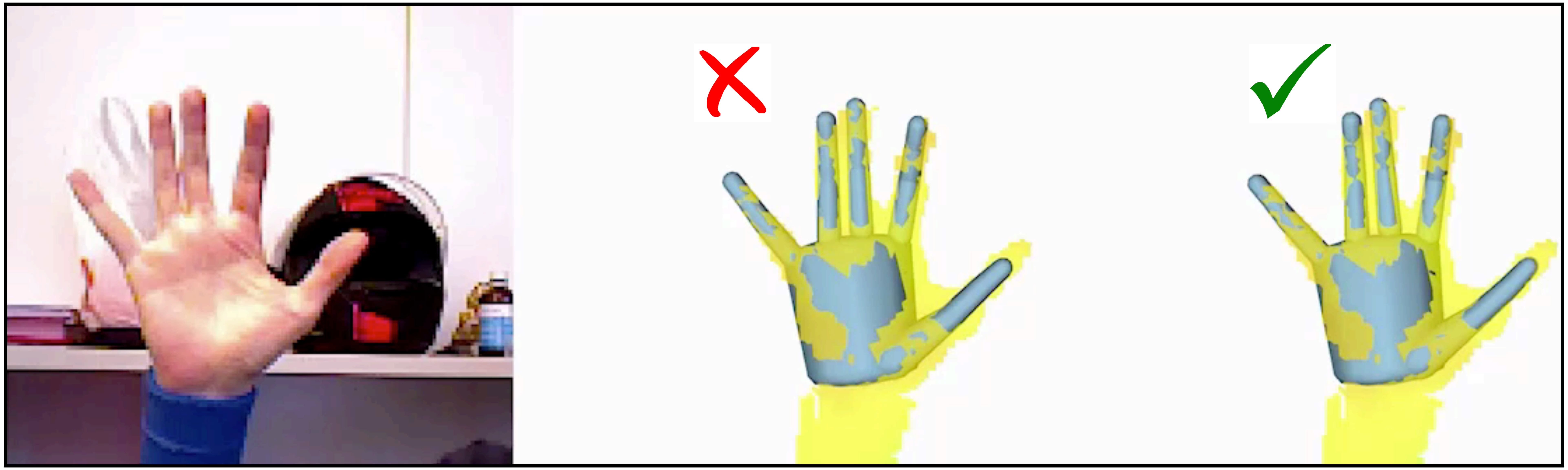
Collision Energy



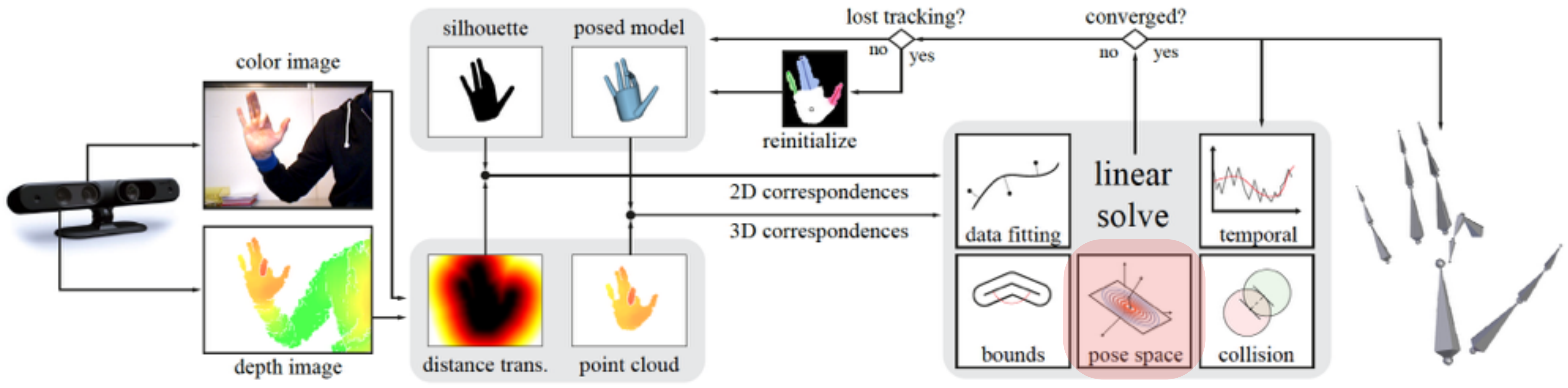


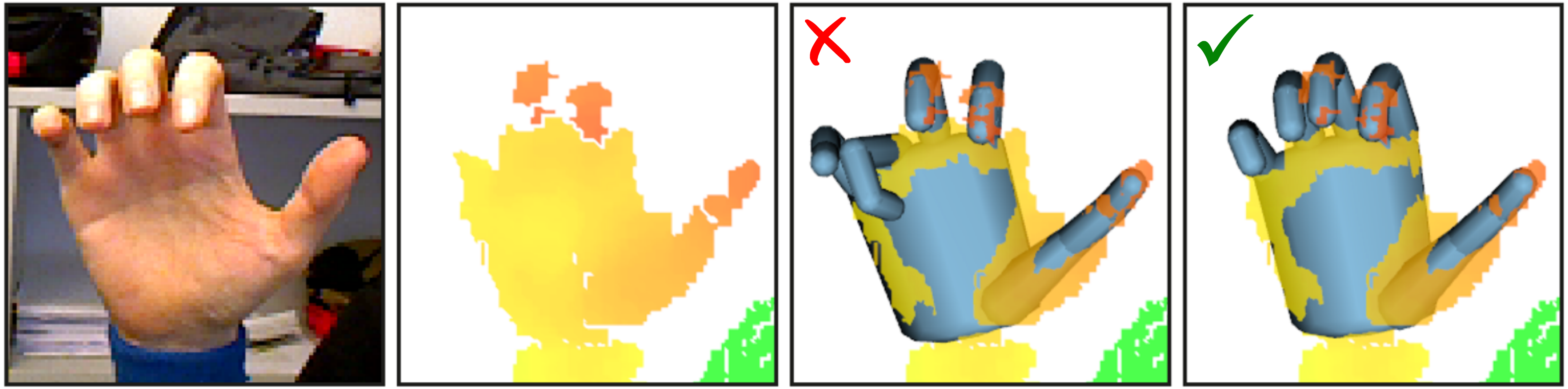
Temporal Coherence Energy





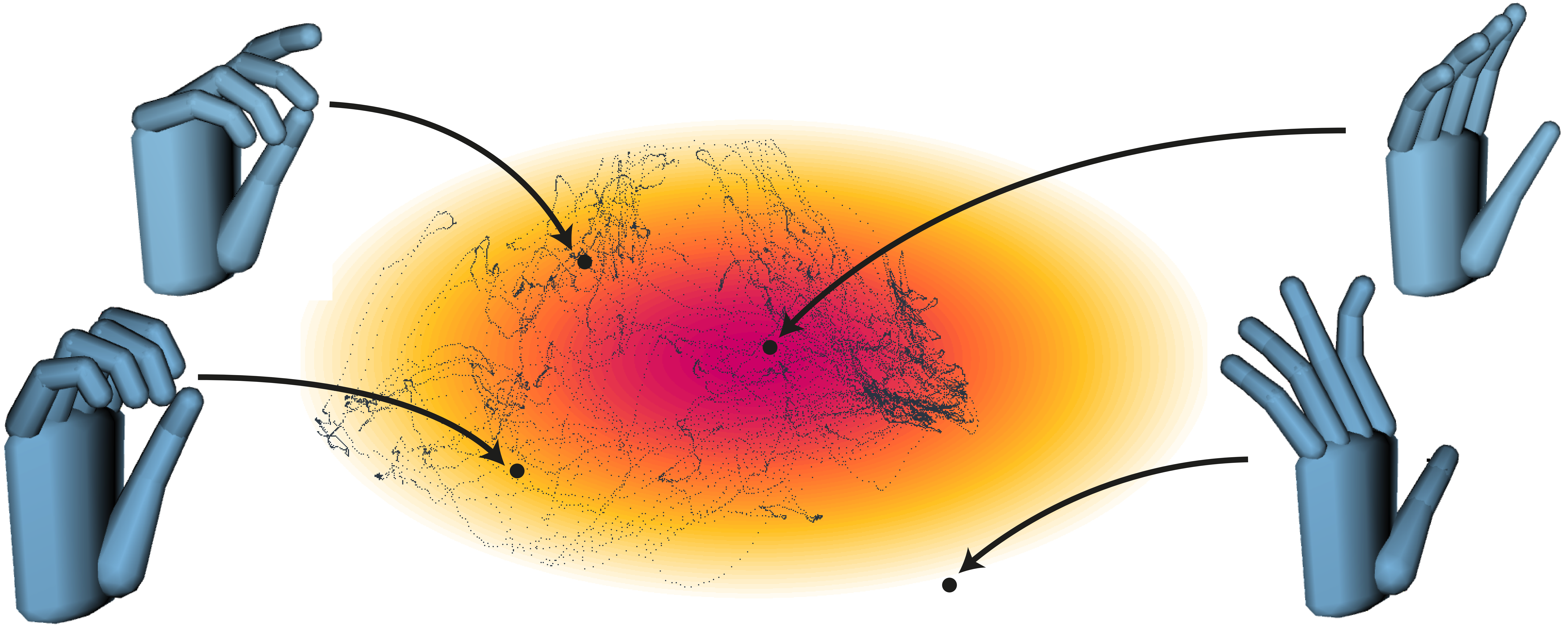
Statistical Pose Prior





encodes correlation across joints

Statistical Pose Prior



Recorded by VICON tracking system [Schroder ICRA'14]
(they are accurate… for the person that have been recorded for)

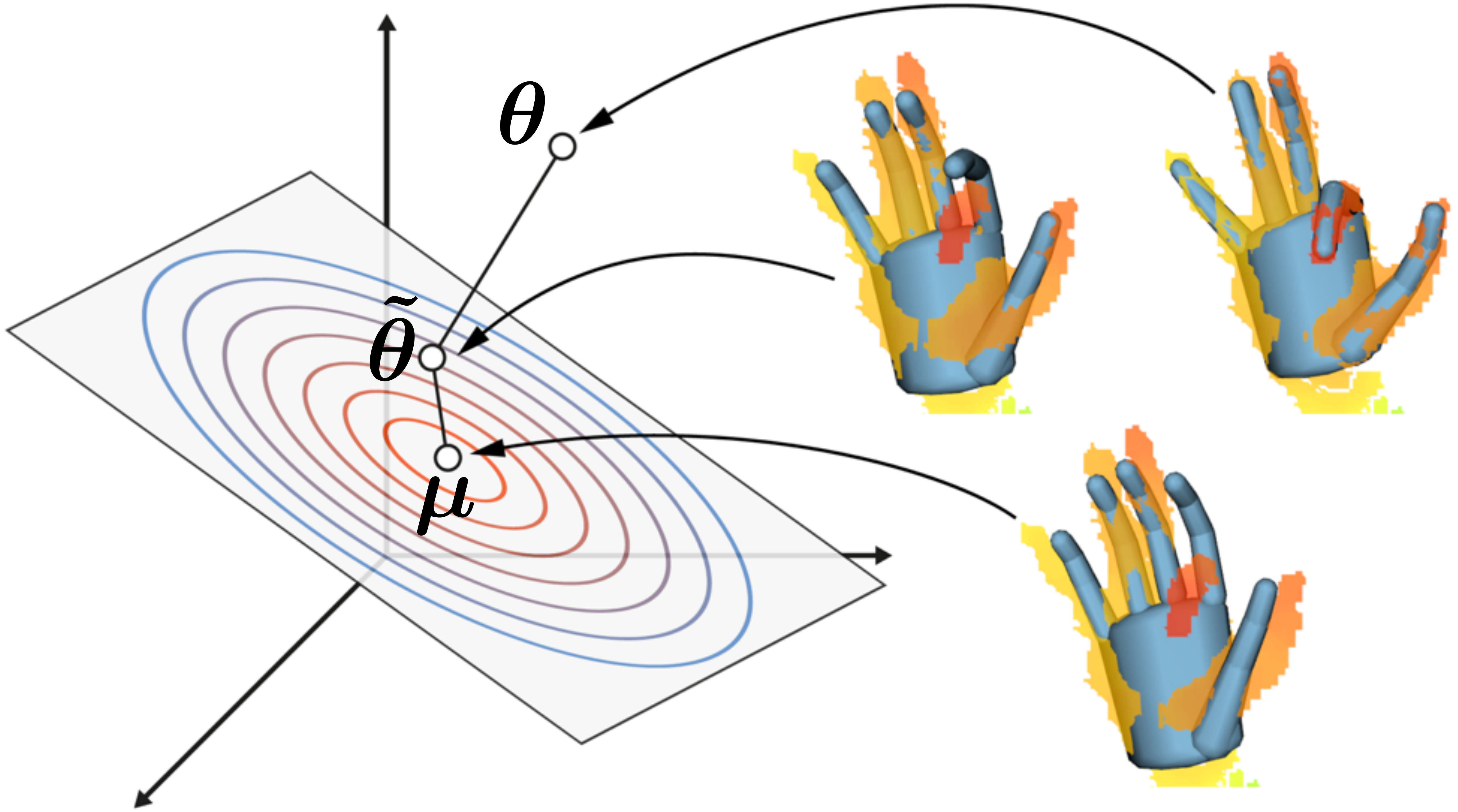
Statistical Pose Prior (Soft)

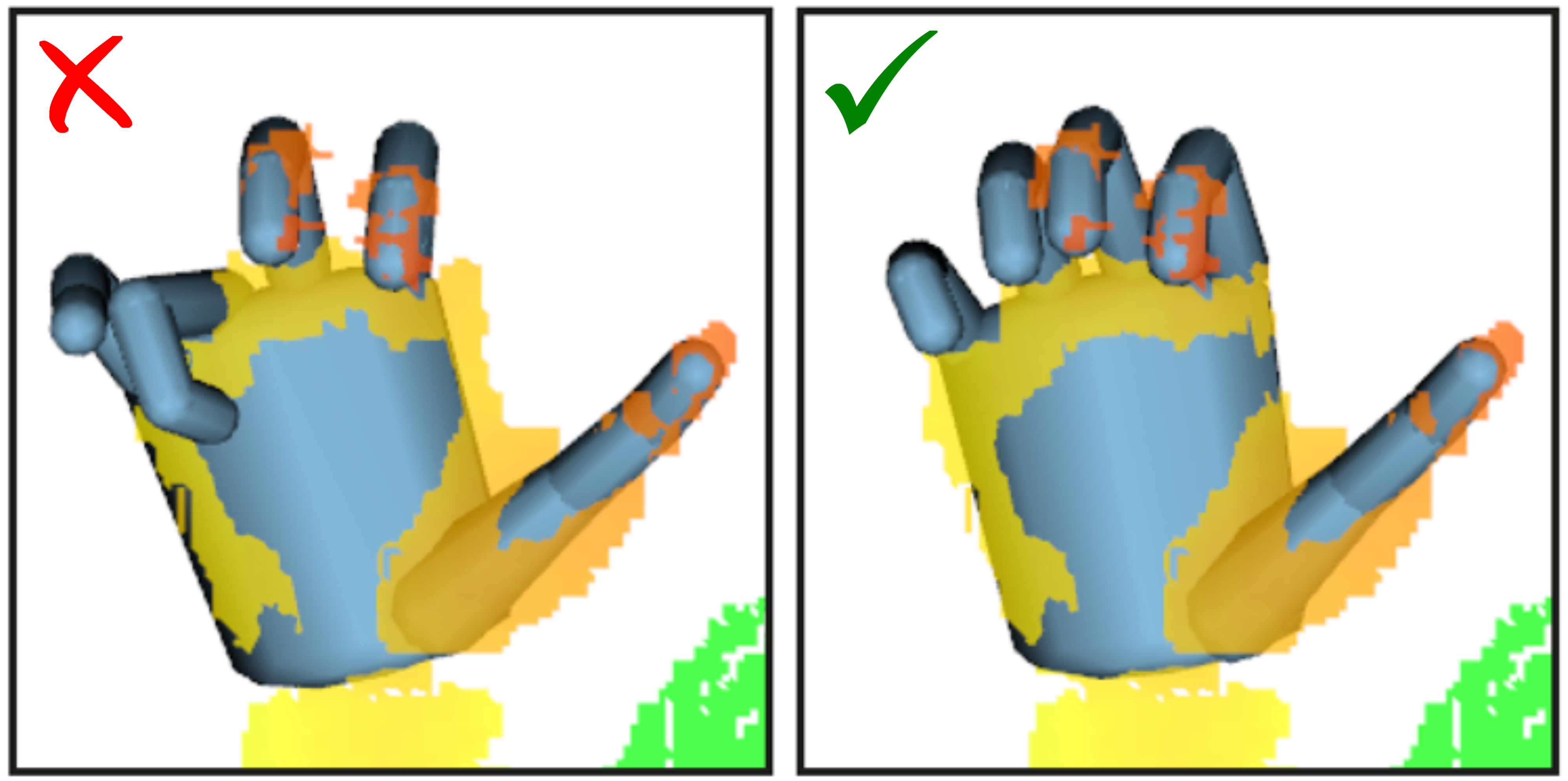
$$E_{\text{pose}} = \omega_4 \|\boldsymbol{\theta} - (\boldsymbol{\mu} + \Pi_{\mathcal{P}} \tilde{\boldsymbol{\theta}})\|_2^2 + \omega_5 \|\boldsymbol{\Sigma} \tilde{\boldsymbol{\theta}}\|_2^2$$

we optimize the current pose

So that it is **similar** to a reconstructed pose from the low dimensional subspace

but when DOF are unconstrained we would like to restore the neutral (i.e. mean) pose.

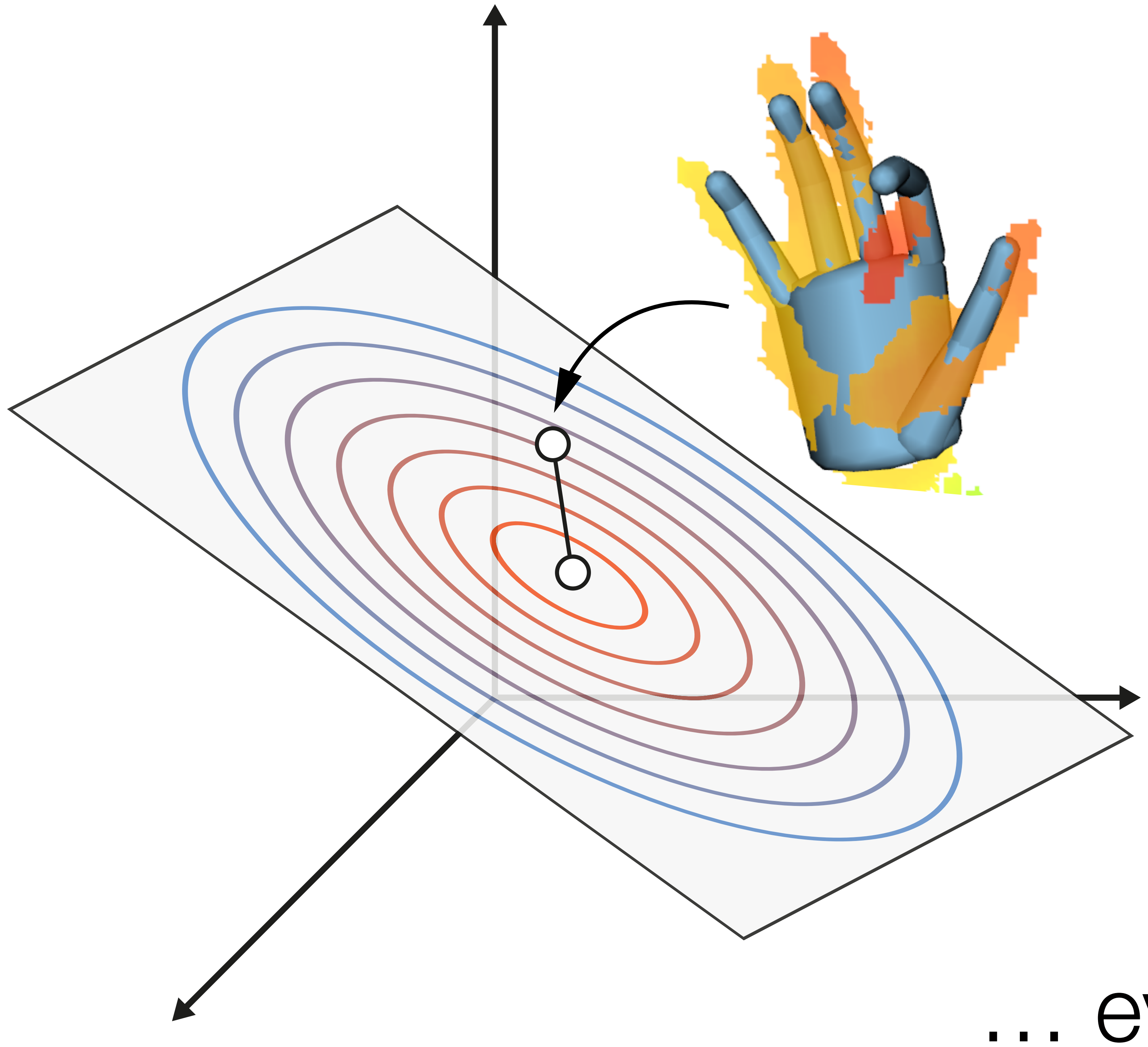


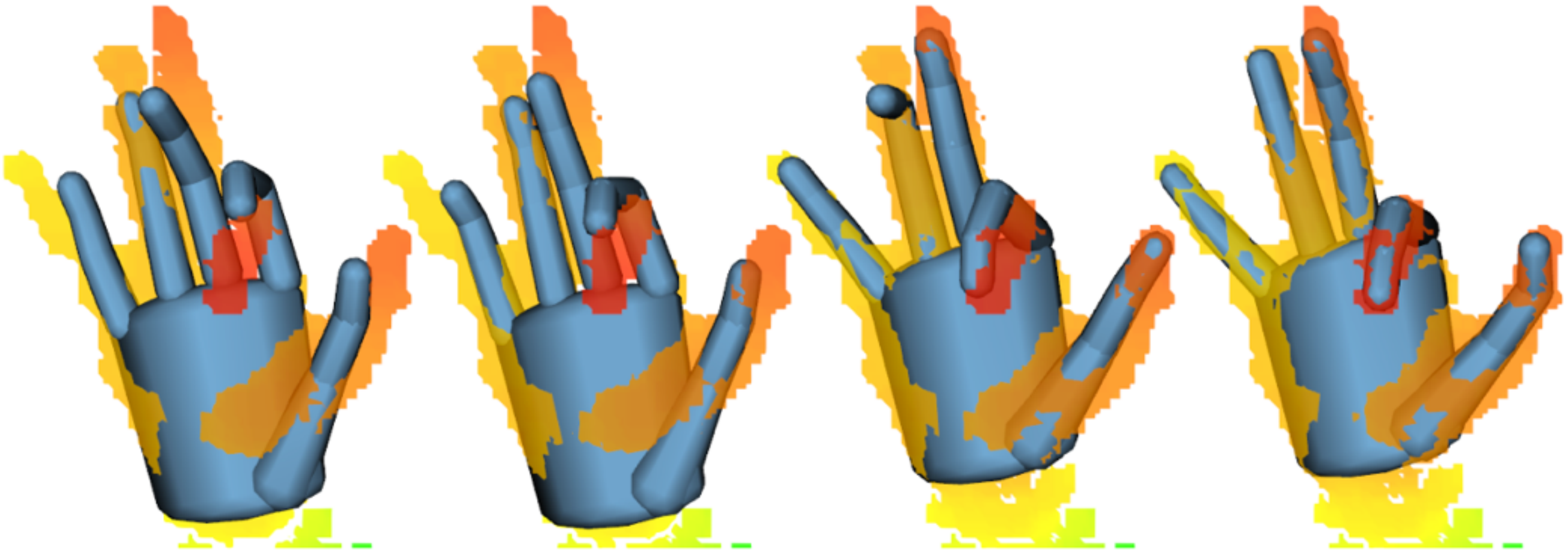


Statistical Pose Prior (Hard)

$$E_{\text{pose}} = \quad \omega_4 \|\boldsymbol{\theta} - (\boldsymbol{\mu} + \Pi_{\mathcal{P}} \tilde{\boldsymbol{\theta}})\|_2^2$$

$\omega_4 = \infty$ only optimize in the subspace (i.e. variable replacement)

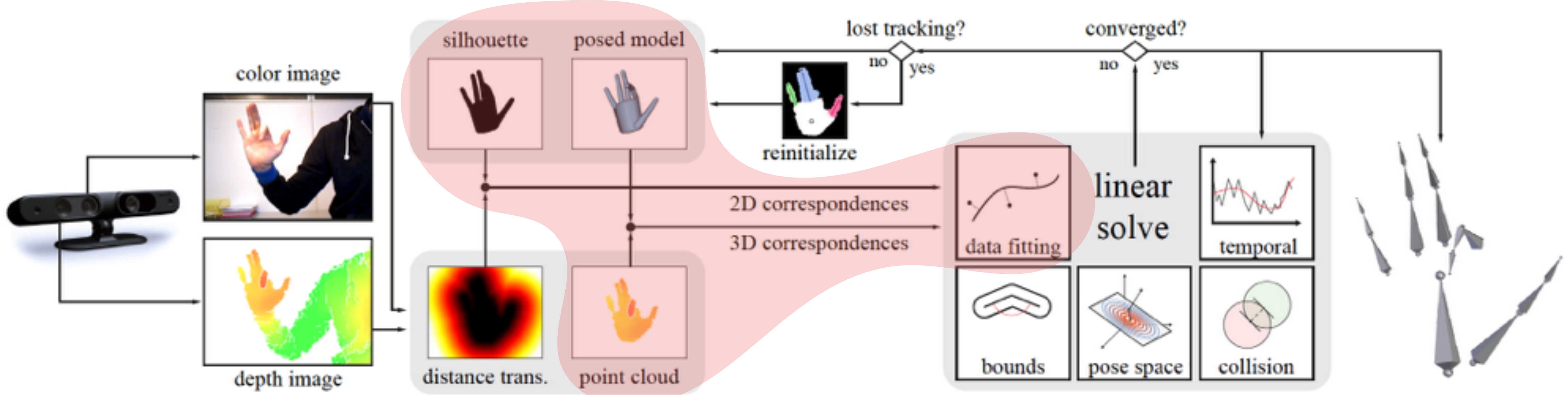




89%, #PCA=4	96%, #PCA=6	99%, #PCA=9	79%, #PCA=2
$\omega_4 = 10^8$	$\omega_4 = 10^8$	$\omega_4 = 10^8$	$\omega_4 = 10^2$

… even with a high number of bases the animation remains **stiff**!!!

CPU/GPU Optimization



$$\min_{\delta\theta} \sum_i \|\mathbf{f}_i + \mathbf{J}_i \delta\theta\|_2^2 \qquad \delta\theta = \Big(\sum_i \mathbf{J}_i^T \mathbf{J}_i\Big)^{-1} \Big(\sum_i -\mathbf{J}_i^T \mathbf{f}_i\Big)$$

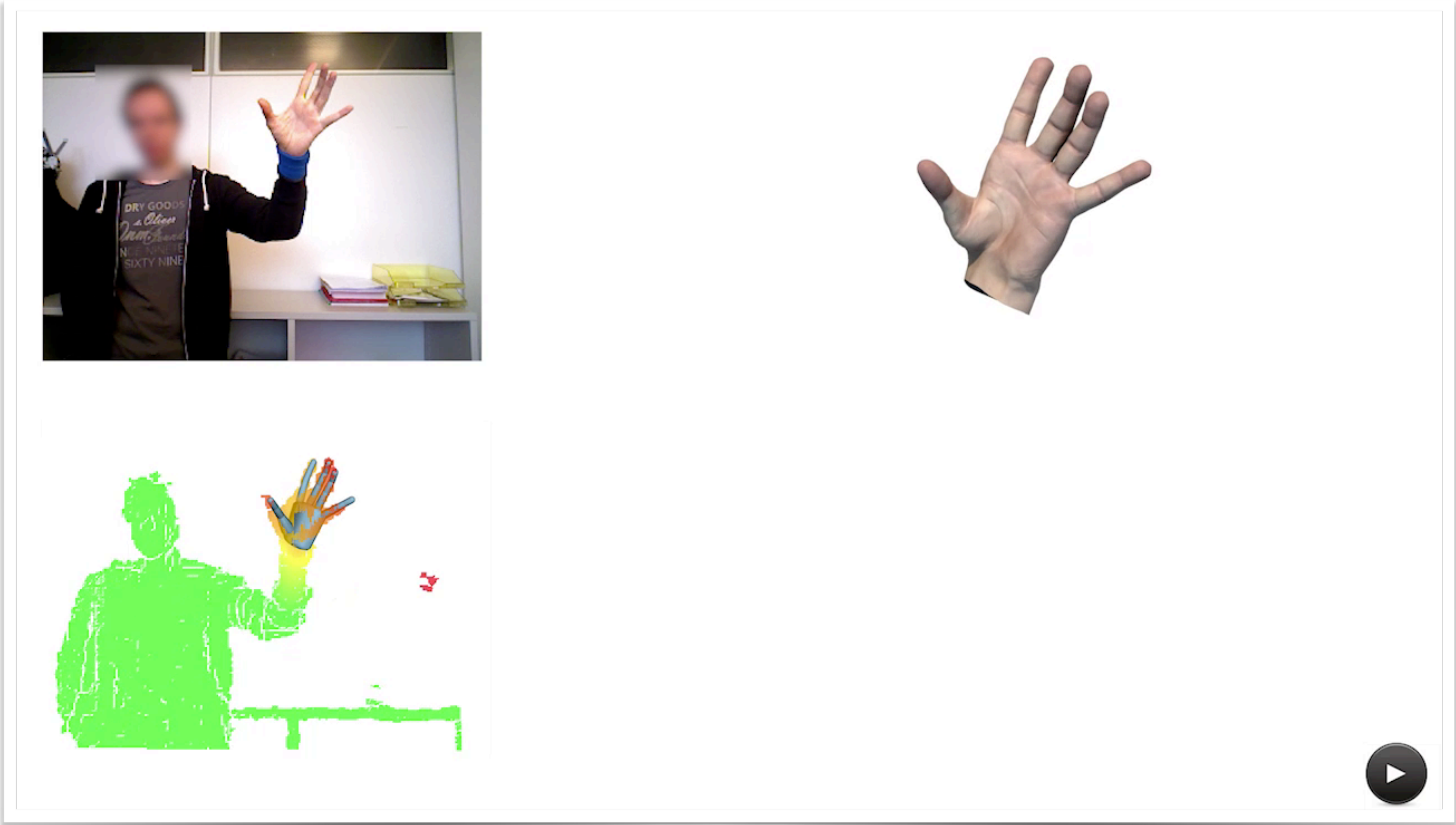
$$\bar{E}_{\text{silh.}} = \omega_2 \sum_{\mathbf{p} \in \mathcal{S}_r} \left(\mathbf{n}^T \left(\mathbf{J}_{\text{persp}}(\mathbf{x}) \mathbf{J}_{\text{skel}}(\mathbf{x}) \delta\boldsymbol{\theta} + (\mathbf{p} - \Pi_{\mathcal{S}_s}(\mathbf{p}, \boldsymbol{\theta}))\right)\right)^2$$

$|\mathcal{S}_r| \approx 20k$!!!!

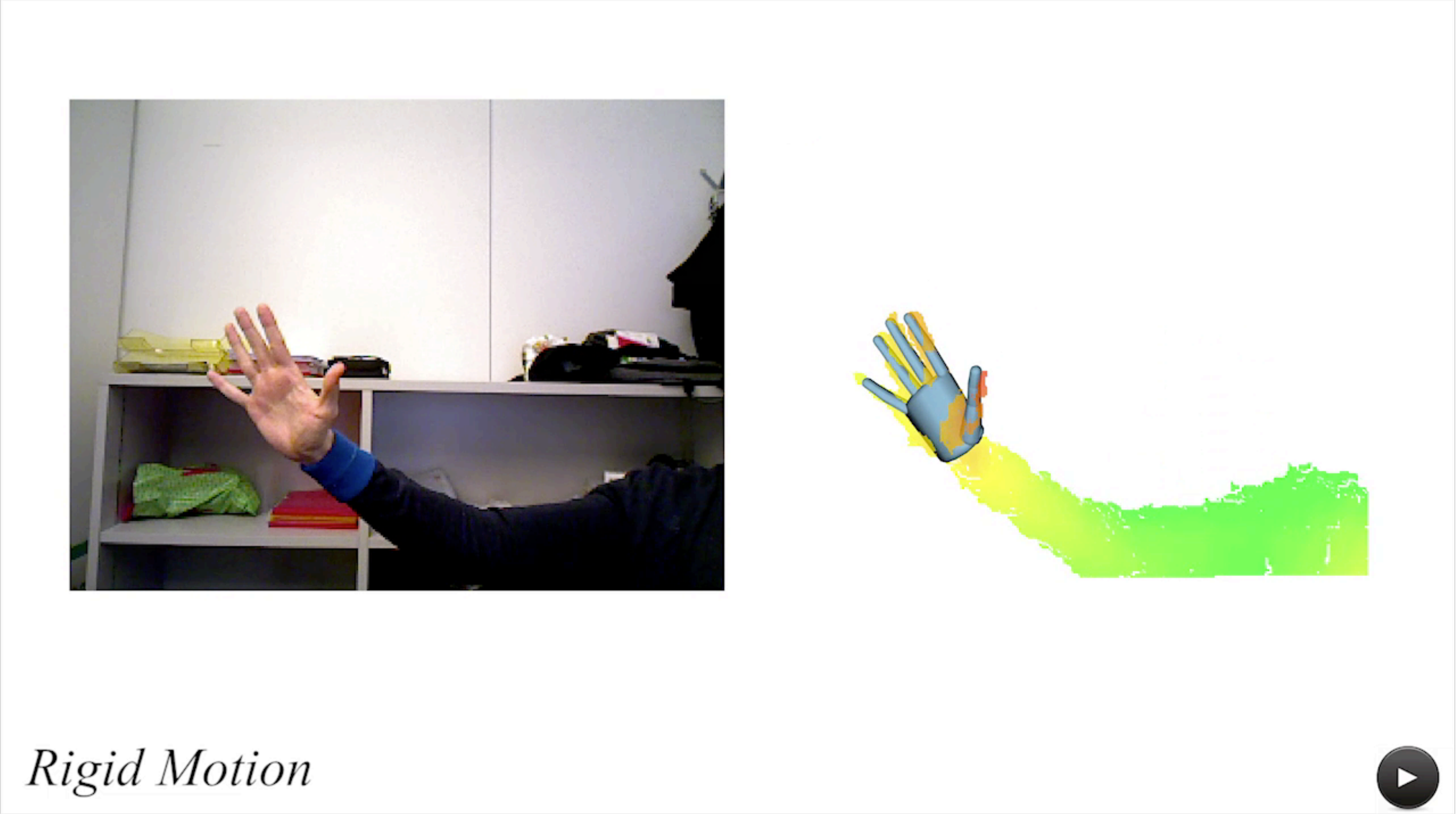
$|\mathbf{J}_{\text{silh}}| \approx 20k \times 26$ $\qquad$ $|\mathbf{J}_{\text{silh}}^T \mathbf{J}_{\text{silh}}| = 26 \times 26$

Results and Limitations

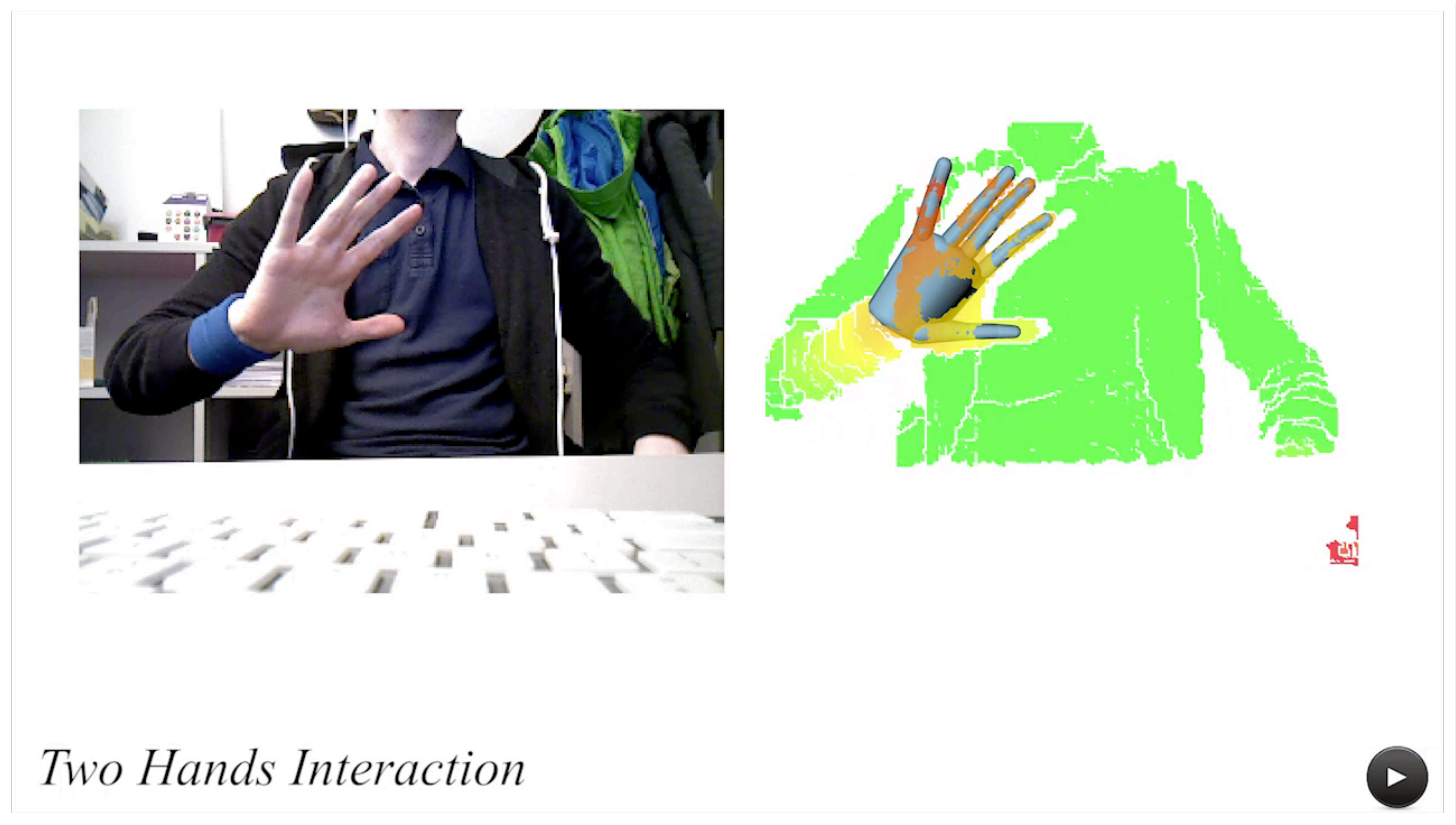
Motion Transfer to Rig



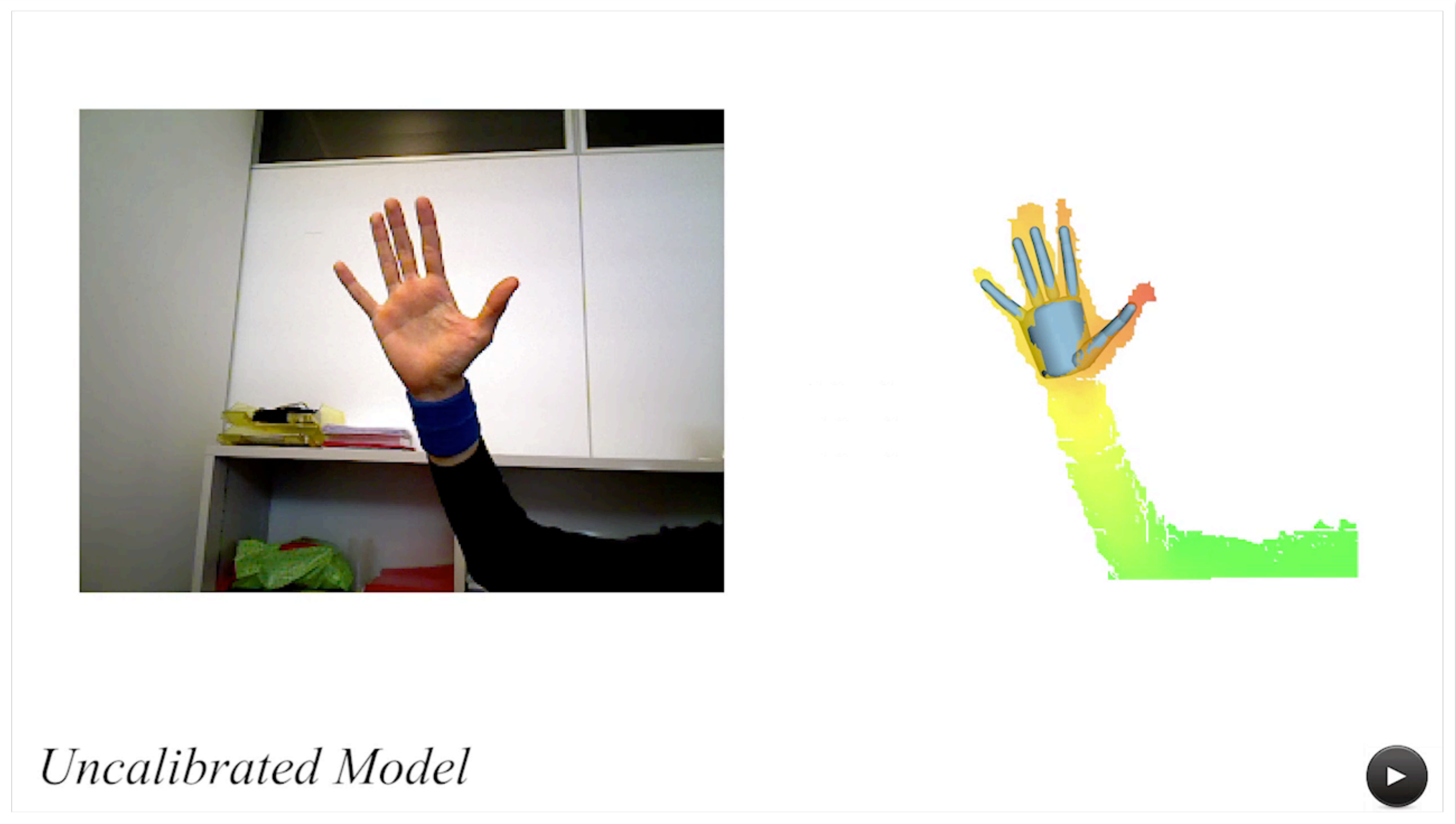
Tracking with Fast Motion



Tracking of Interacting Hands



Limitation: Calibration



Limitations: Fist Rotation

State of the Art - Evaluations


Real-Time Hand Tracking using Synergistic Inverse Kinematics

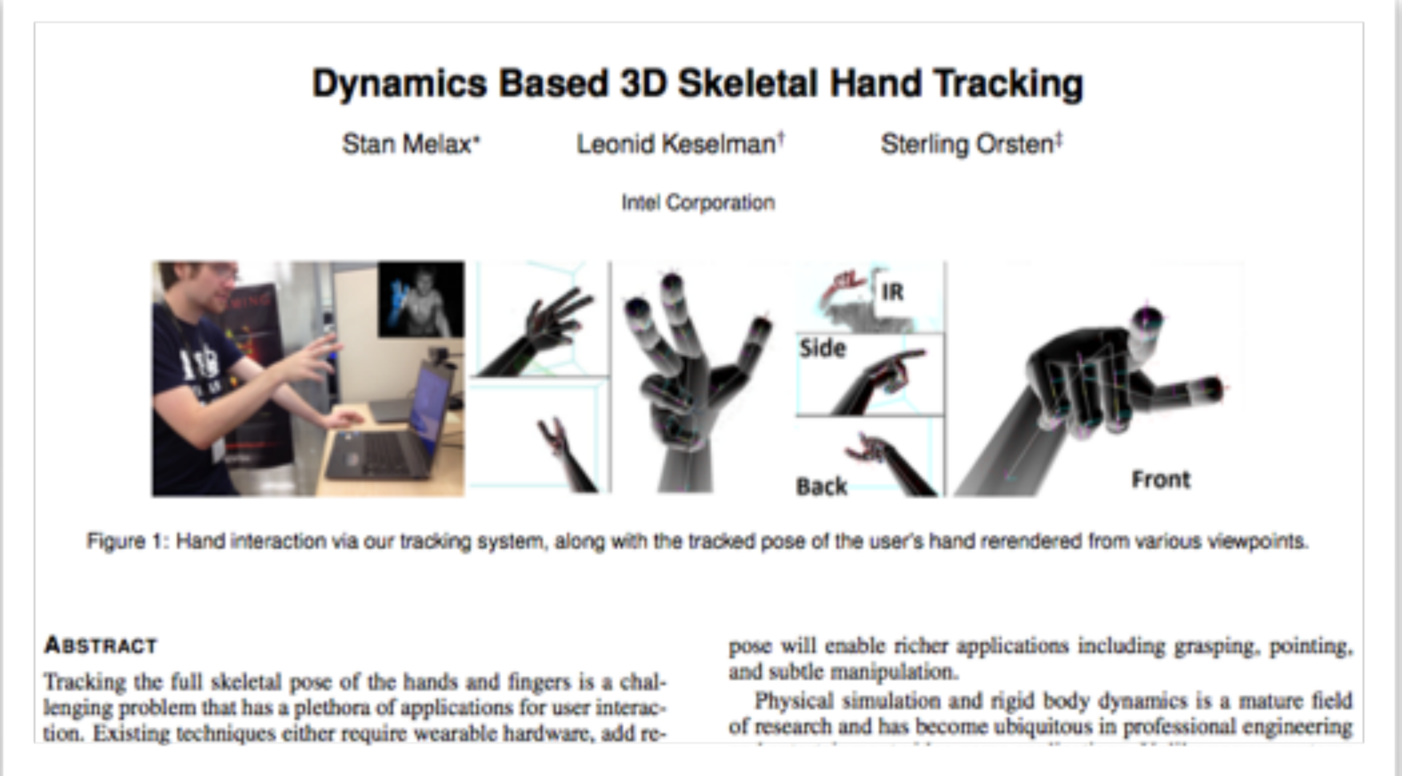
Matthias Schröder[1], Jonathan Maycock[2], Helge Ritter[2], Mario Botsch[1]
[1]Computer Graphics & Geometry Processing Group, [2]Neuroinformatics Group, Bielefeld University, Germany

Abstract—We present a method for real-time bare hand tracking that utilizes natural hand synergies to reduce the complexity and improve the plausibility of the hand posture estimation. The hand pose and posture are estimated by fitting a virtual hand model to the 3D point cloud obtained from a Kinect camera using an inverse kinematics approach. We use real human hand movements captured with a Vicon motion tracking system as the ground truth for deriving natural hand synergies based on principal component analysis. These synergies are integrated in the tracking scheme by optimizing the posture in a reduced parameter space. Tracking in this reduced space combined with joint limit avoidance constrains the posture estimation to natural hand articulations. The information loss associated with dimension reduction can be dealt with by employing a hierarchical optimization scheme. We show that our synergistic hand tracking approach improves runtime performance and increases the quality of the posture estimation.

Fig. 1. The user's hand posture is estimated in real-time by fitting a 3D hand model to the Kinect point cloud using inverse kinematics.

with a linear prediction step and a cylinder model; then, we employ joint limit avoidance to ensure physically plausible hand postures. We also propose a highly robust hierarchical optimization scheme. Finally we analyze the performance of

[Shroder ICRA'14] Subspace ICP


Dynamics Based 3D Skeletal Hand Tracking

Stan Melax* Leonid Keselman† Sterling Orsten‡
Intel Corporation

Figure 1: Hand interaction via our tracking system, along with the tracked pose of the user's hand rerendered from various viewpoints.

ABSTRACT

Tracking the full skeletal pose of the hands and fingers is a challenging problem that has a plethora of applications for user interaction. Existing techniques either require wearable hardware, add re-

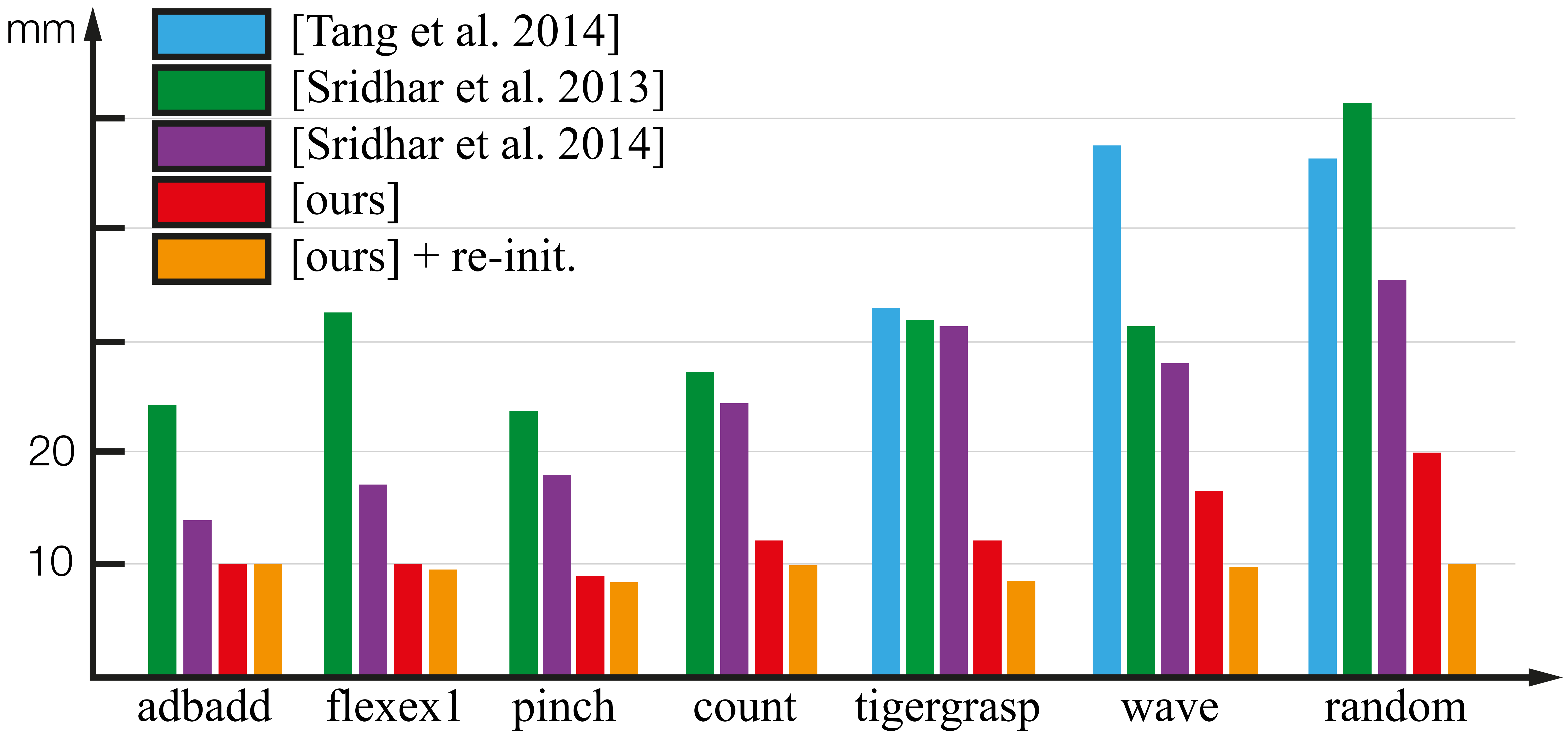
pose will enable richer applications including grasping, pointing, and subtle manipulation.

Physical simulation and rigid body dynamics is a mature field of research and has become ubiquitous in professional engineering

[Melax'14] Intel Perceptual SDK

Qualitative

Dexter-1 Dataset (MPI)

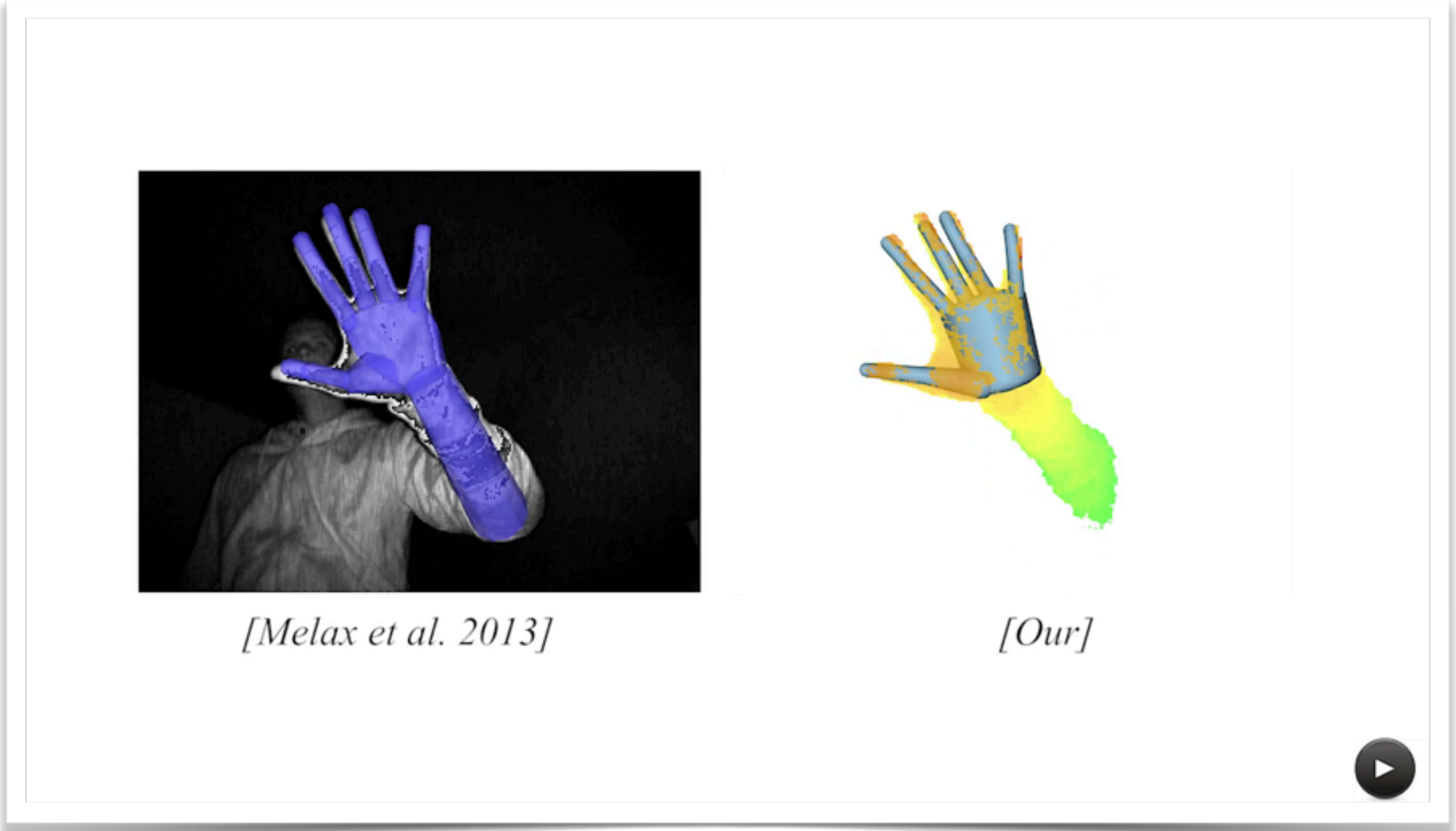


(re-initialization helps because Dexter-1 is a low frame-rate dataset… only 30Hz)

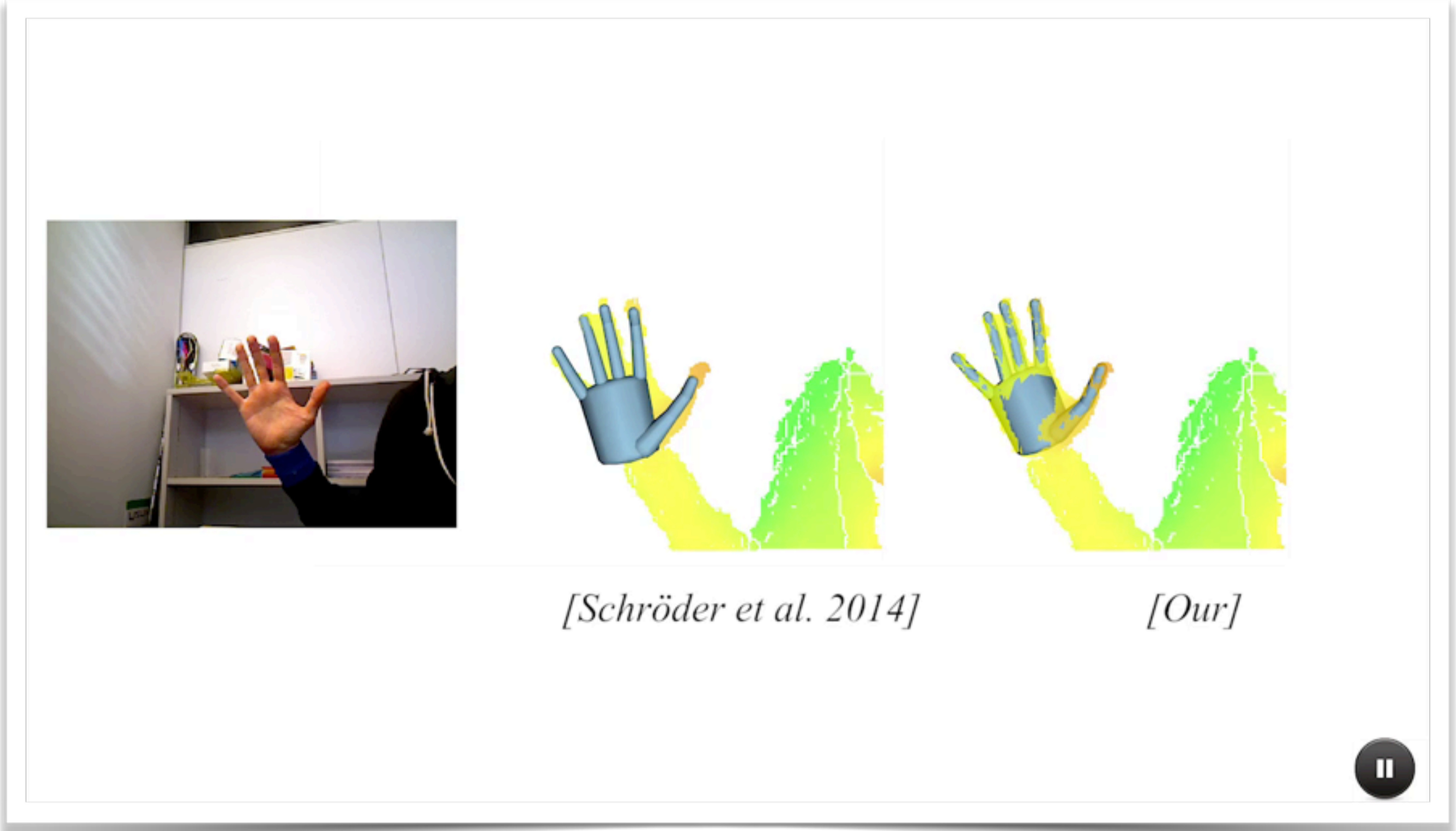
Quantitative

Qualitative Comparisons

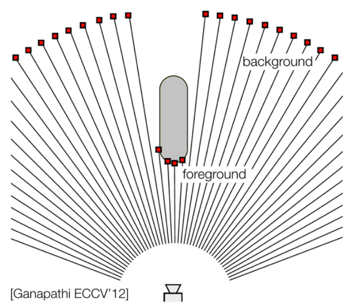
Convex Dynamics (Intel SDK)



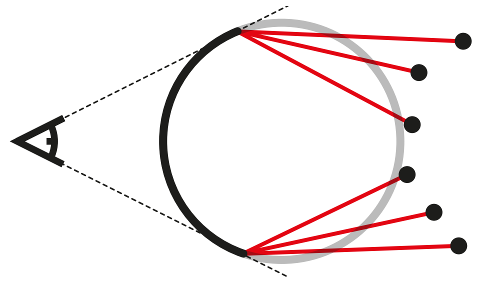
Subspace ICP



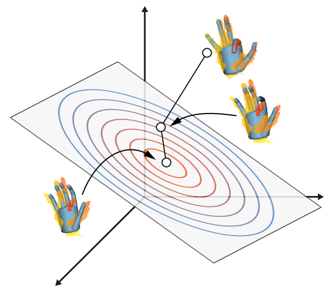
Conclusions



- combined 2D/3D registration (within ICP)



- occlusion-aware correspondences (ICP)



- regularization with statistical pose-space prior

- extensible and unified real-time solver (>60fps)

GitHub

- fully open source!

https://github.com/OpenGP/htrack

Shameless Advertisement

Who? Prof. Andrea Tagliasacchi and Brian Wyvill

What? MSc (PhD)

Where? University of Victoria

Who pays? Google

Language? English

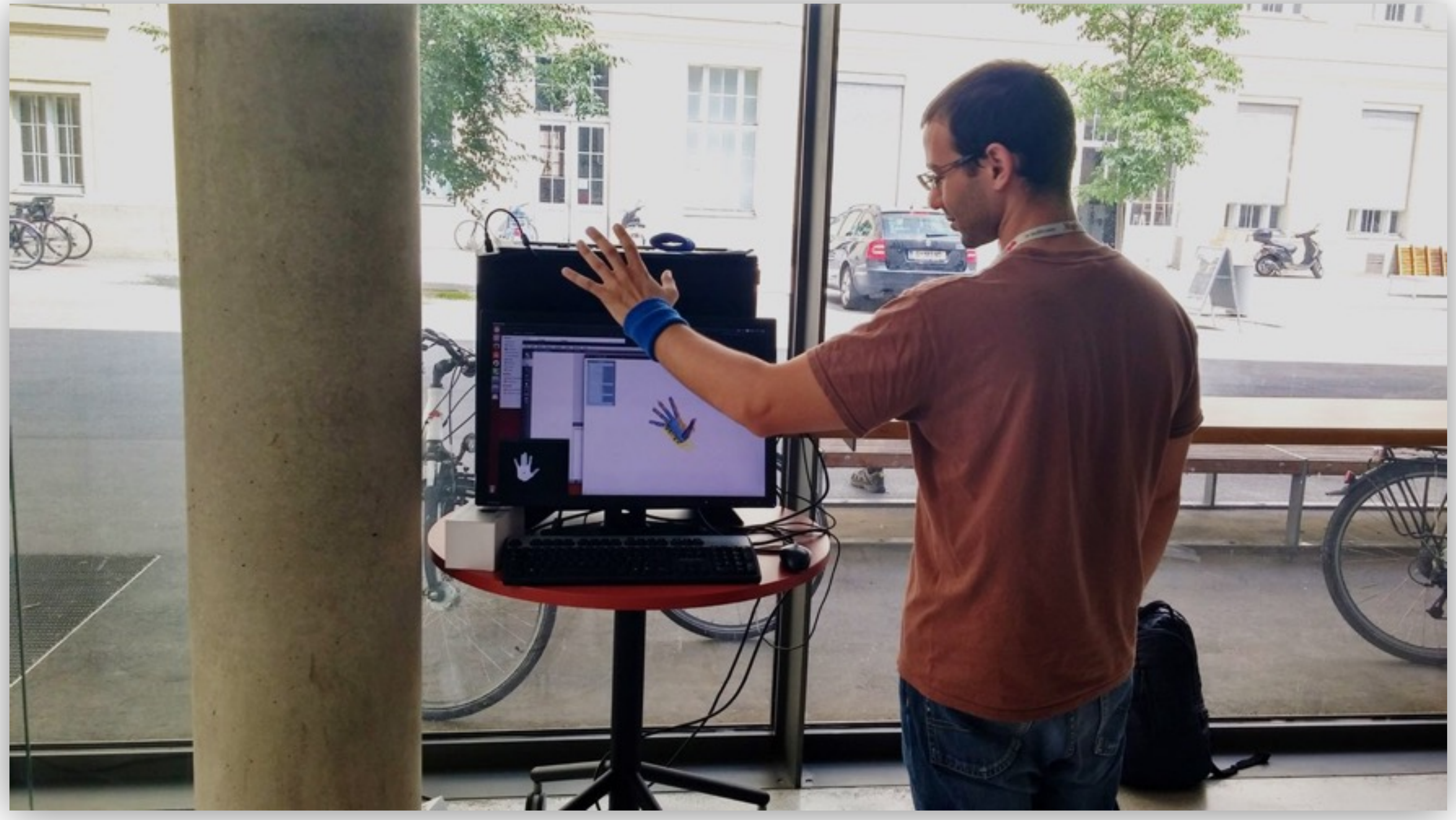
https://www.csc.uvic.ca/Program_Information/Graduate_Studies/msc_program.htm

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Universität Bielefeld

Thank you!!!

Live demo at SGP’15!!!



https://github.com/OpenGP/htrack

Don’t be shy!!

Sofien

Matthias

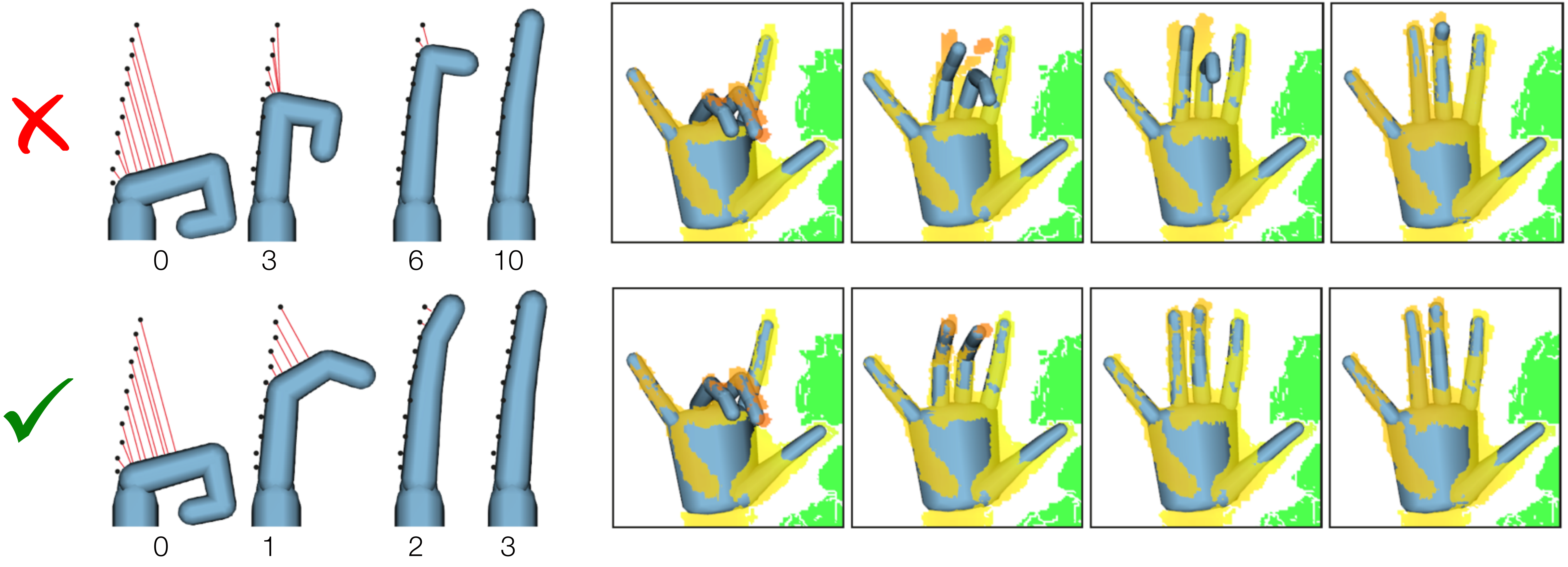
Mark

Anastasia

Extra Slides

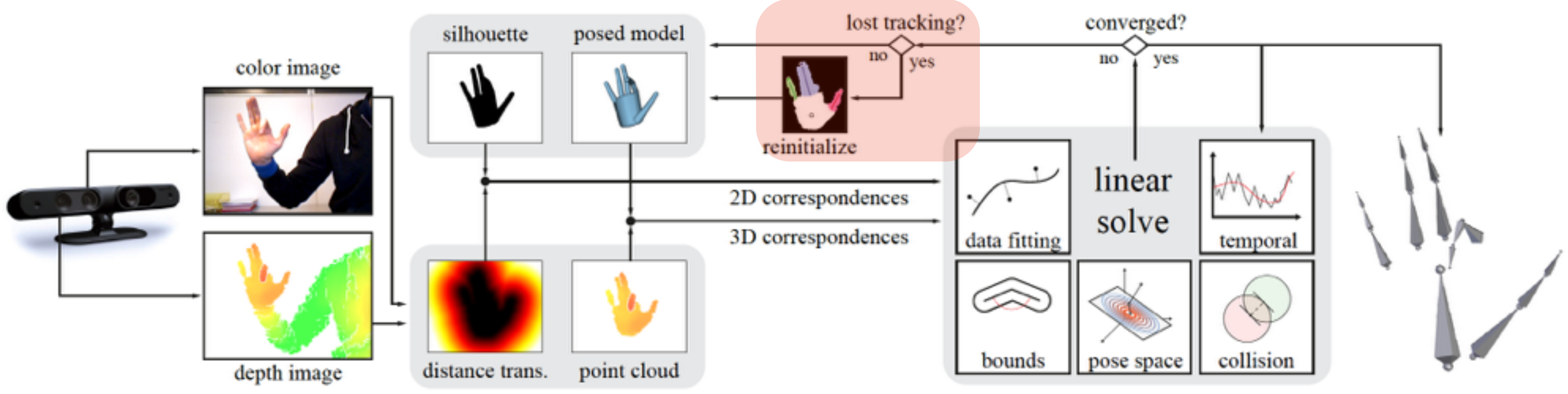
Statistical Prior: Side Effects

$$E_{\text{pose}} = \omega_4 \|\boldsymbol{\theta} - (\boldsymbol{\mu} + \Pi_{\mathcal{P}} \tilde{\boldsymbol{\theta}})\|_2^2 + \omega_5 \|\boldsymbol{\Sigma} \tilde{\boldsymbol{\theta}}\|_2^2$$



as this prior correlates joint angles the convergence speed increases by about 3x

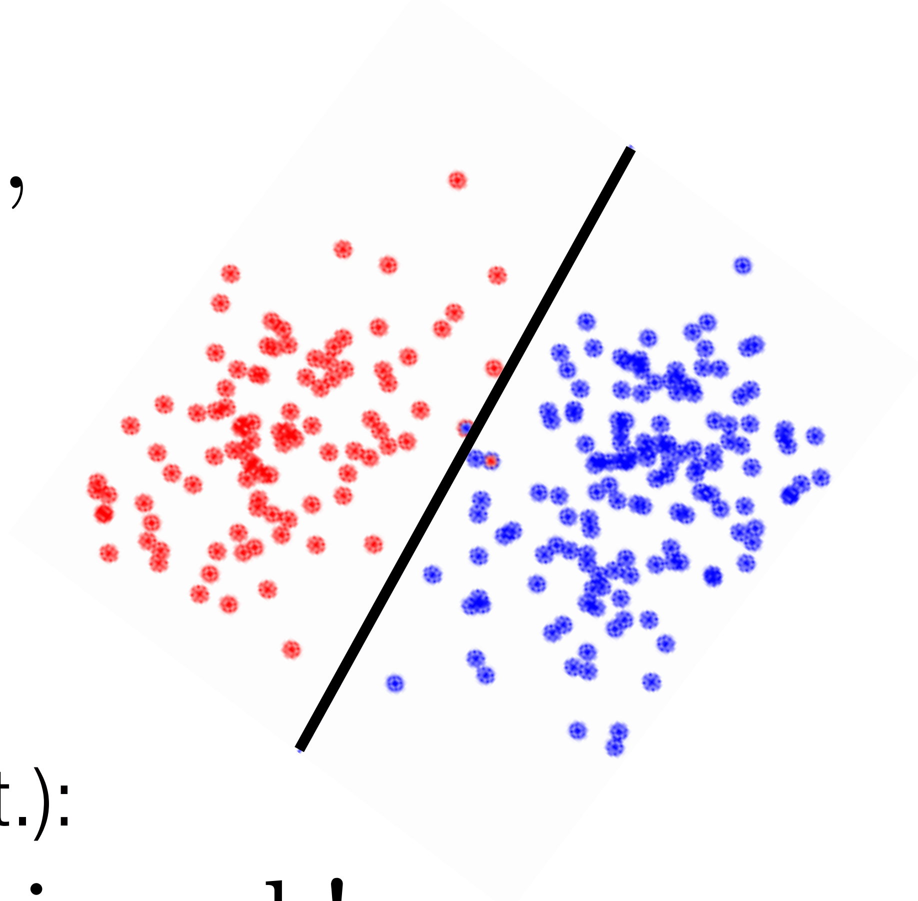
Reinitialization from Failure



Determine tracking failure [Melax'13]

$$\epsilon_1 = \sum_{\mathbf{x} \in \mathcal{X}_s} \|\mathbf{x} - \Pi_{\mathcal{M}}(\mathbf{x}, \boldsymbol{\theta})\|_2,$$

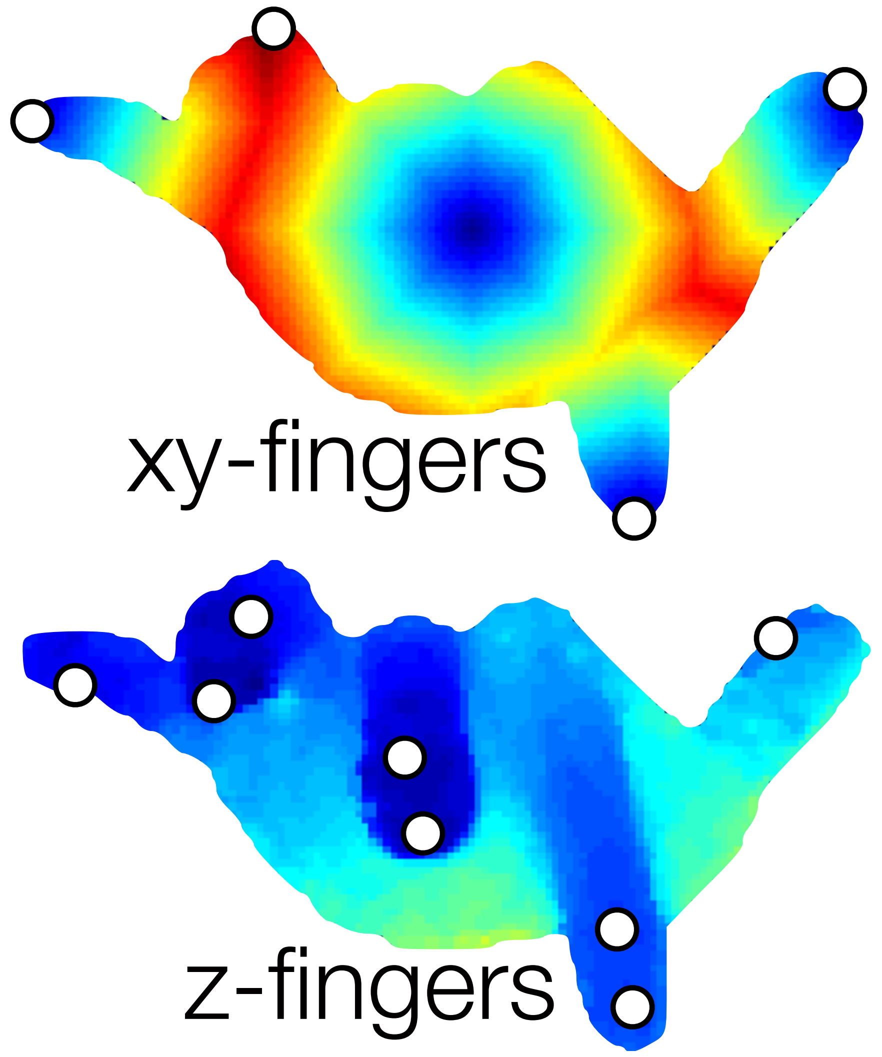
$$\epsilon_2 = \frac{\mathcal{S}_r \cap \mathcal{S}_s}{\mathcal{S}_r \cup \mathcal{S}_s}$$

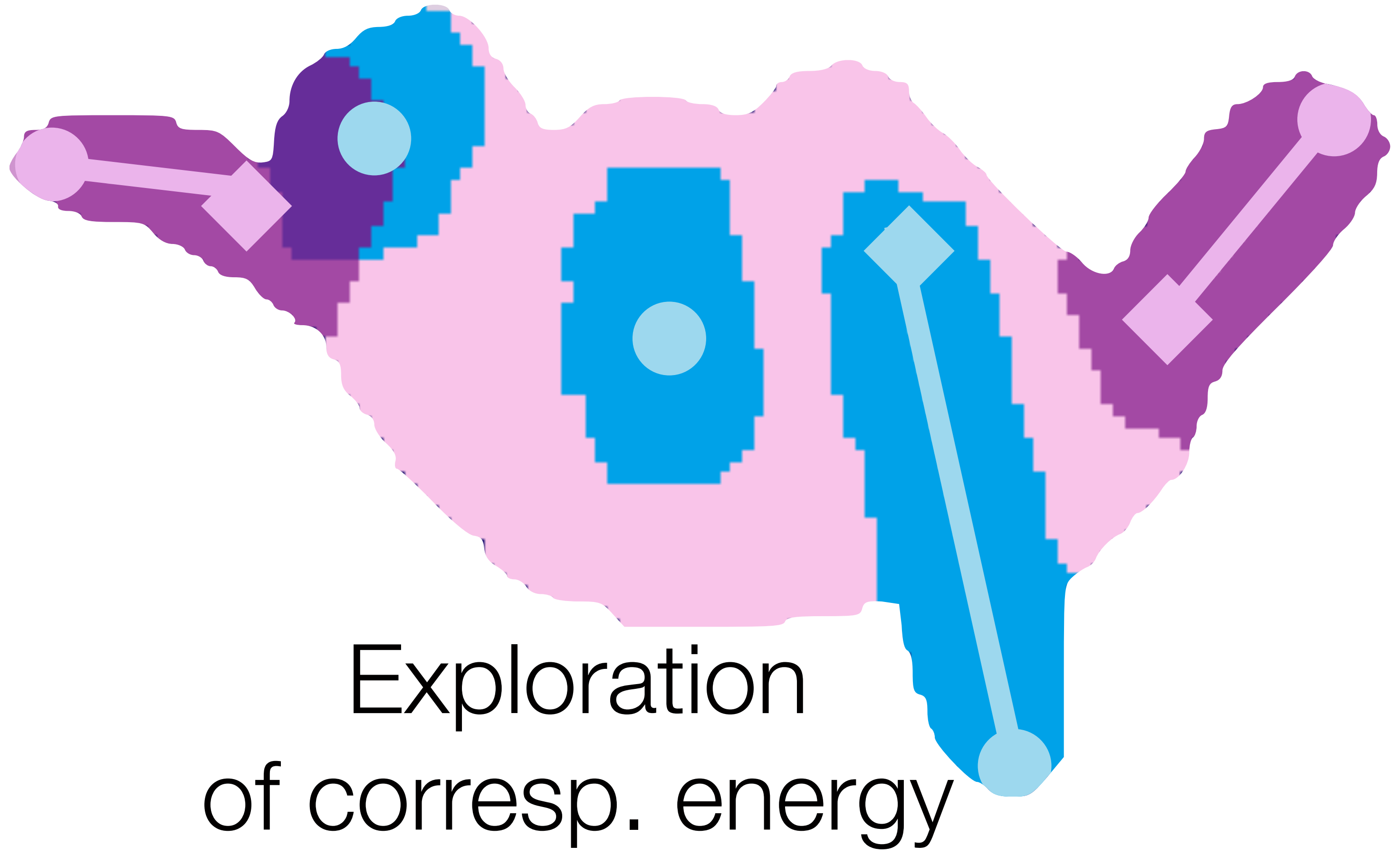


Logistic Regressor (optimize alpha s.t.):

$$2 \geq 1 + e^{-\alpha^T [\epsilon_1, \epsilon_2]} \rightarrow \text{tracking ok!}$$

Detection by ~[Qian et al. CVPR'14]





Particle Swarm