

ICC Real-time Hand Tracking under Occlusion from an Egocentric RGB-D Sensor



Franziska Mueller^{1,2} Dushyant Mehta^{1,2} Oleksandr Sotnychenko¹

Srinath Sridhar¹ Dan Casas³ Christian Theobalt¹

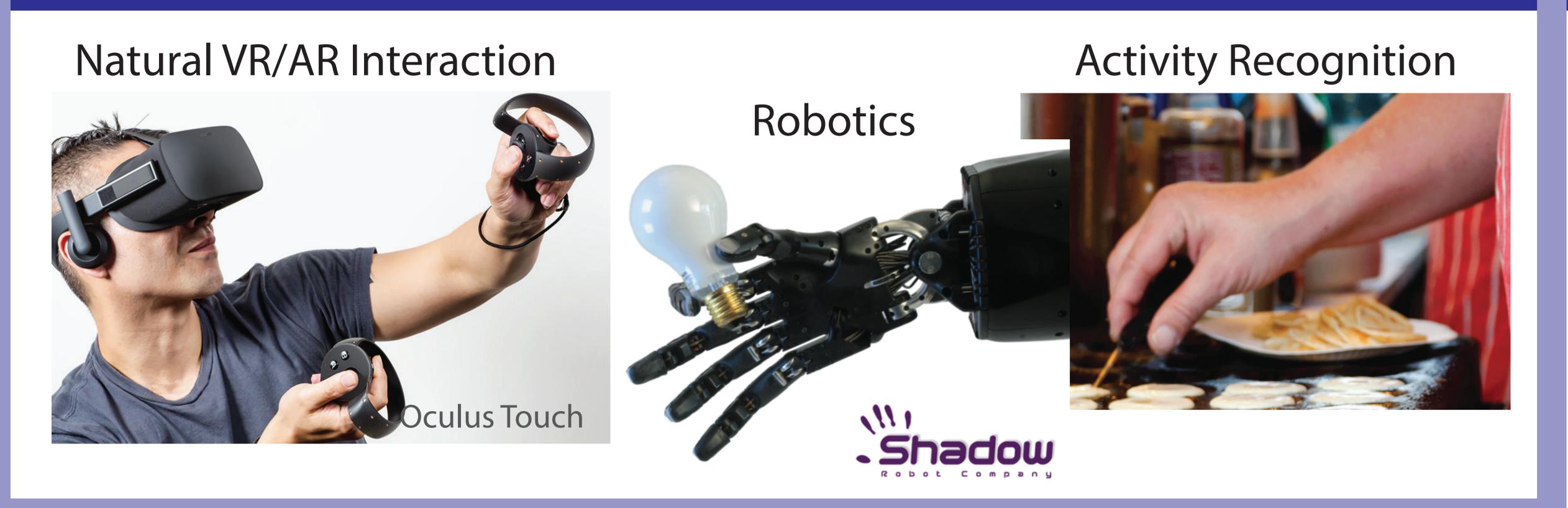
SAARLAND SAARLAND WINIVERSITY WINIVERS







Motivation



Challenges



- Strong **self-occlusions** especially due to **egocentric** camera view
- Extreme occlusions by objects
- Complex and fast motions
- Segmentation of hand from object and cluttered background
- High dimensional problem
- Runtime constraint

Contributions

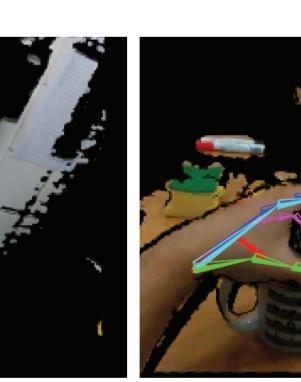
Real-time hand tracking under occlusion in cluttered environments



Single RGB-D camera

@ 50 FPS



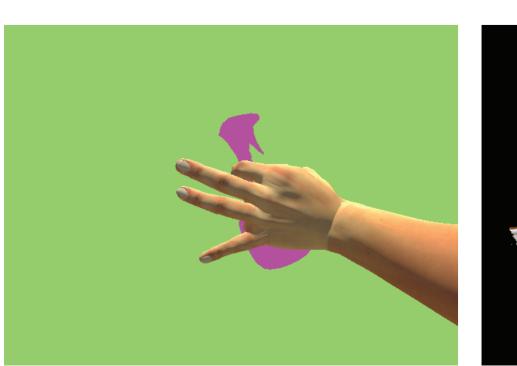


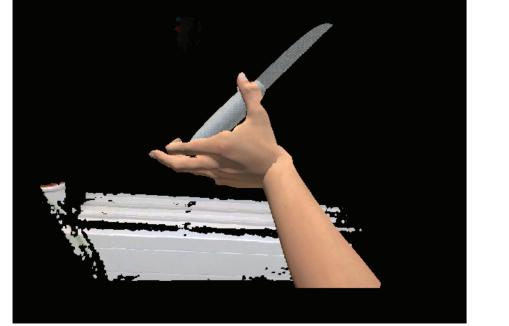
Supports hands of different shapes and colors in interaction with arbitrary objects

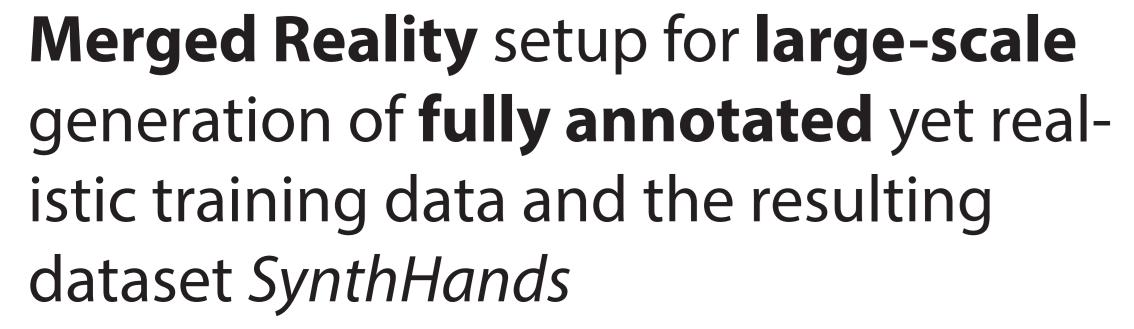
TECHNICAL

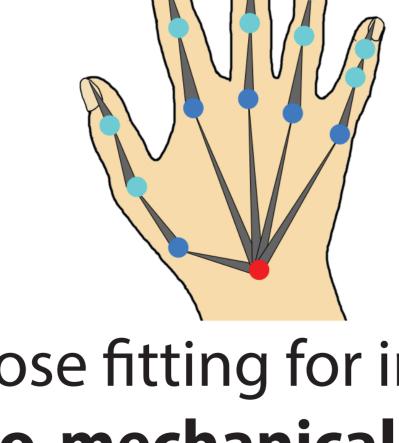


Two-step CNN framework for accurate regression of 2D and 3D joint locations even under occlusion

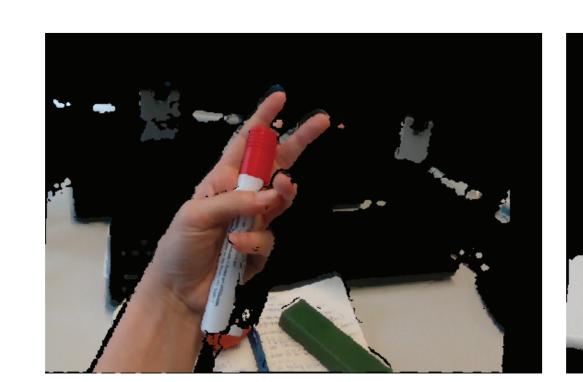








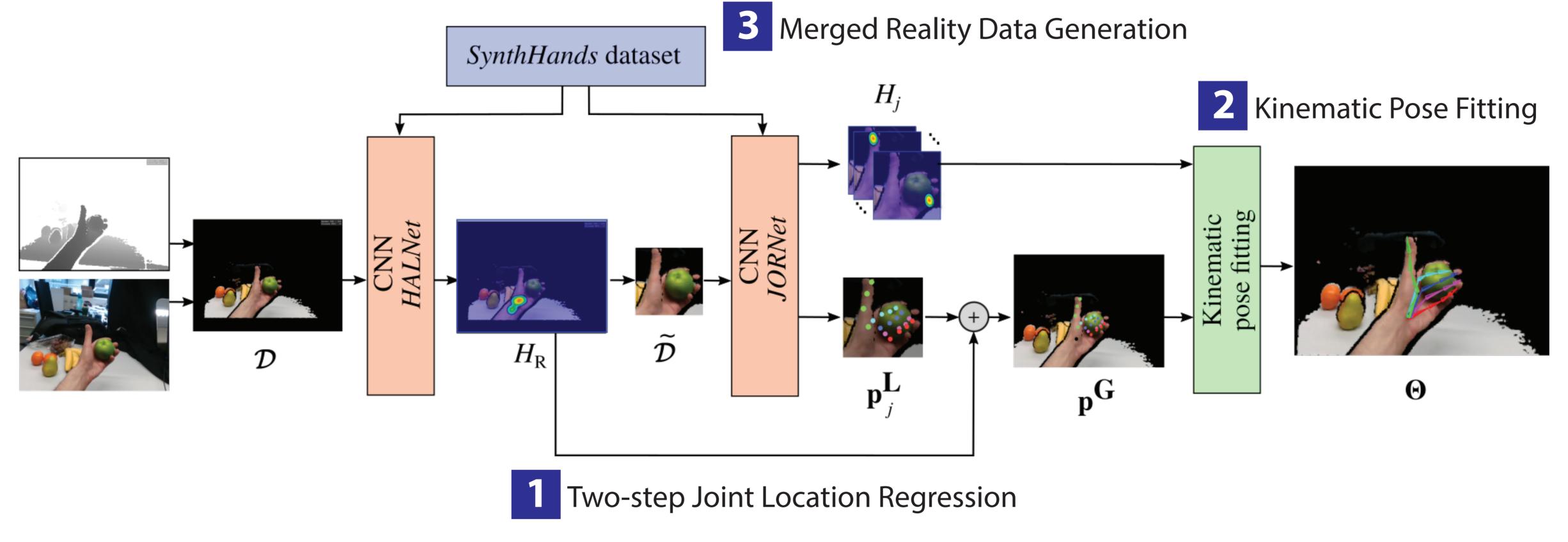
Kinematic pose fitting for improved temporal and bio-mechanical plausibility





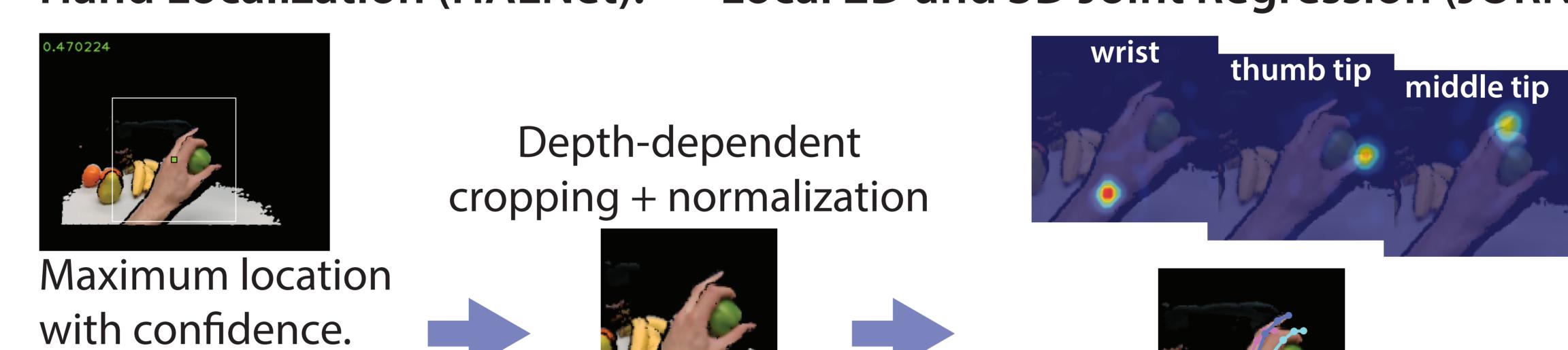
Challenging new real benchmark Ego-Dexter with annotated 3D fingertip positions featuring a variety of objects, clutter, and users

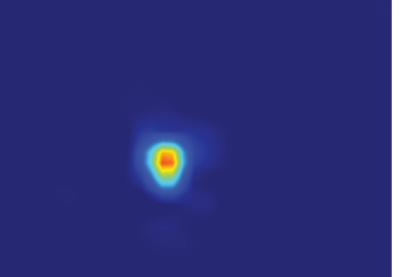
Overview



. Two-step Joint Location Regression

Local 2D and 3D Joint Regression (JORNet): Hand Localization (HALNet): L

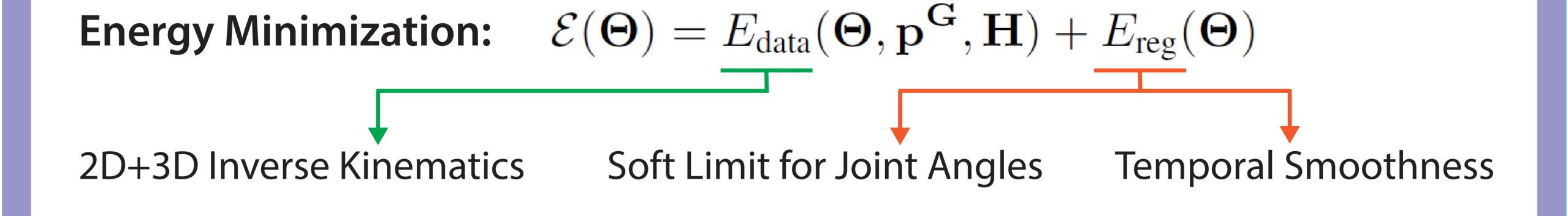




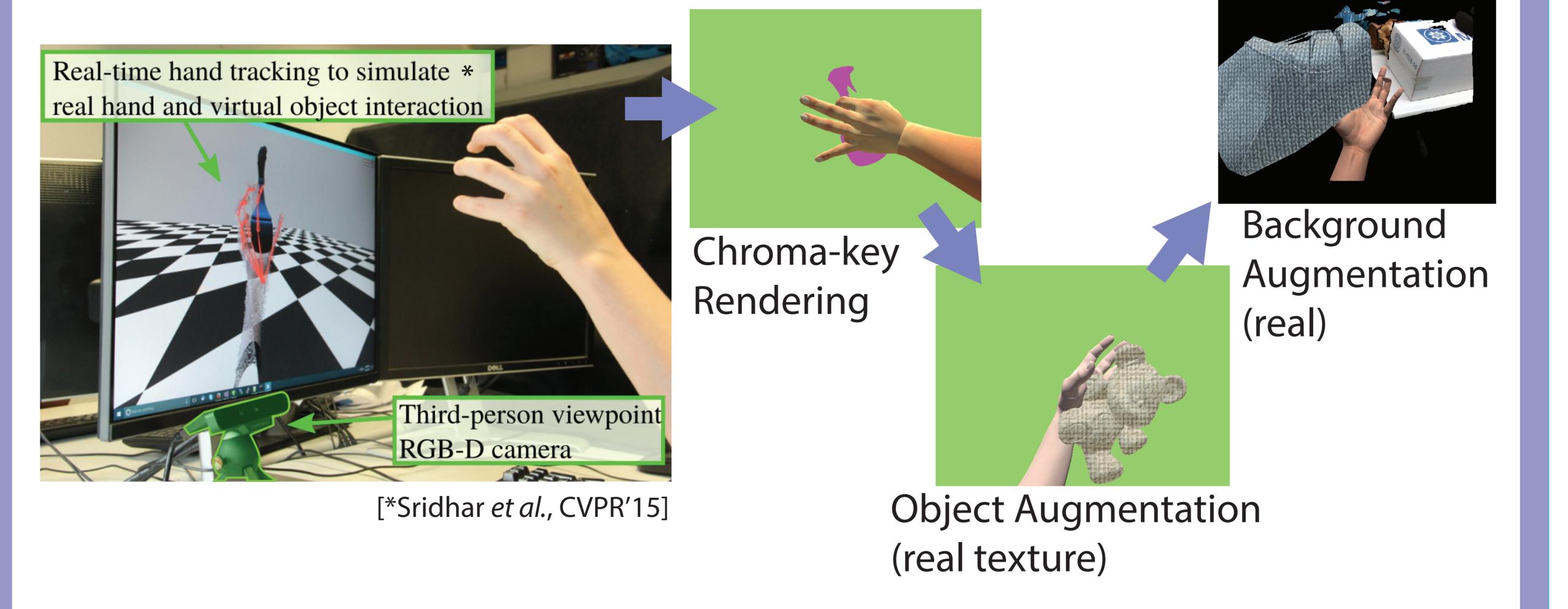


2D heatmaps and 3D positions for all joints. Global 3D locations are obtained from center-relative 3D predictions and global center position.

2. Kinematic Pose Fitting



3. Merged Reality Data Generation

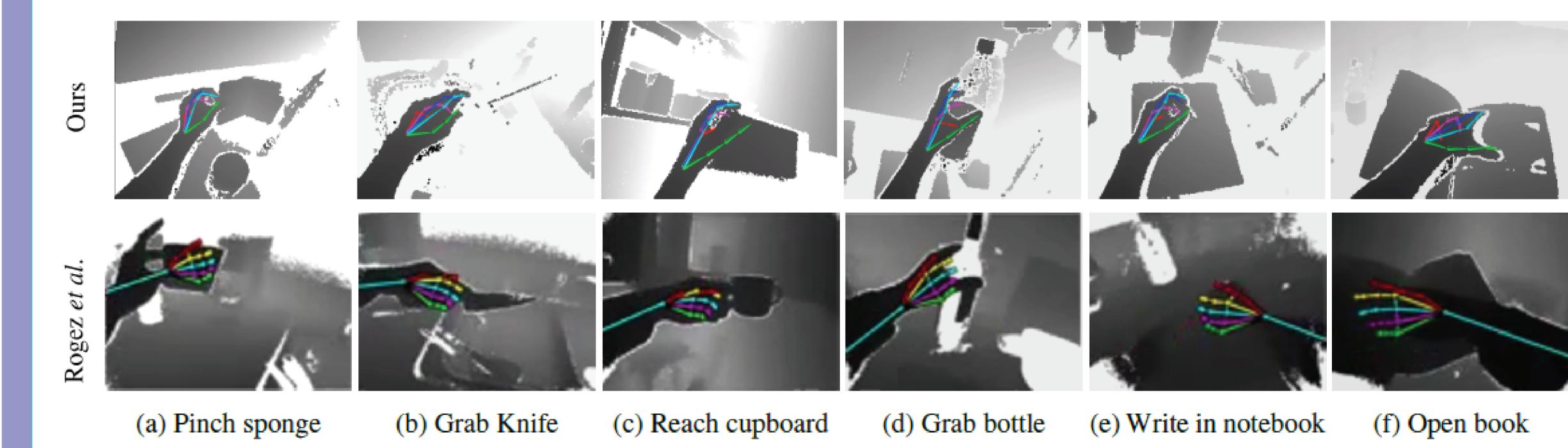


SynthHands and EgoDexter datasets available! handtracker.mpi-inf.mpg.de

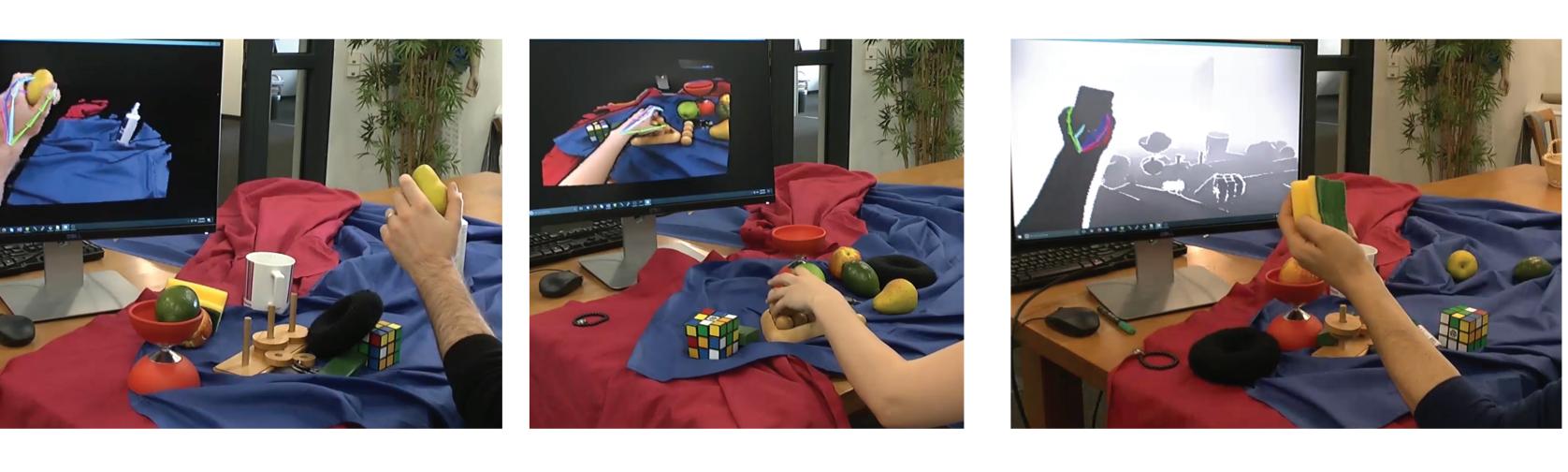
For access to a wide range of human shape and performance capture datasets, please visit: Graphics gvvperfcapeva.mpi-inf.mpg.de

Results

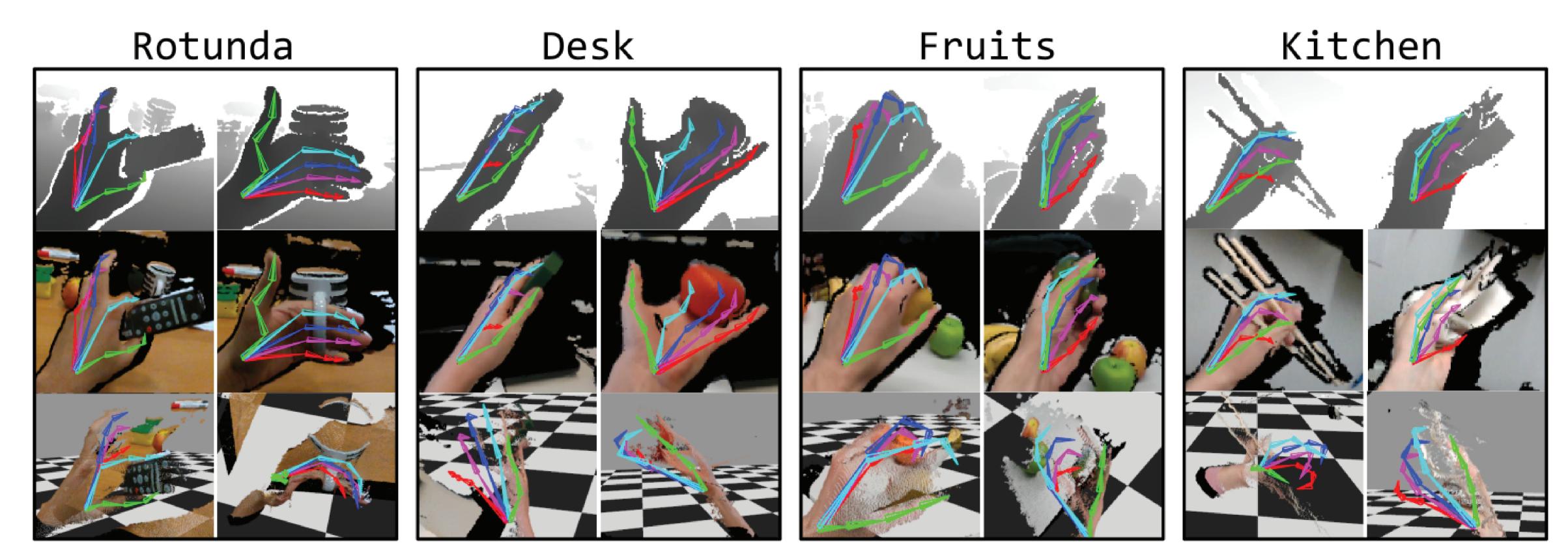
Comparison to Rogez at al., CVPR 2015



Live Capture Setup

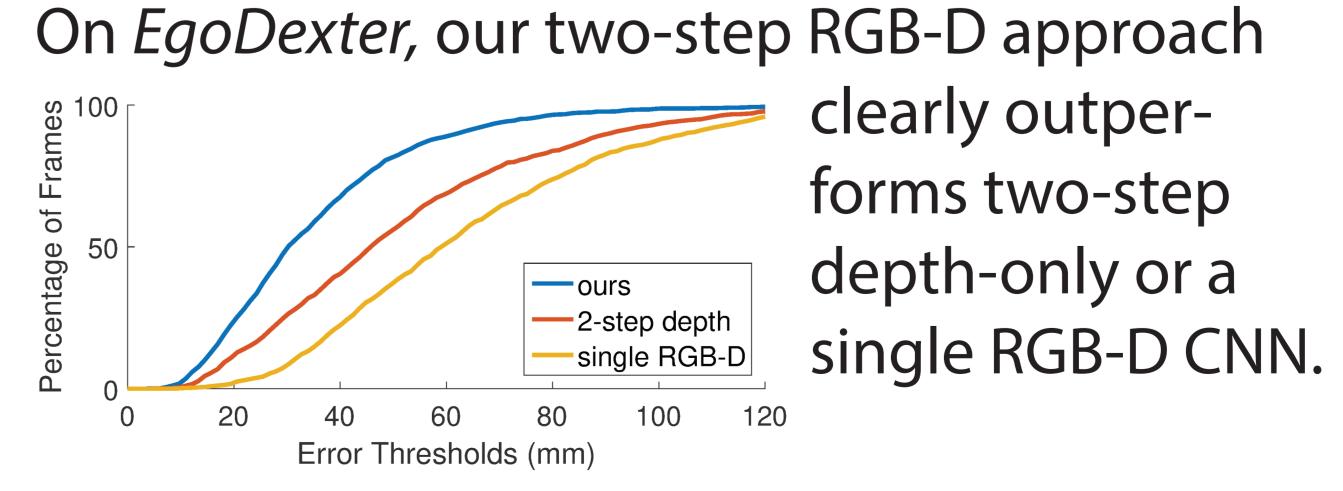


Results on our New Real Benchmark EgoDexter



Quantitative Analysis

On the SynthHands test set, predicting all 21 keypoints (JORNet) improves over only predicting fingertips and wrist (JOR-Net light).



Ablative Analysis

