Performance Capture with Depth Cameras

Christian Theobalt

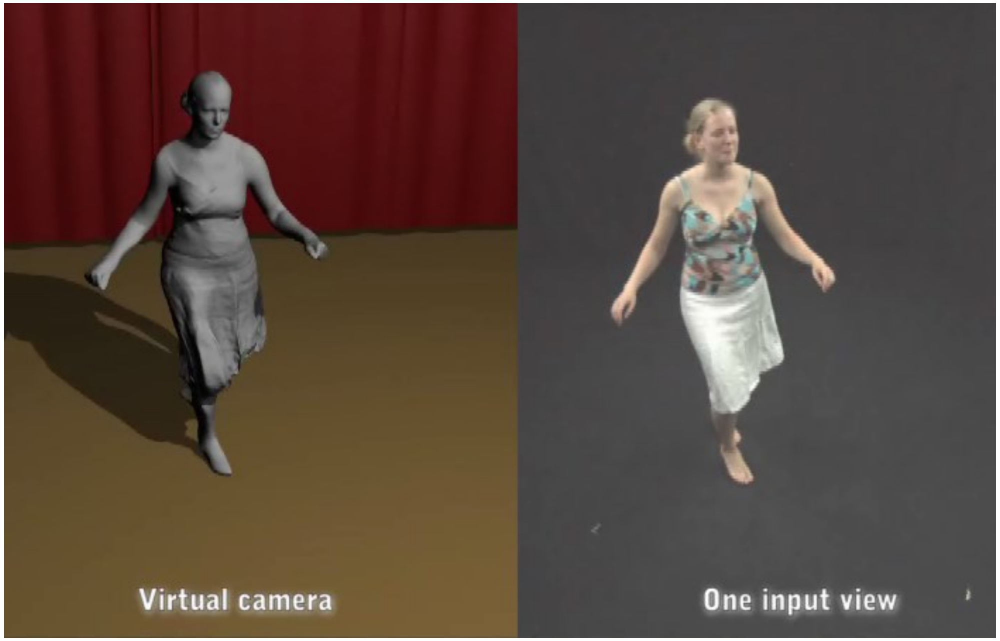
Graphics, Vision & Video

max planck institut informatik

UNIVERSITÄT DES SAARLANDES

VISUAL COMPUTING INSTITUTE

Marker-less Performance Capture



[de Aguiar et al,, SIGGRAPH 2008]

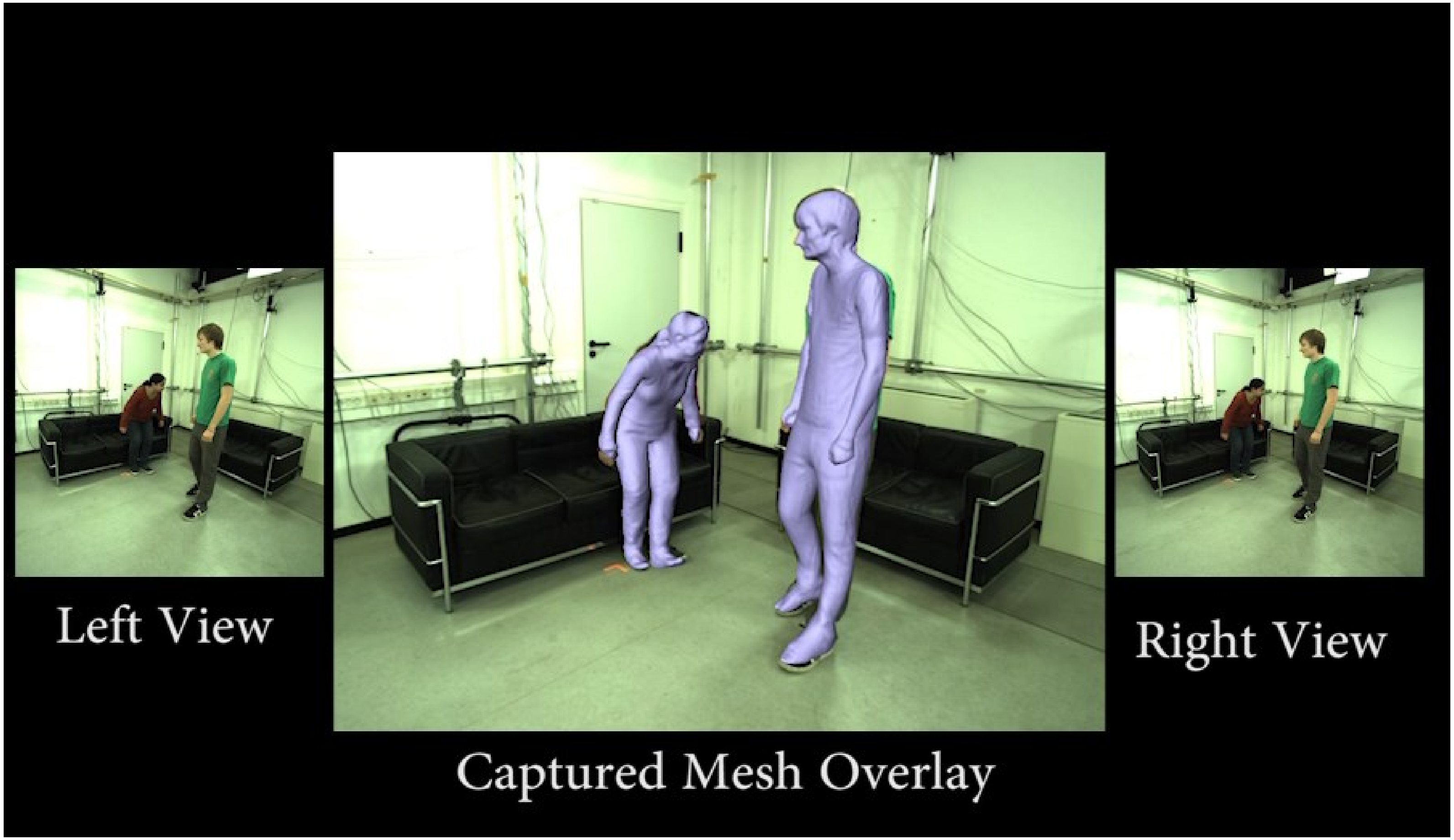
[Liu et al. CVPR 2011, PAMI 2013]

- High-quality shape, motion and texture of people in general clothing
- Input: multi-view video, no markers
- Exceed capabilities of existing Mocap technology by far

High-quality Binocular Face Capture

L.Valgaerts, C. Wu, A. Bruhn, H.-P. Seidel, C. Theobalt, Binocular Facial Performance Capture under Uncontrolled Illuminationm, SIGGRAPH Asia 2012

On-Set Performance Capture with a Stereo Camera

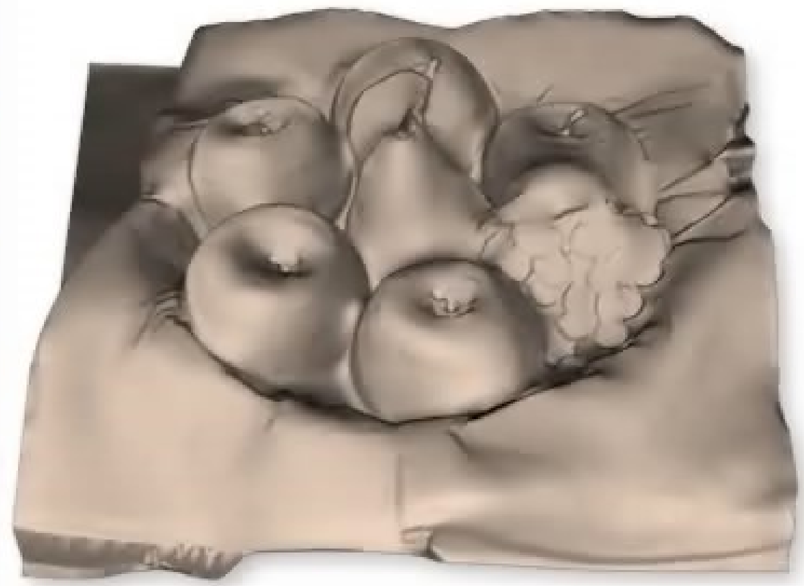


C. Wu, C. Stoll, L. Valgaerts, C. Theobalt, On-set Performance Capture of Multiple Actors with a Stereo Camera, SIGGRAPH Asia 2013

Space-Time Upsampling

- Space-Time Joint Upsamling

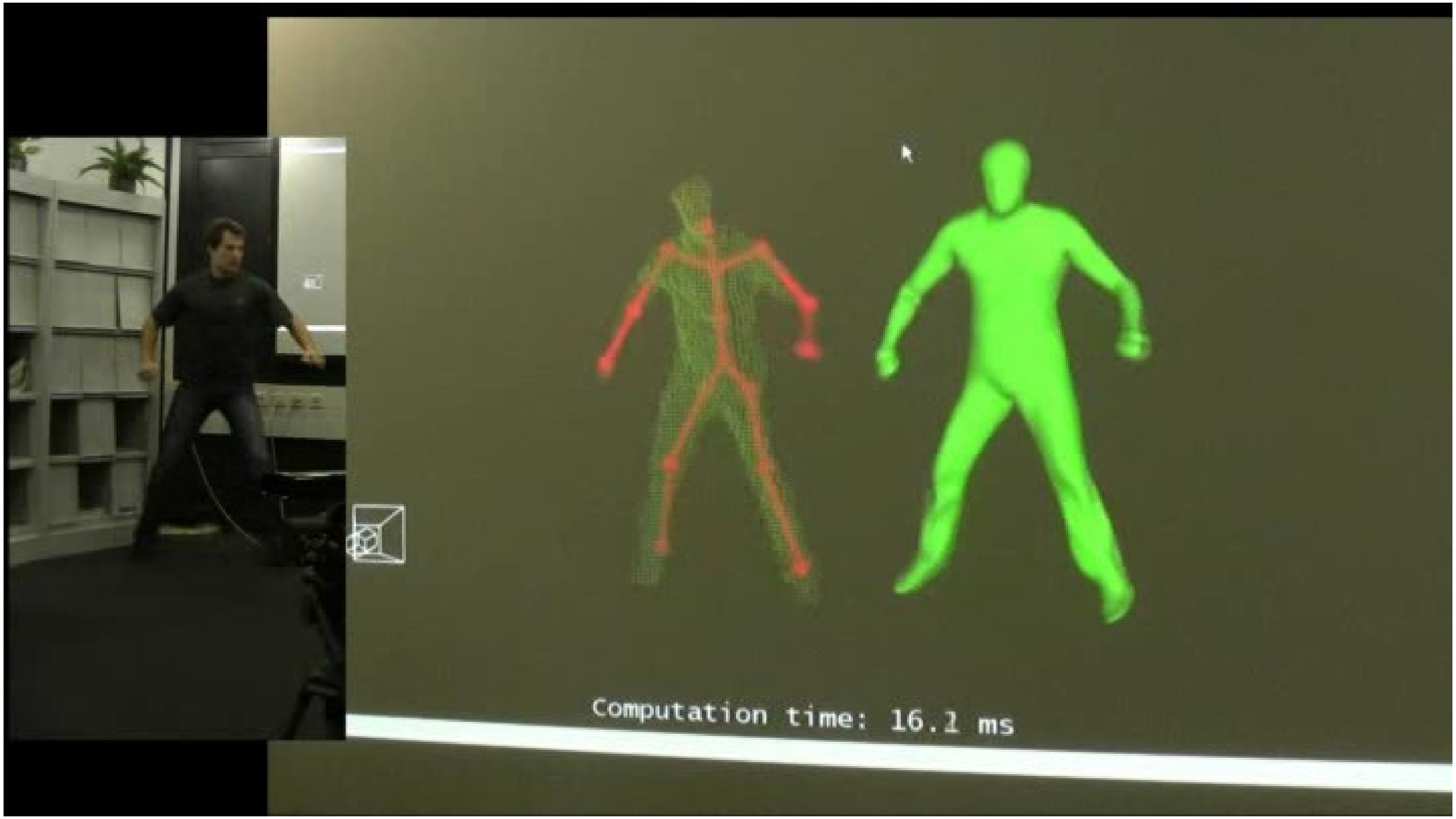
Aligned Video



Filtered Distance Map
(rendered as mesh)

C. Richardt, C. Stoll, N. Dogdgson; H.-P. Seidel, C. Theobalt, *Coherent Spatiotemporal Filtering, Upsampling and Rendering of RGBZ Videos*, in CGF (Proc. EUROGRAPHICS), 2012.

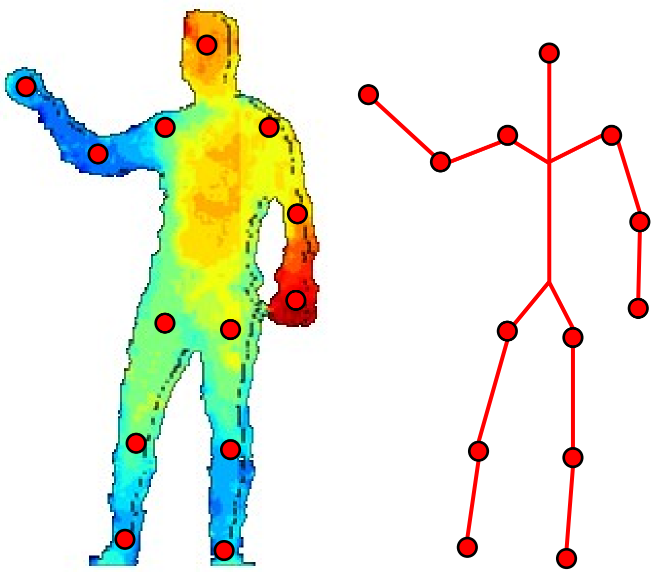
Real-time Skeletal Tracking – Single Depth Camera



A. Baak, M. Müller, G. Bharaj, H.-P. Seidel, C. Theobalt, *A Data-Driven Approach for Real-Time Full Body Pose Reconstruction from a Depth Camera*, Proc. ICCV, 2011.

Real-time Skeleton Tracking from Depth

Discriminative approaches

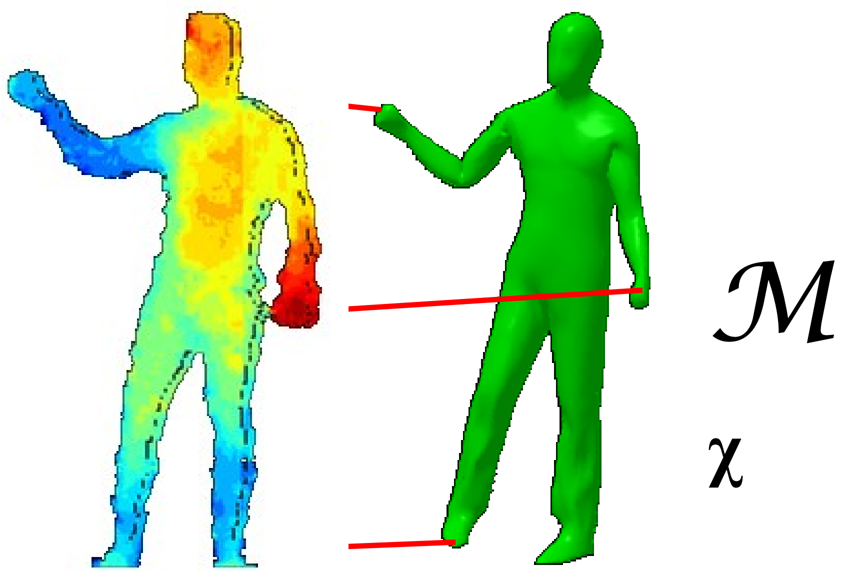


Shotton et al.,*CVPR,* 2011

Taylor et al.,*CVPR*, 2012

....

Generative approaches



Bleiweiss et al.,*SIGGRAPH ASIA Sketches*, 2009

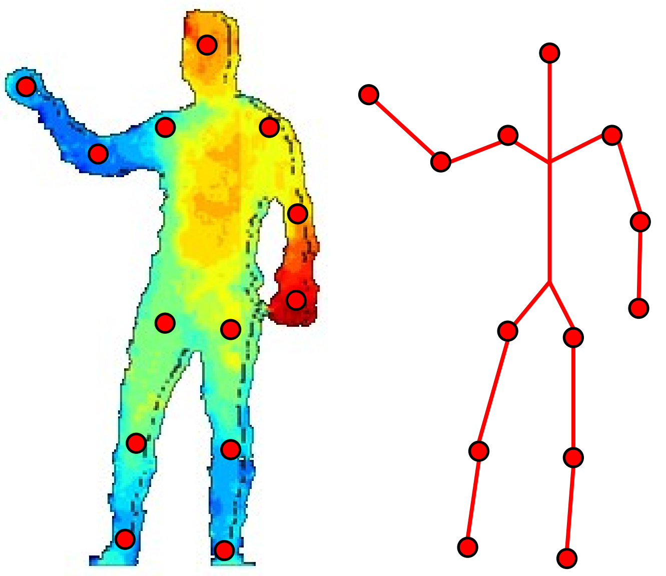
Knoop et al., *Robotics and Autonomous Systems,* 2009

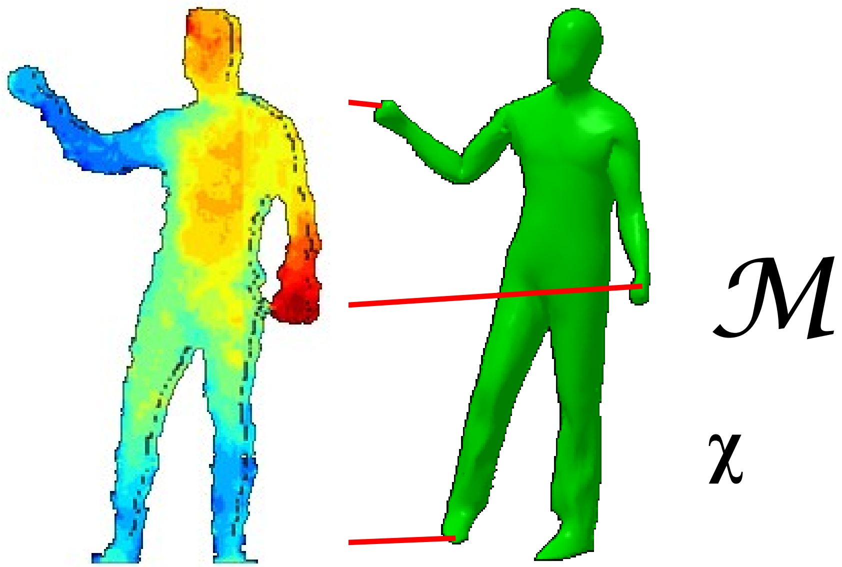
Ganapathi et al. CVPR 2010, ECCV 2012...

Real-time Skeleton Tracking from Depth

Discriminative approaches

Generative approaches



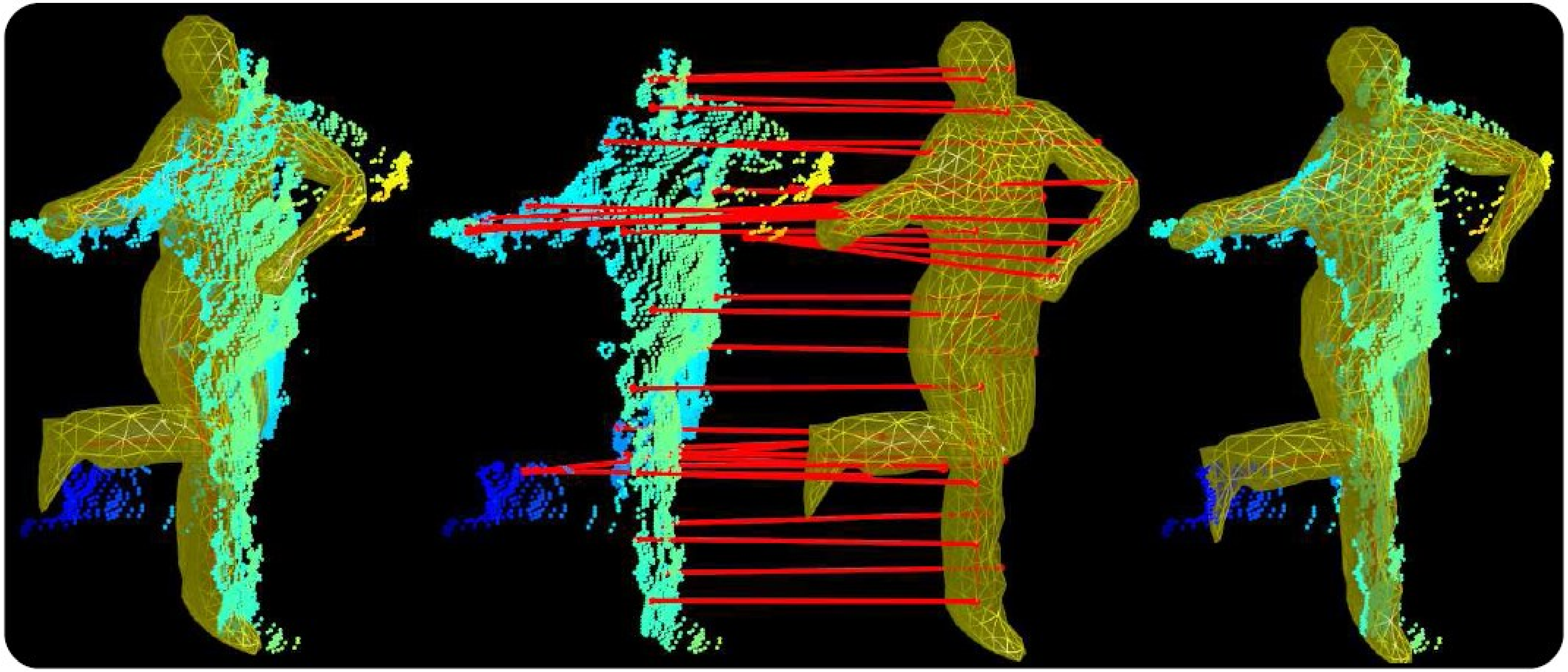


Baak et al., *ICCV* 2011

Helten et al., 3DV 2013

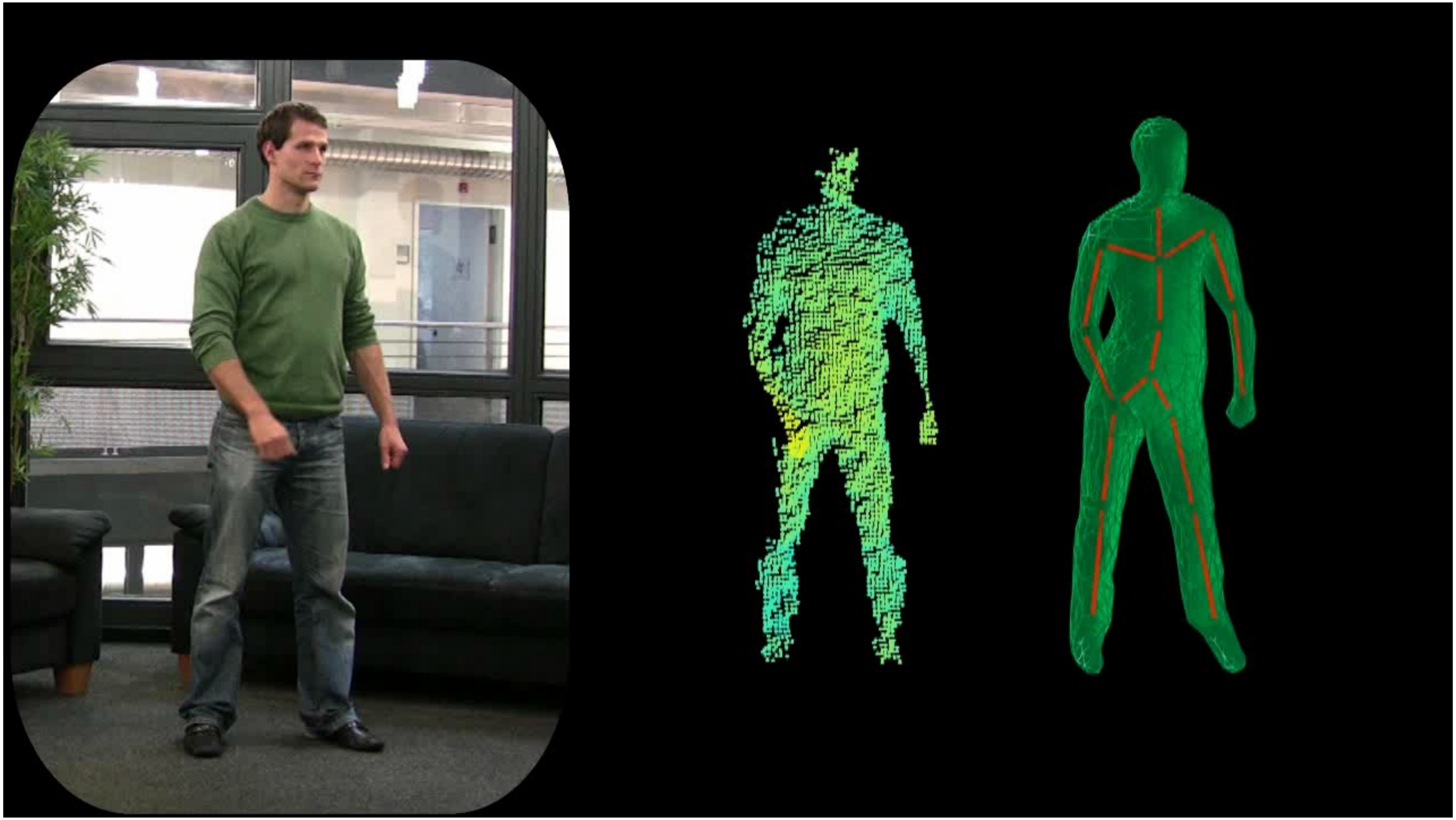
Wei et al., SIGGRAPH Asia 2012

Solver I: Generative Pose Optimization

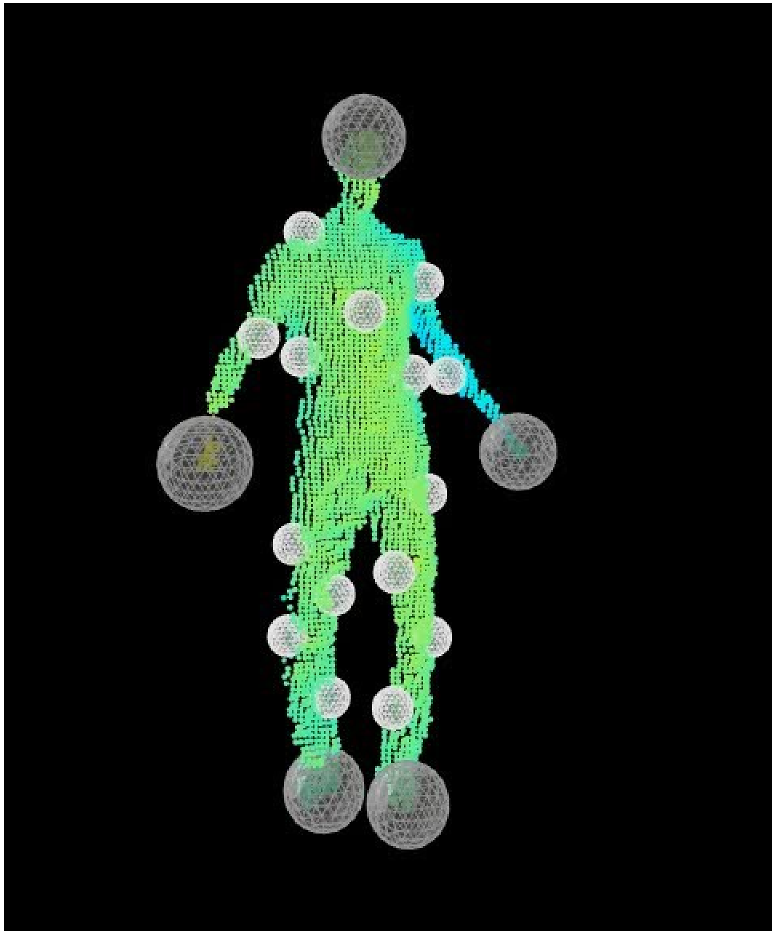


Correspondence hypothesis between model and point cloud →
optimize pose parameters that minimize distance

Solver I: Local Pose Minima

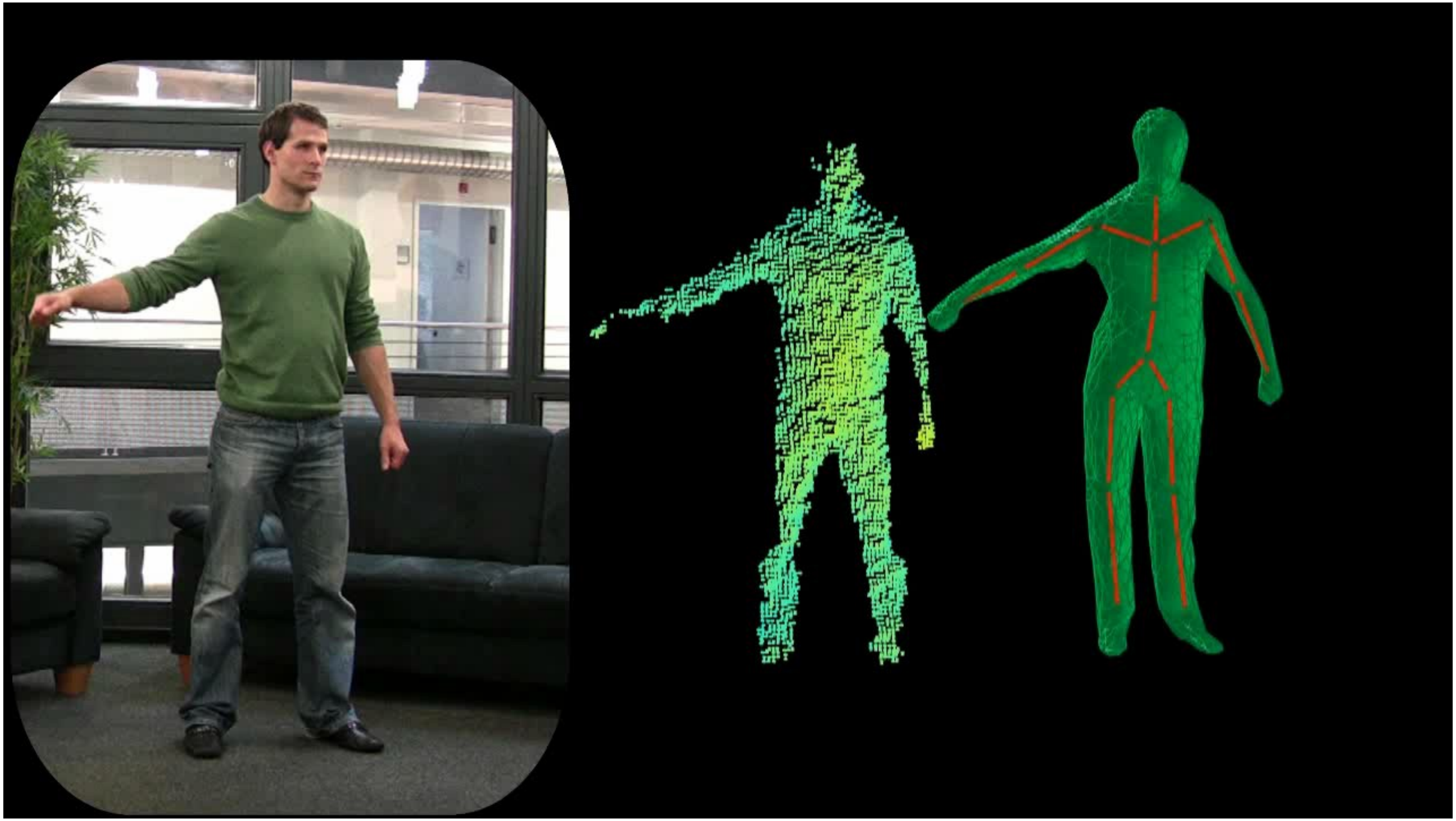


Solver II: Geodesic Extrema



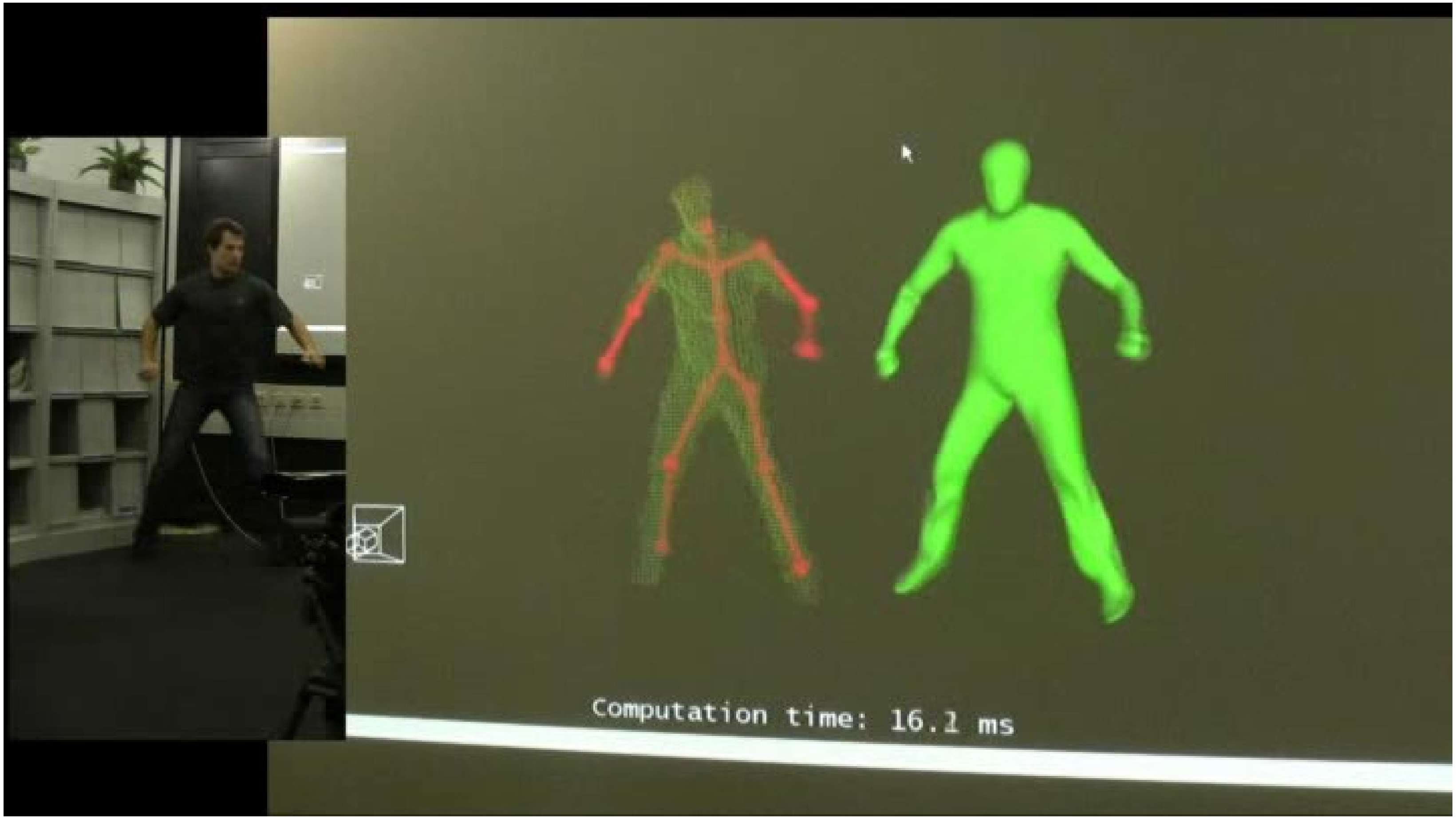
Label 5 end-effectors + temporal smoothness

Solver II:
(Discriminative) Database Lookup



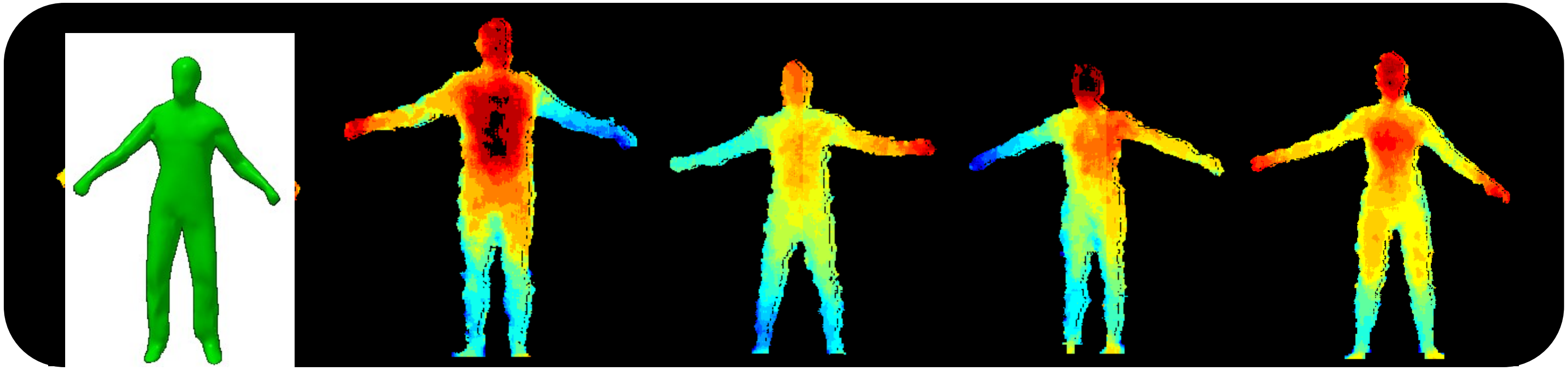
Kd-tree: Find closest pose from 23000 in a database

Fusion: Hypothesis Selection



A. Baak, M. Müller, G. Bharaj, H.-P. Seidel, C. Theobalt, *A Data-Driven Approach for Real-Time Full Body Pose Reconstruction from a Depth Camera*, Proc. ICCV, 2011.

Personalization of Body Model

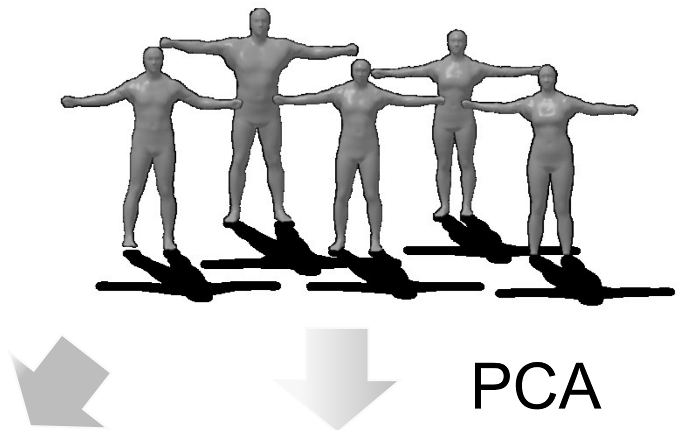


- Fast fitting of parametric human body model
- Personalized Hybrid Depth Tracker

T. Helten, A. Baak, G. Bharaj, M. Müller, H.-P. Seidel, C. Theobalt, *Personalization and Evaluation of a Real-time Depth-based Full Body Tracker* , Proc. of 3DV, 2013.

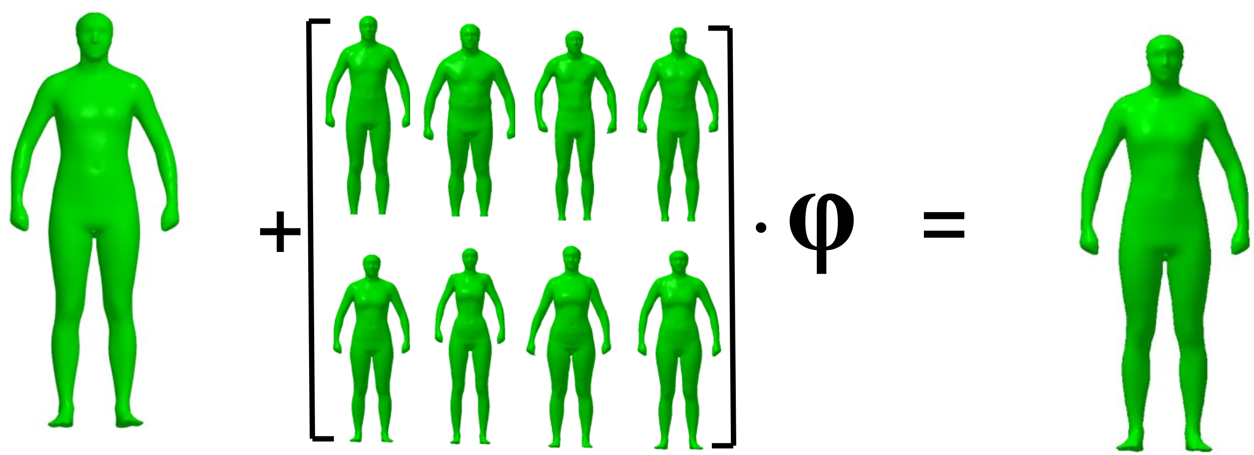
max planck institut informatik

Shape-parameterized Model



Pre-recorded laser scans

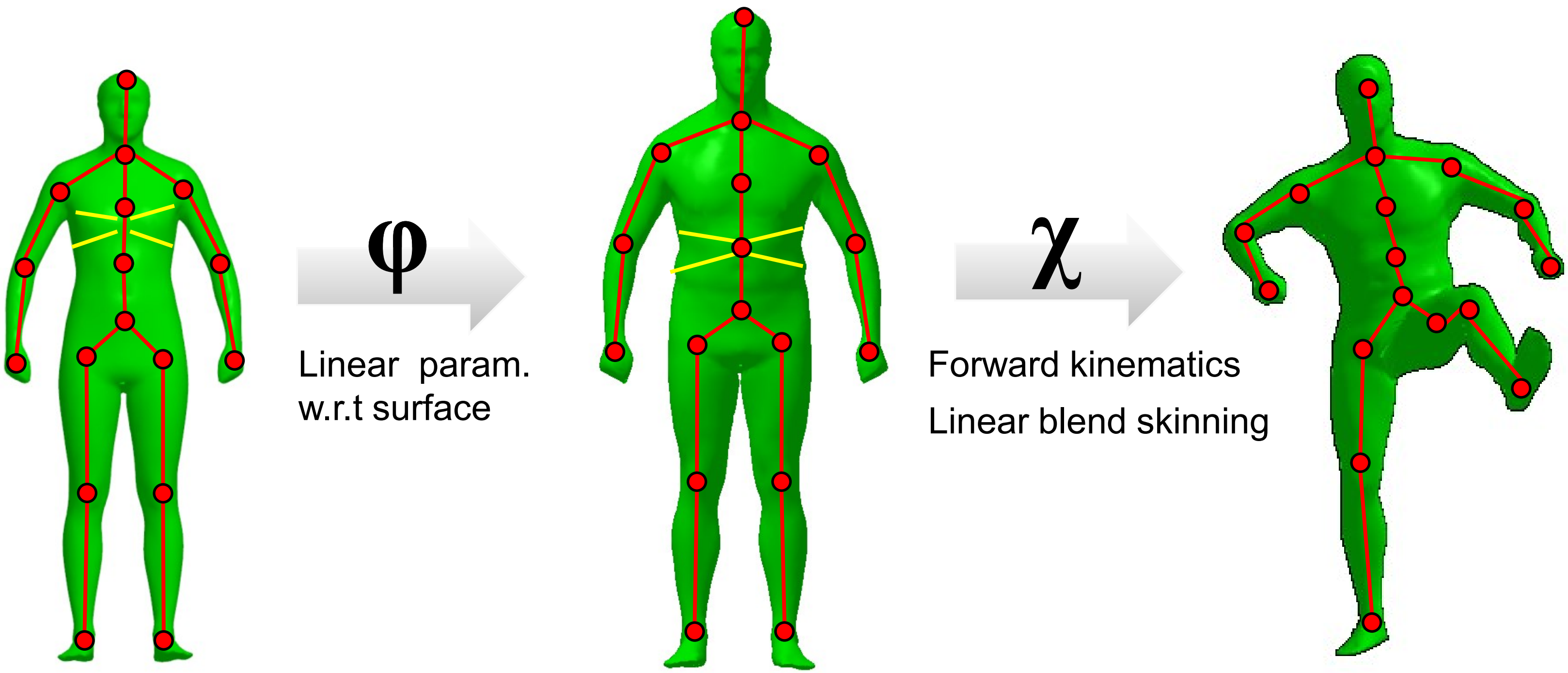
PCA



$$\mathcal{M}_{\chi_0,0} + \Phi \cdot \varphi = \mathcal{M}_{\chi_0,\varphi}$$

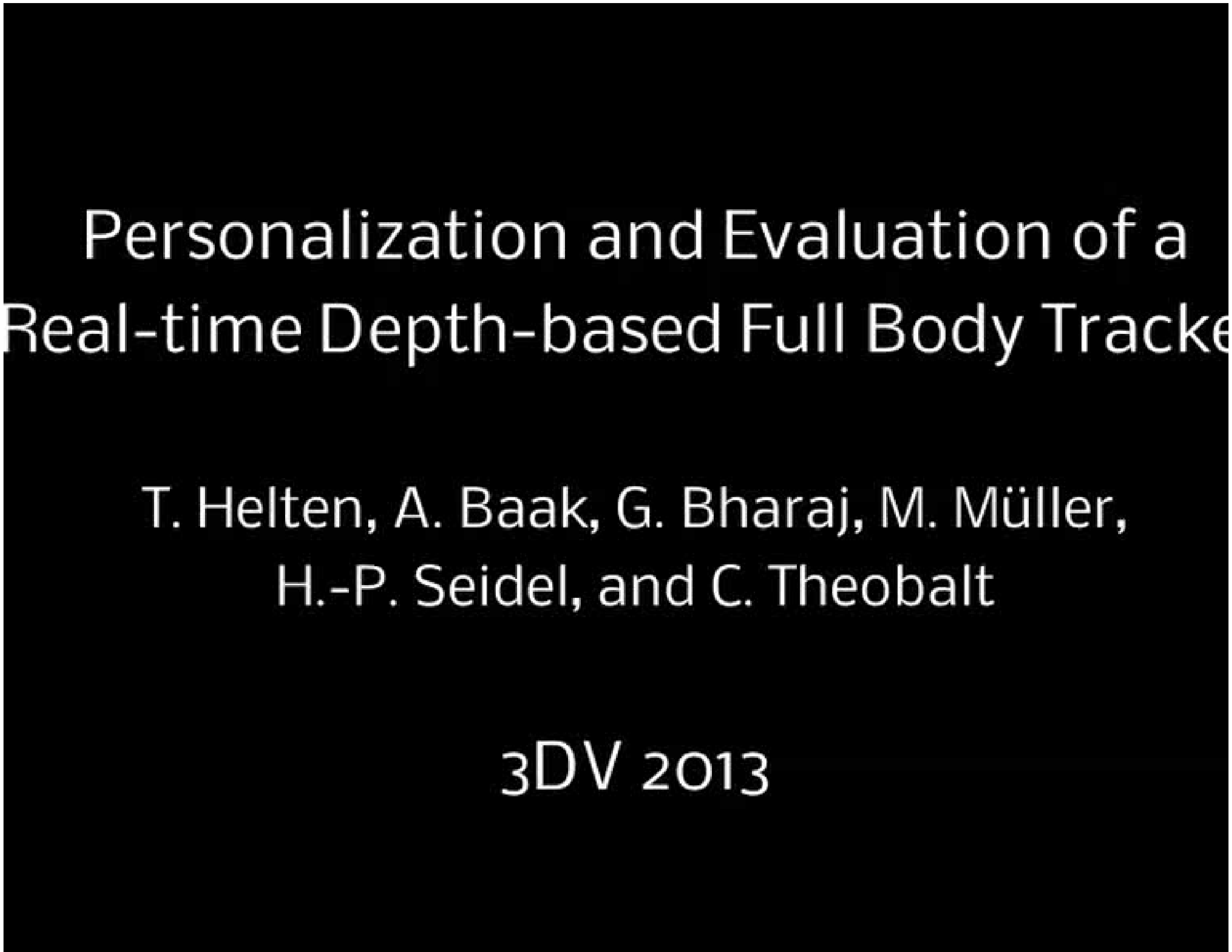
Pose Parameterization

max planck institut
informatik

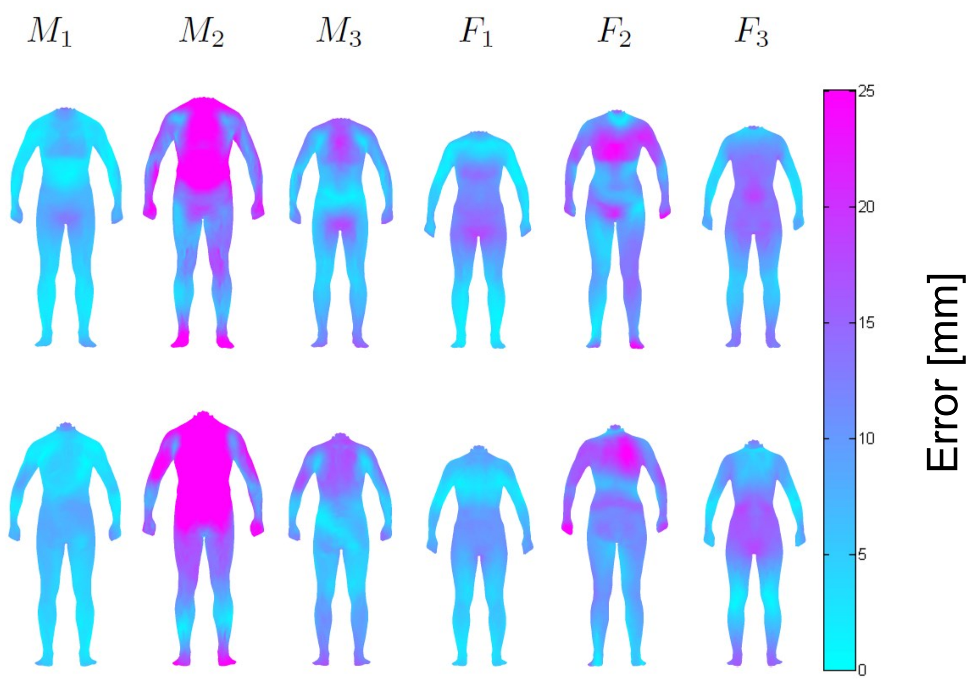


$\mathcal{M}_{\chi^0,0}$ $\mathcal{M}_{\chi^0,\varphi}$ $\mathcal{M}_{\chi,\varphi}$

Shape Estimation

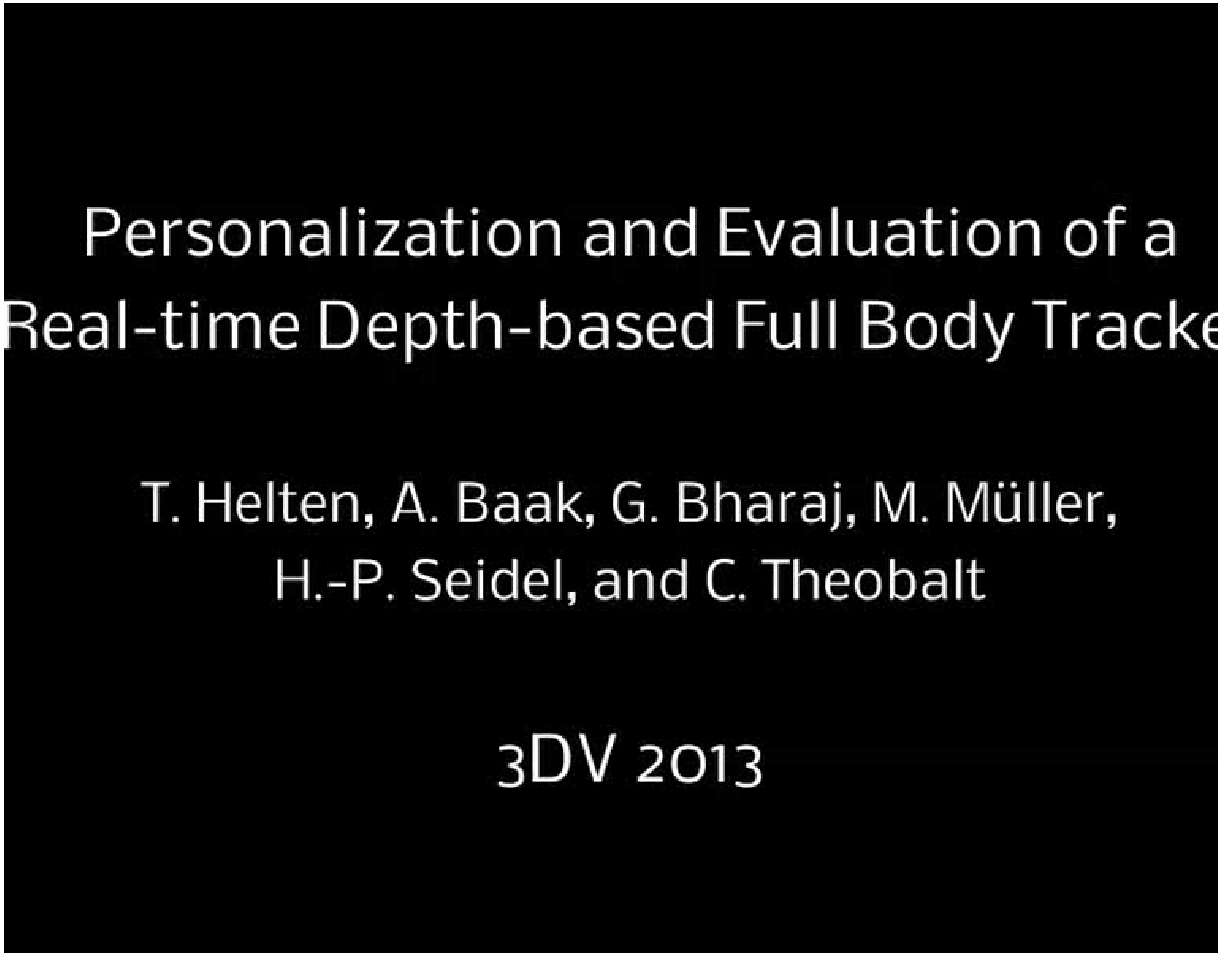


Shape Estimation Results



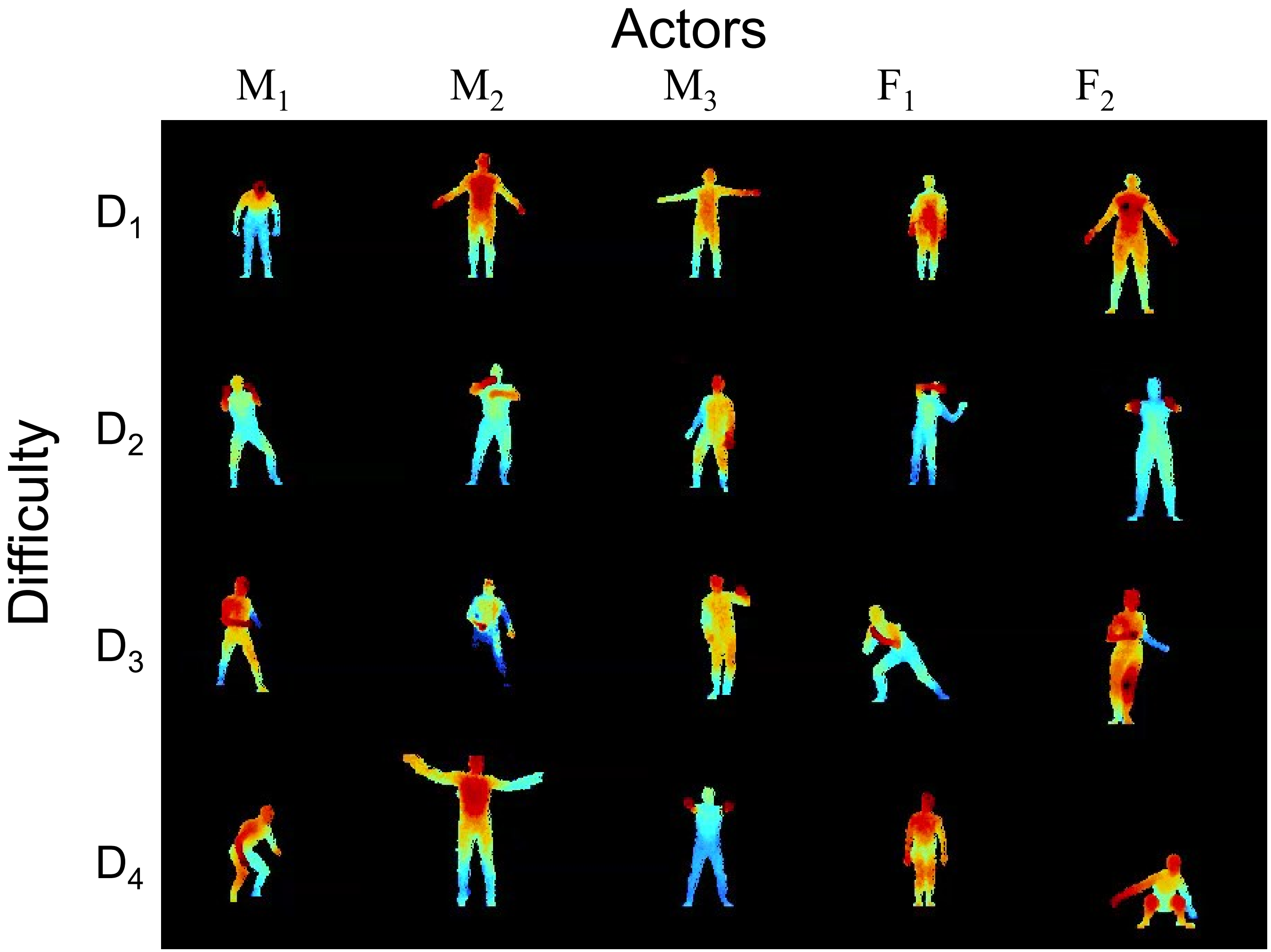
	Weiss et al., ICCV 2011	Our Approach
Avg. error	~1 cm	~1 cm
Used images	4	2
Runtime	~60 minutes	~1 minute

Personalized Depth Tracker



T. Helten, A. Baak, G. Bharaj, M. Müller, H.-P. Seidel, C. Theobalt, *Personalization and Evaluation of a Real-time Depth-based Full Body Tracker* , Proc. of 3DV, 2013.

Evaluation Dataset



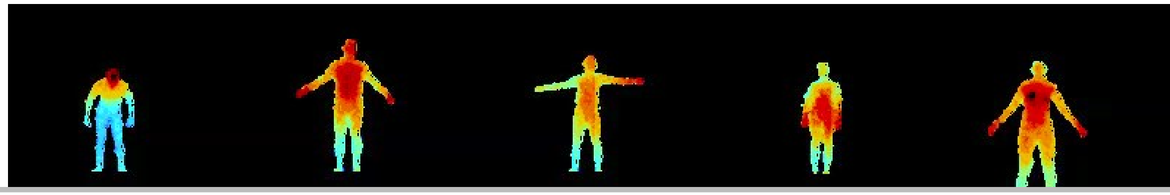
- Calibrated Depth data & marker positions
- Ground-truth marker positions

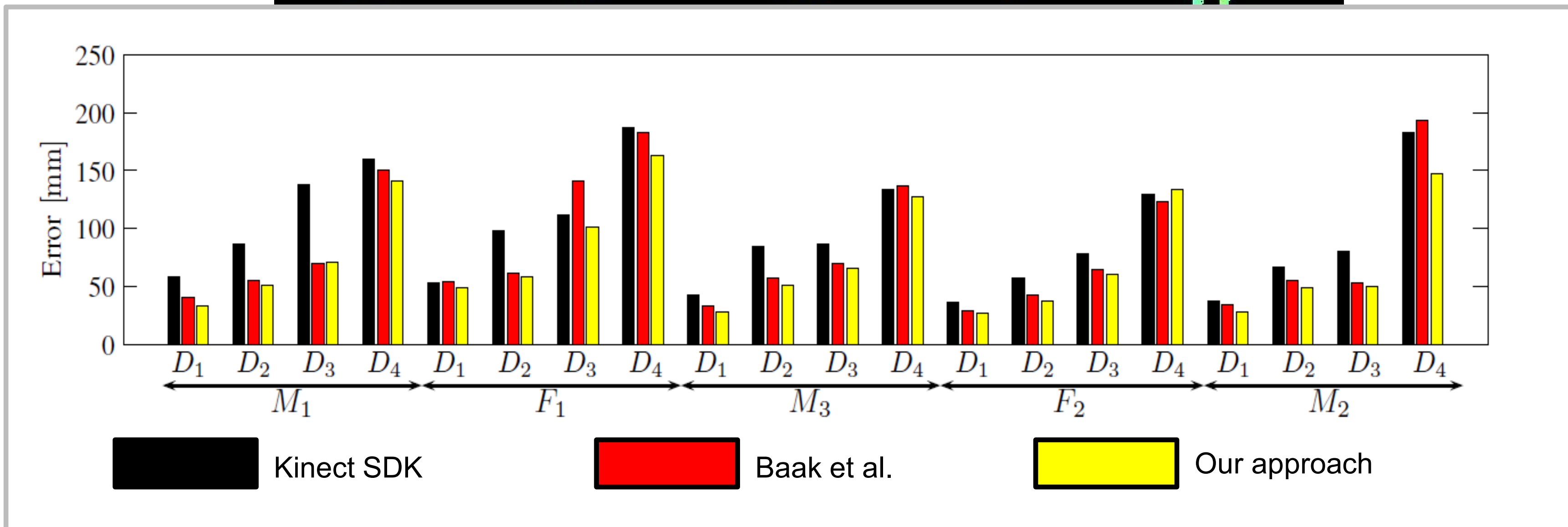
Evaluation Dataset

Actors

M_1 M_2 M_3 F_1 F_2

D_1





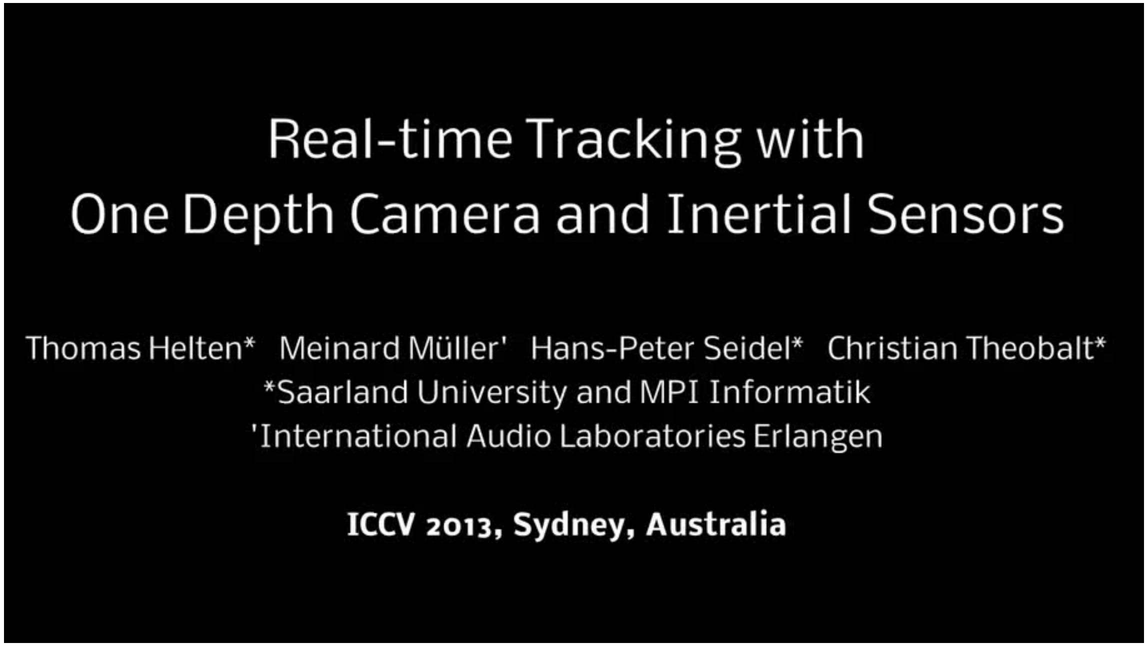
D_4

- Calibrated Depth data & marker positions
- Ground-truth marker positions

Results

T. Helten, M. Müller, and H.-P. Seidel, and C. Theobalt, *Real-time Body Tracking with One Depth Camera and Inertial Sensors*, ICCV 2013.

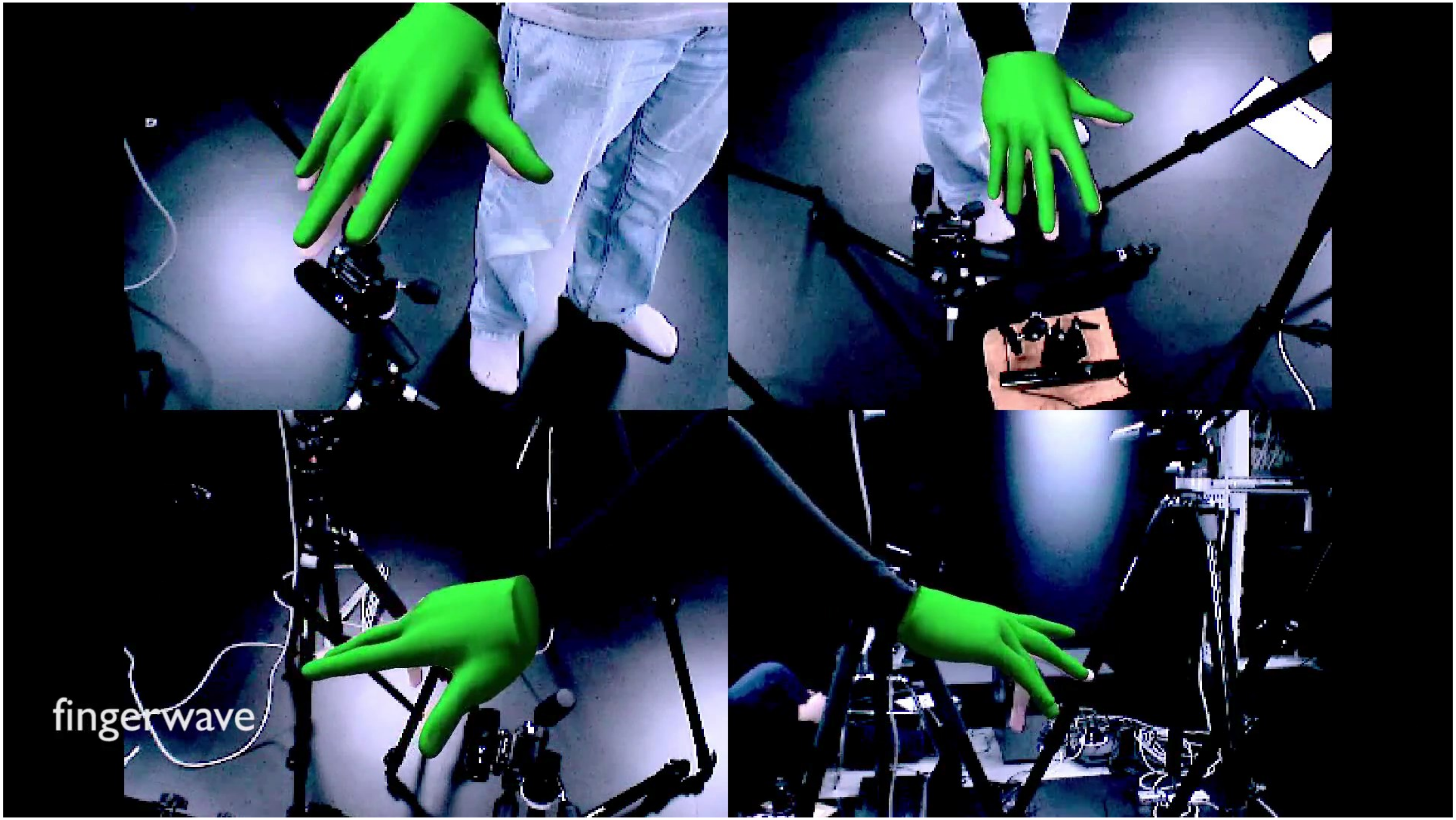
Results – Evaluation Dataset



6 sequences of varying difficulty
- Kinect
- XSENS IMUs
- Ground truth: 38 markers

Real-time Articulated Hnad Tracking

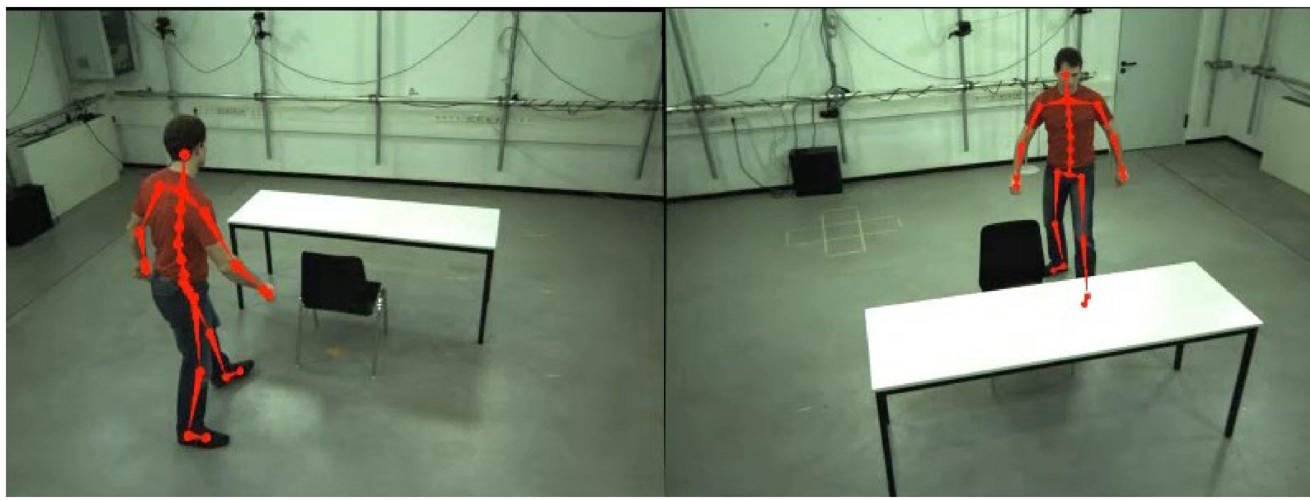
max planck institut
informatik



S. Sridhar, A. Oulasvirta, and C. Theobalt, *Interactive Markerless Articulated Hand Motion Tracking using RGB and Depth Data*, ICCV 2013

Real-time Full-Body Motion Capture

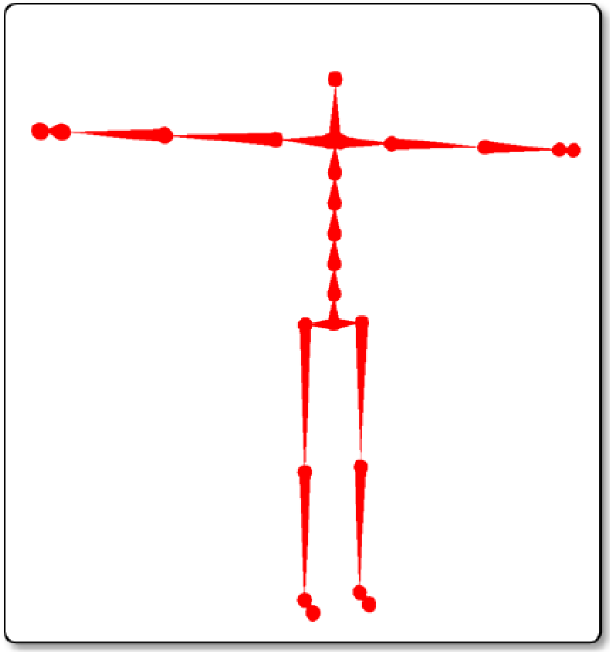
Occlusions,
fast motions,
dynamic backgroud...

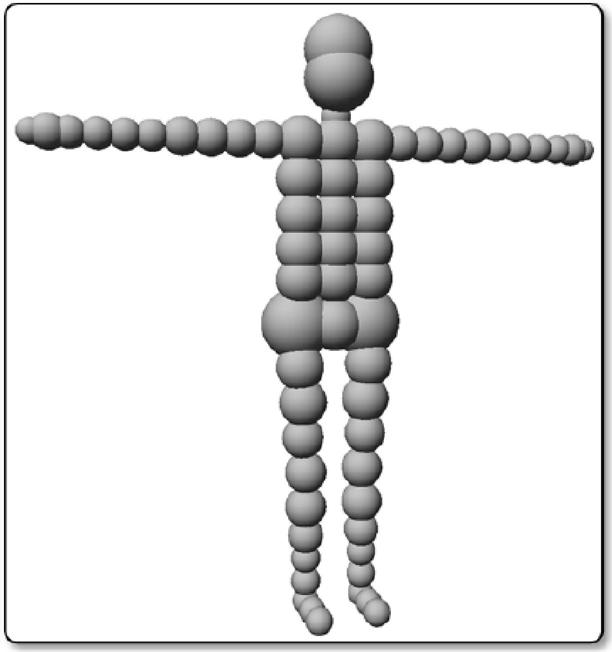


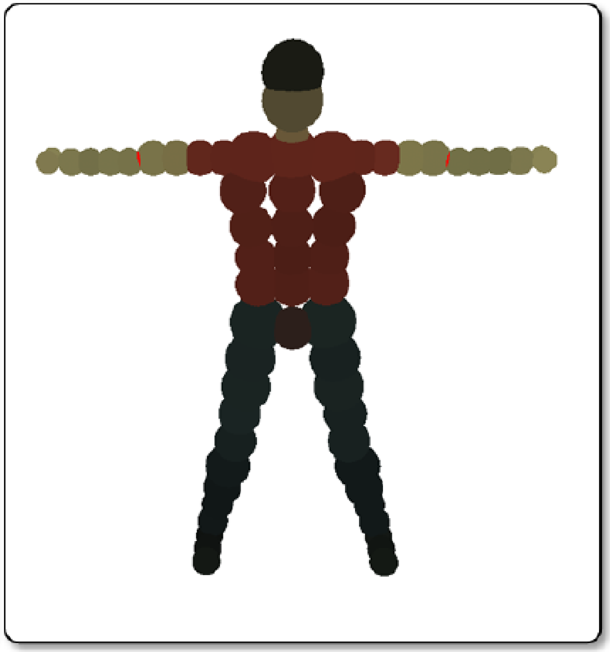
C. Stoll, N. Hasler, J. Gall, H.-P. Seidel, C. Theobalt, *Fast Articulated Motion Tracking using a Sums of Gaussians Body Model*, Proc. ICCV, 2011.

3D Sum of Gaussians Body Model

- Skeleton + Gaussians with color model





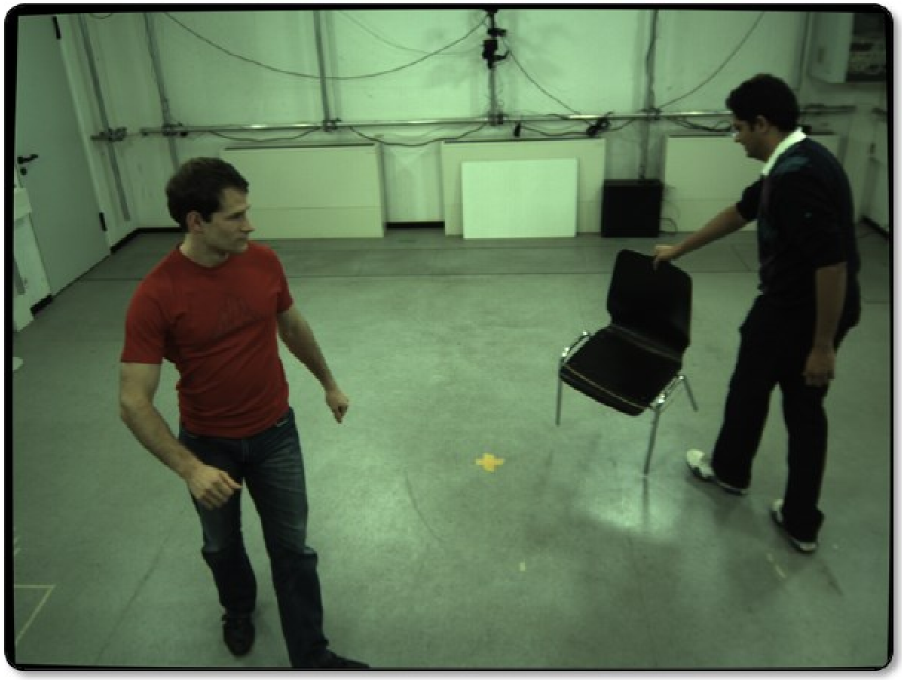


58 DoFs

63 Gaussians

2D Sum of Gaussians Body Model

- Each video frame converted to 2D SoG

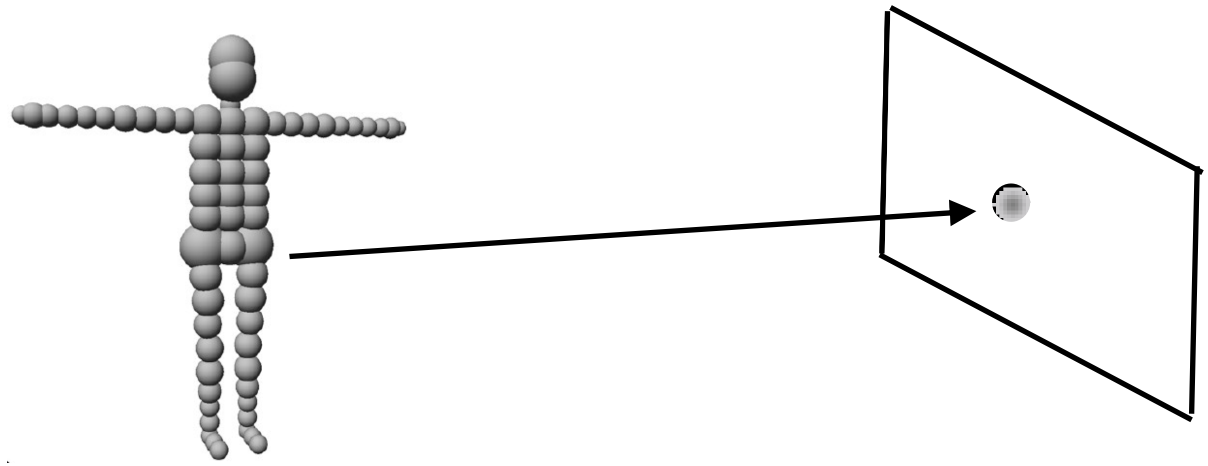


Video frame

Quad tree –
each region one 2D Gaussian

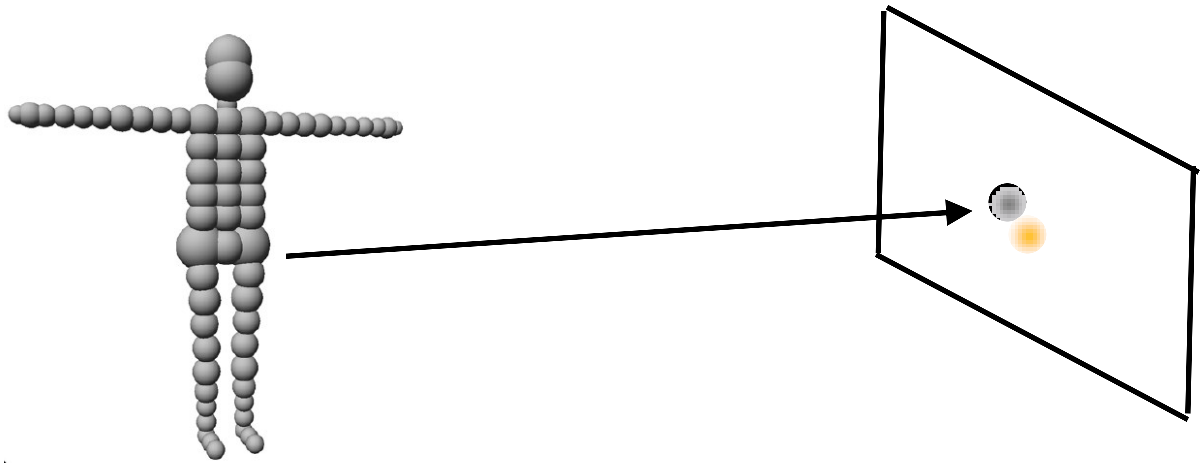
3D-2D Similarity Function

- Projection of 3D Gaussians into cameras



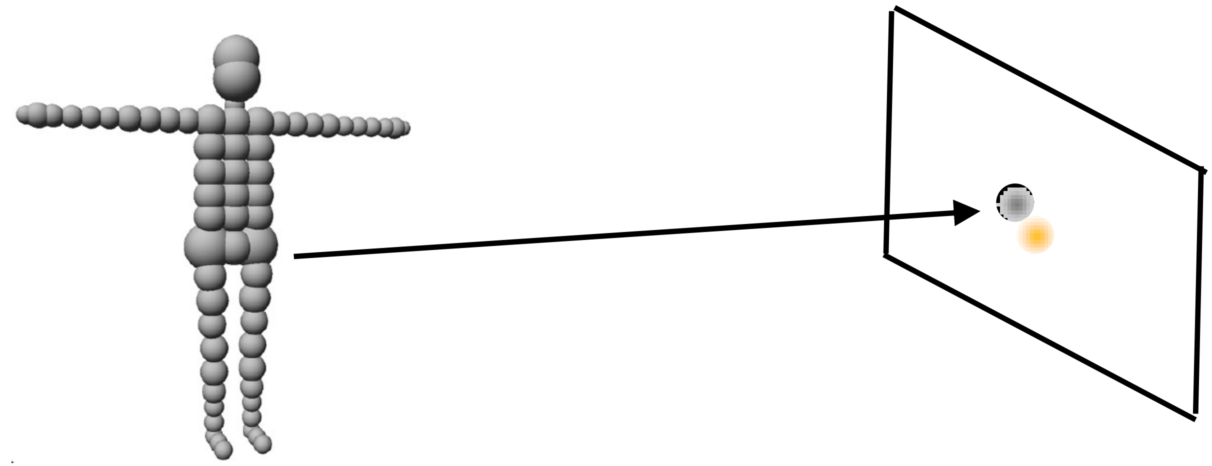
3D-2D Similarity Function

- Projection of 3D Gaussians into cameras



3D-2D Similarity Function

- Projection of 3D Gaussians into cameras

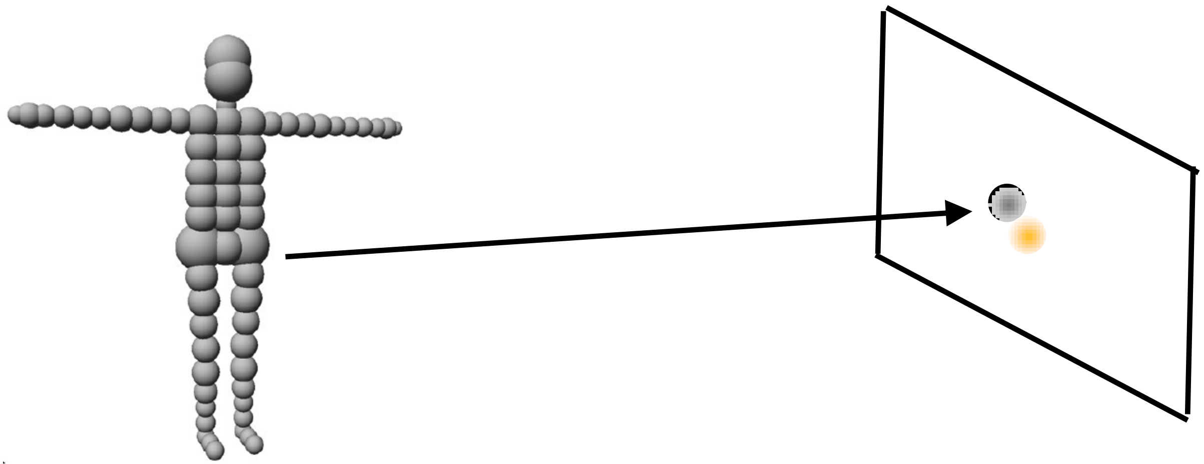


- 2D-2D SoG similarity in one camera view

$$E_{ij} = \underbrace{d(\mathbf{c}_i, \mathbf{c}_j)}_{\text{Color similarity}} \int_{\Omega} \underbrace{\mathcal{B}_i(x)}_{\text{Projected model Gaussian}} \underbrace{\mathcal{B}_j(x)}_{\text{Image Gaussian}} \mathrm{d}\mathbf{x}$$

3D-2D Similarity Function

- Projection of 3D Gaussians into cameras

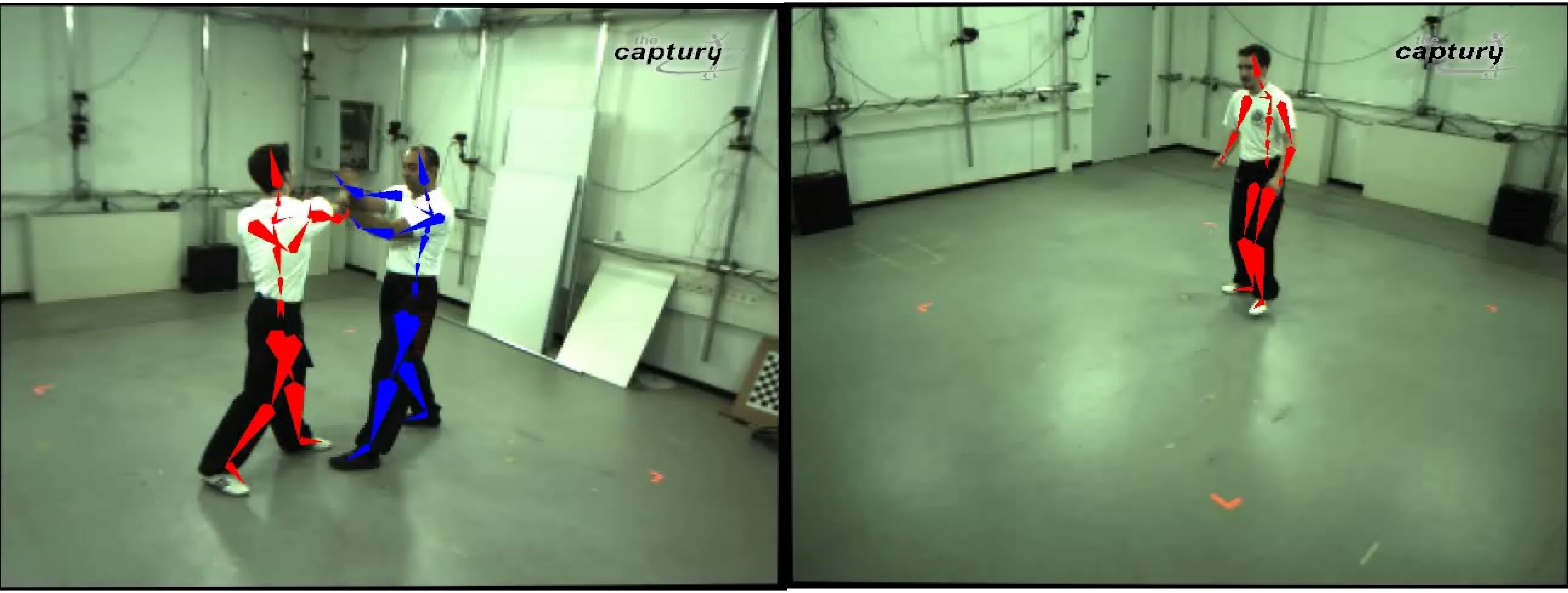


- 2D-2D SoG similarity – closed form solution

$$E_{ij} = d(\mathbf{c}_i, \mathbf{c}_j) \int_{\Omega} \mathcal{B}_i(x) \mathcal{B}_j(x) \mathrm{d}\mathbf{x}$$

$$= d(\mathbf{c}_i, \mathbf{c}_j) 2\pi \frac{{\sigma_i}^2 {\sigma_j}^2}{{\sigma_i}^2 + {\sigma_j}^2} \exp\left(- \frac{\| \mu_i - \mu_j \|^2}{{\sigma_i}^2 + {\sigma_j}^2} \right)$$

More Results



Min nr. of cameras: 5
Pose estimation > 30 fps

Space-time Reconstruction

8 GoPros, 30 fps, not frame synchronized

[Elhayek et al., CVPR 2012]

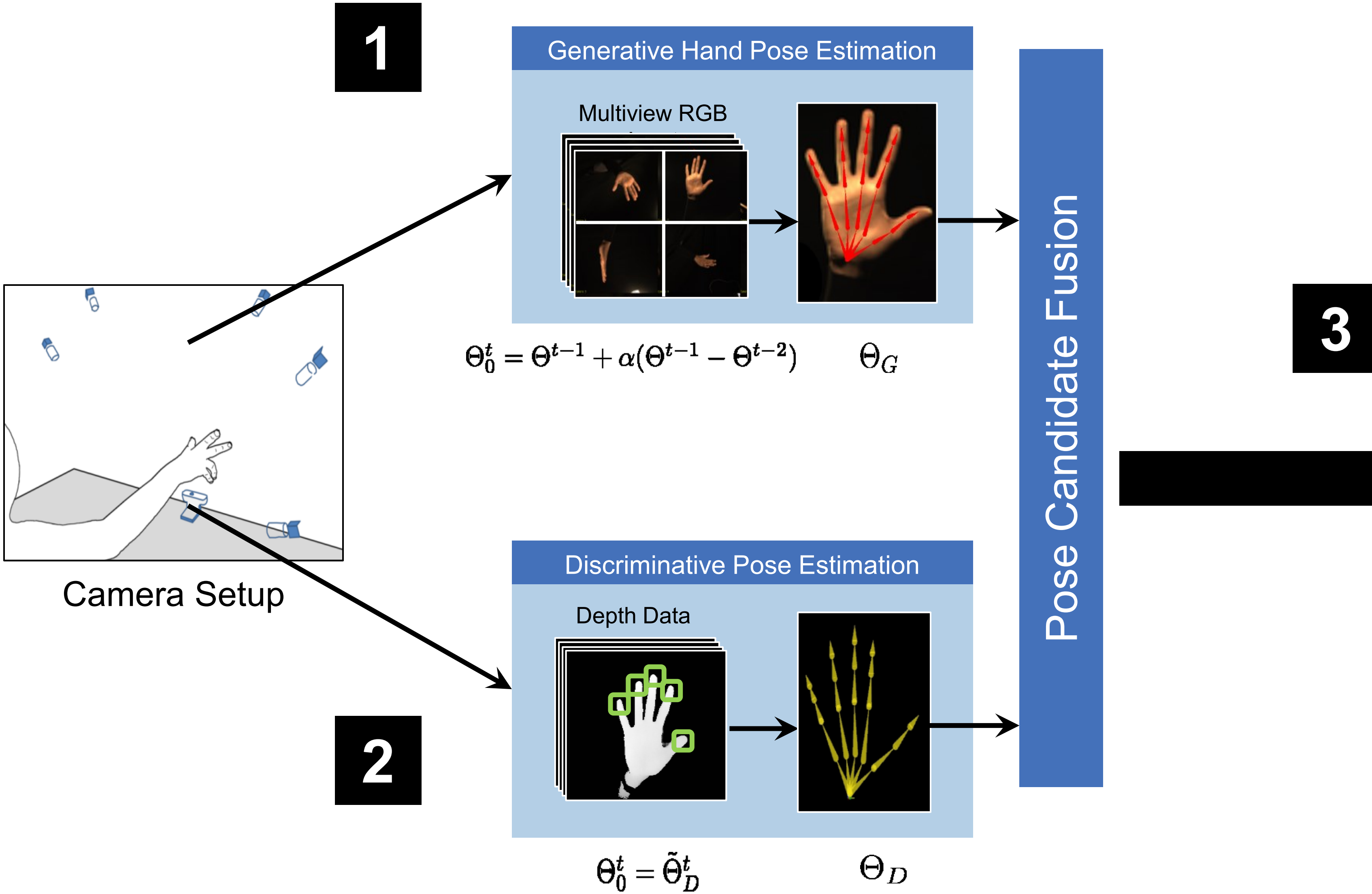
www.thecaptury.com

Real-time Hand-Tracking: Setup

5 video cameras

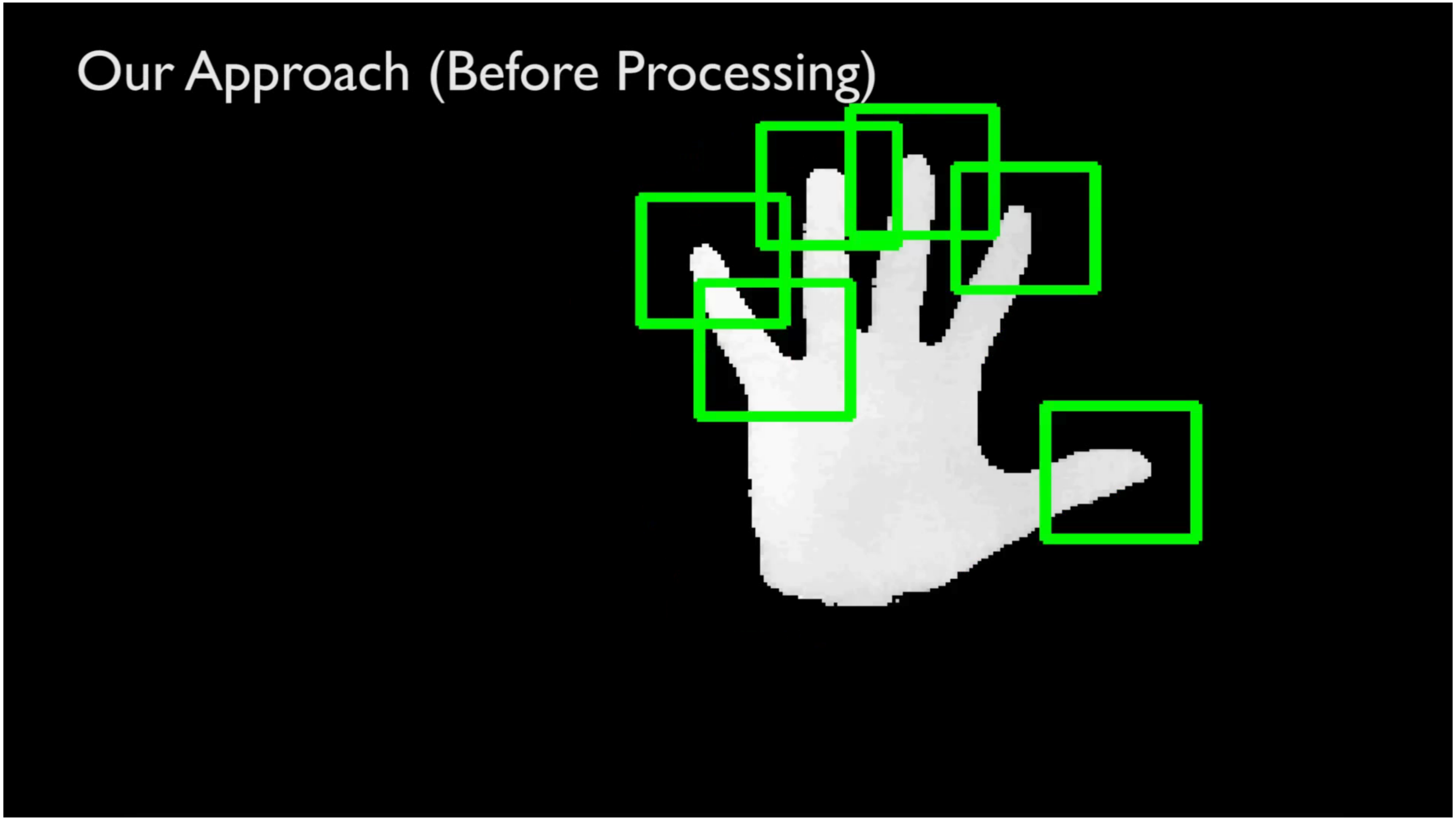
1 Creative TOF Camera

Fusion of two Tracking Strategies

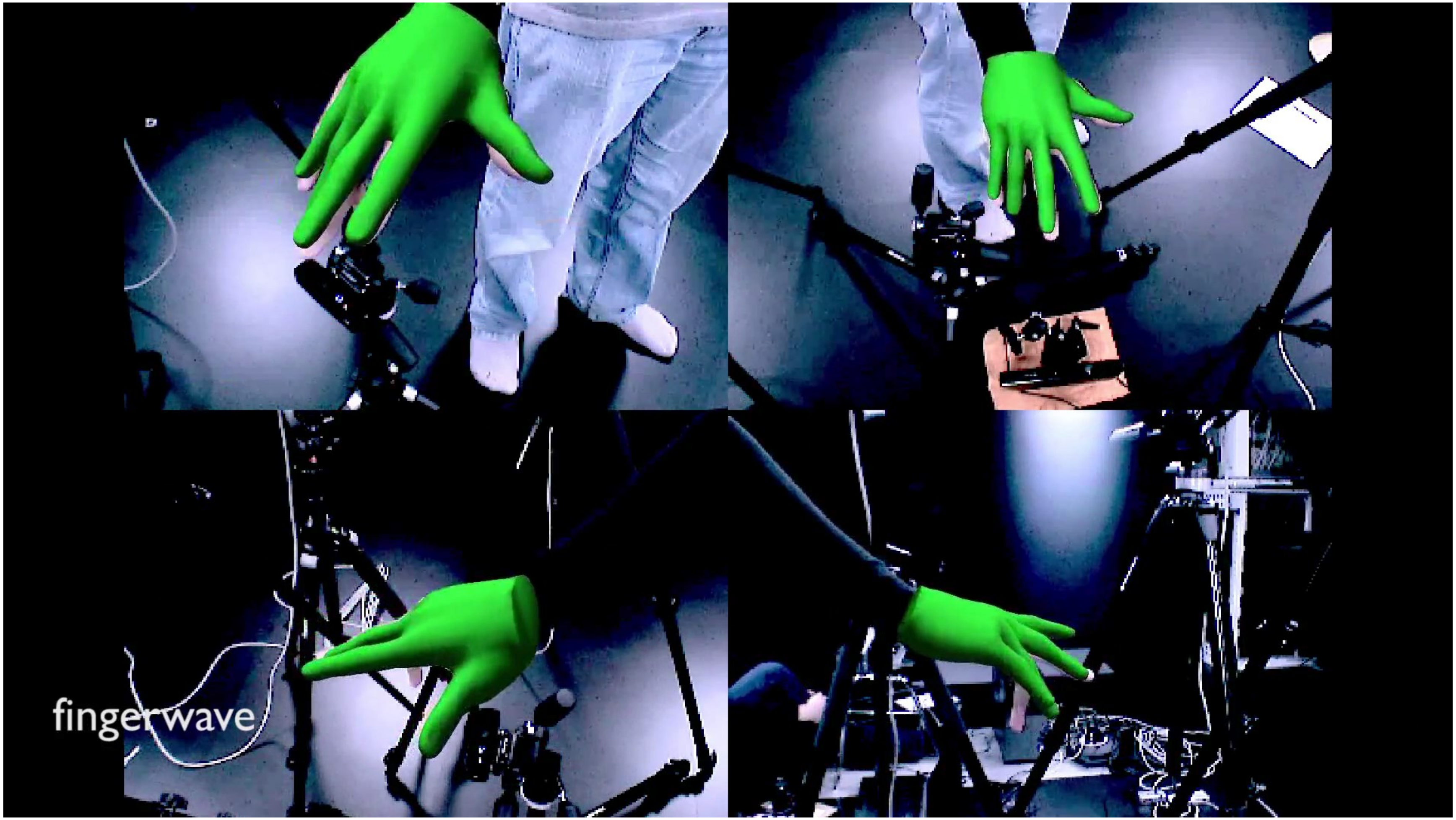


[redacted]

Discriminative Pose Estimation

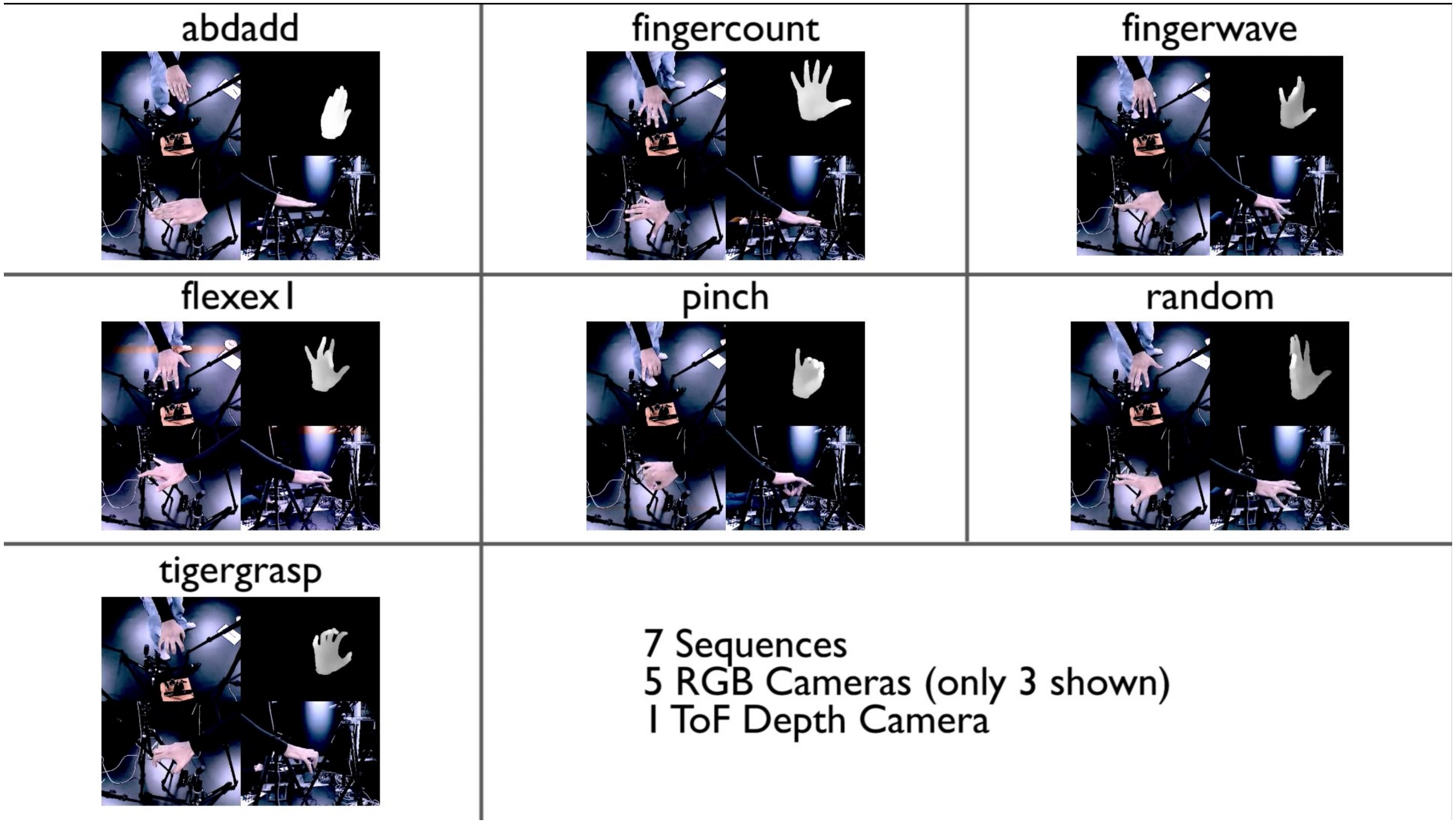


Real-time Hand Tracking



> 30 fps

Evalutation Dataset: Dexter 1



- Accuracy on annotated finger tip position: ~ 13 mm

Thank You !

gvv.mpi-inf.mpg.de

www.thecaptury.com

Thanks to:

ERC Starting Grant
CapReal