

Tangible Interaction for AR/VR

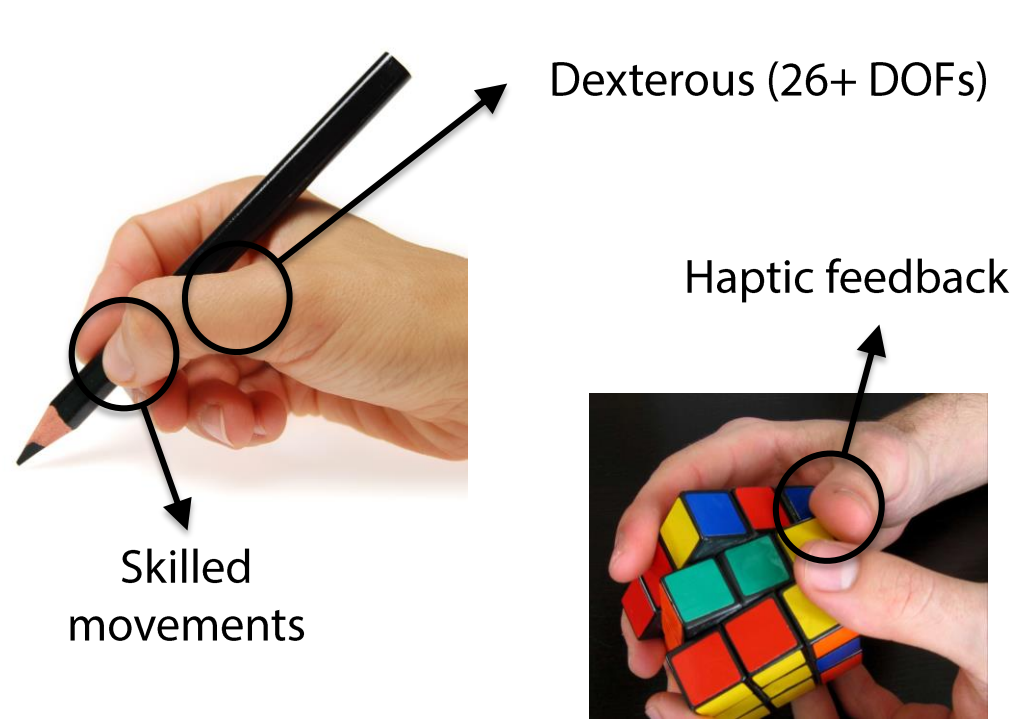
AR/VR devices need **new forms of input**.



We need to **increase the expressiveness** of interaction beyond controllers and simple gestures.

Joint Hand-Object Tracking

ADVANTAGES

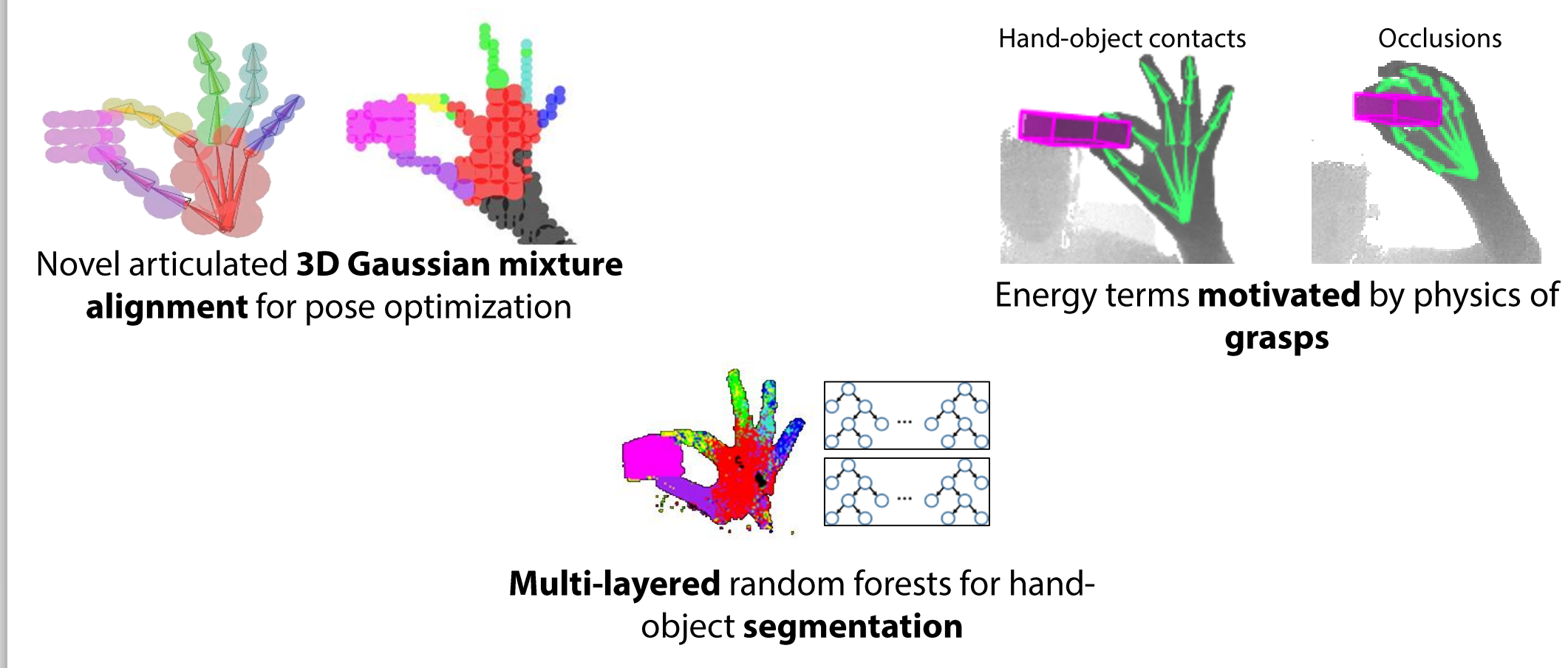


CHALLENGES

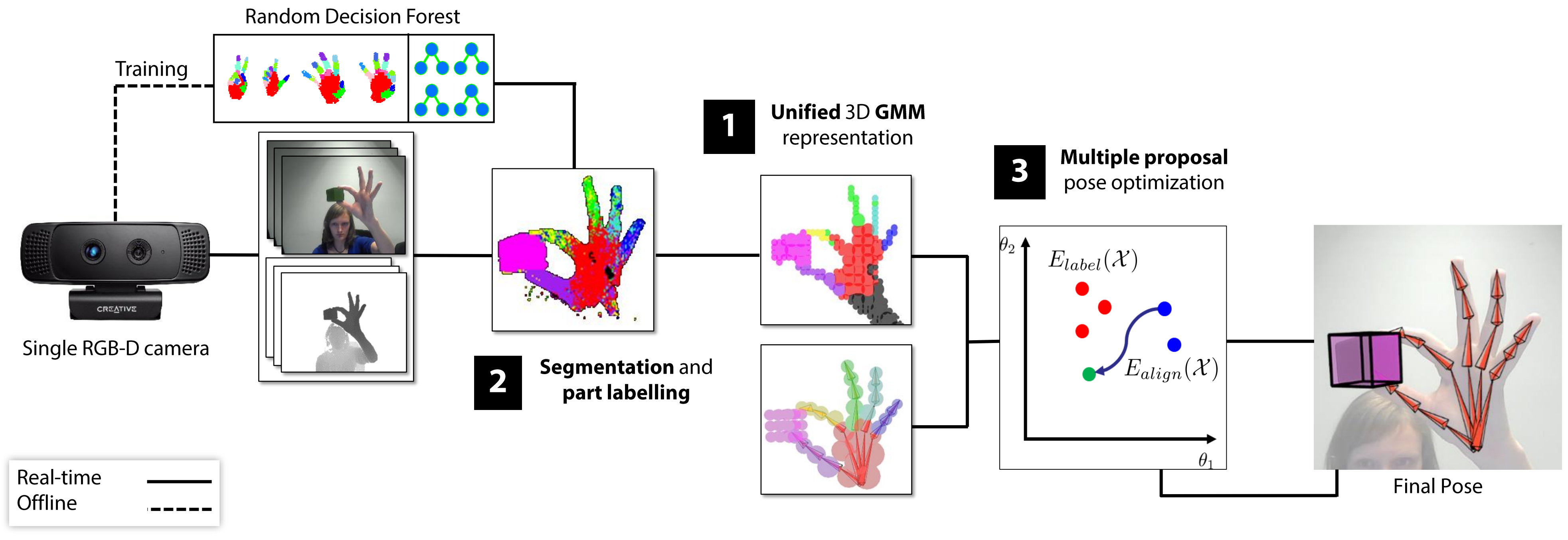
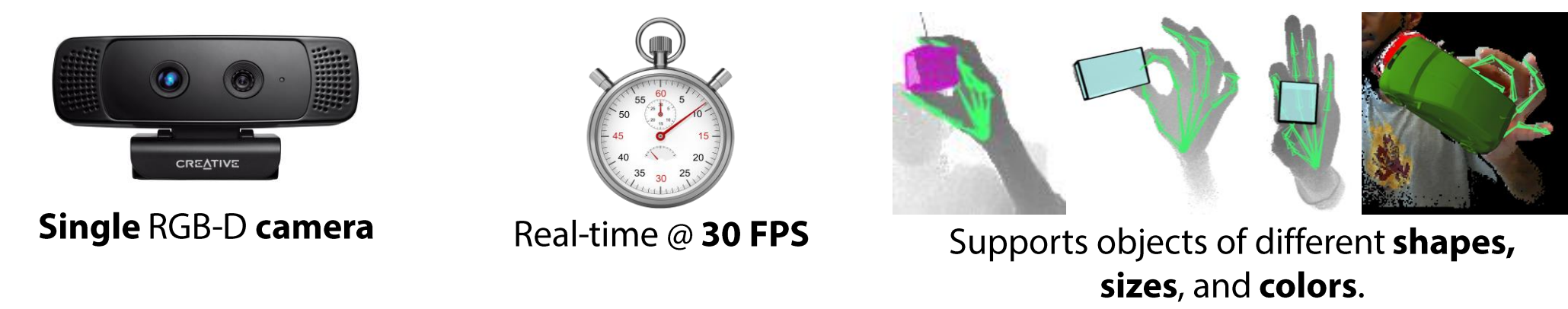
- **Extreme occlusions** by objects
- **Complex and fast motions**
- **Segmentation** of hand from object
- **High dimensional** problem
- **Run time** constraint

Contributions

NOVEL FRAMEWORK FOR REAL-TIME HAND-OBJECT TRACKING



PRACTICAL



3 Multiple Proposal Pose Optimization

OBJECTIVE 1:

$$E_{align}(\mathcal{X}) = \underbrace{E_a}_{\text{Spatial alignment}} + \underbrace{w_p E_p}_{\text{Anatomical plausibility}} + \underbrace{w_t E_t}_{\text{Temporal smoothness}} + \underbrace{w_c E_c + w_o E_o}_{\text{Contact and occlusion terms}}$$

OBJECTIVE 2:

$$E_{label}(\mathcal{X}) = \underbrace{E_a}_{\text{Spatial alignment}} + \underbrace{w_s E_s}_{\text{Semantic alignment}} + \underbrace{w_p E_p}_{\text{Anatomical plausibility}}$$

SPATIAL ALIGNMENT

$$E_a(\mathcal{X}) = \int_{\Omega} \left[(\mathcal{M}_{d_h}(\mathbf{x}) - \mathcal{M}_h(\mathbf{x}))^2 + (\mathcal{M}_{d_o}(\mathbf{x}) - \mathcal{M}_o(\mathbf{x}))^2 \right] dx$$

Aligns model and input 3D Gaussian mixtures

SEMANTIC ALIGNMENT

$$E_s(\mathcal{X}) = \sum_{i=1}^{N_s} \sum_{j=1}^{N_s} \alpha_{i,j} \cdot \|\mu_i - \mu_j\|_2^2$$

Aligns GMMs with semantic label information

CONTACT POINTS

$$E_c(\mathcal{X}) = \sum_{(k,l) \in \mathcal{T}} \left(\|\mu_k - \mu_l\|_2^2 - t_{kl}^2 \right)^2$$

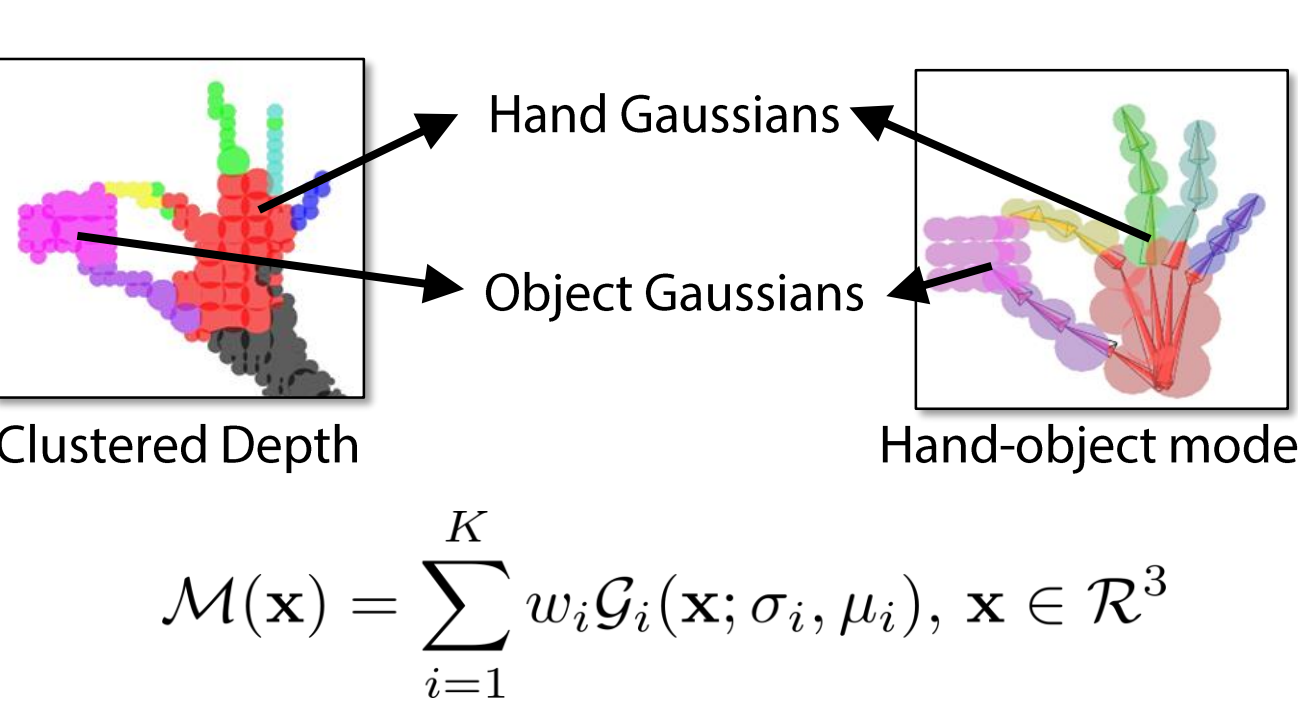
Enforces contact between fingertip and object Gaussians that are close

OCCUSION HANDLING

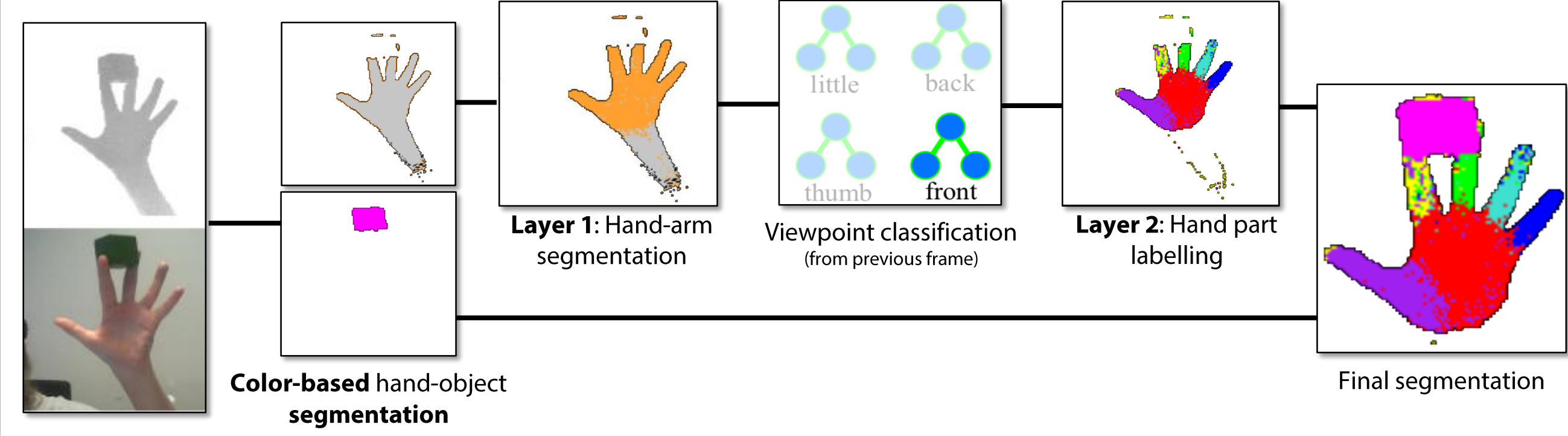
$$E_o(\mathcal{X}) = \sum_{i=0}^{N_h} \sum_{j \in \mathcal{H}_i} (1 - f_i) \cdot \|x_j - x_j^{old}\|_2^2$$

Forces occluded parts of the hand to move consistently with other parts

1 GMM Representation

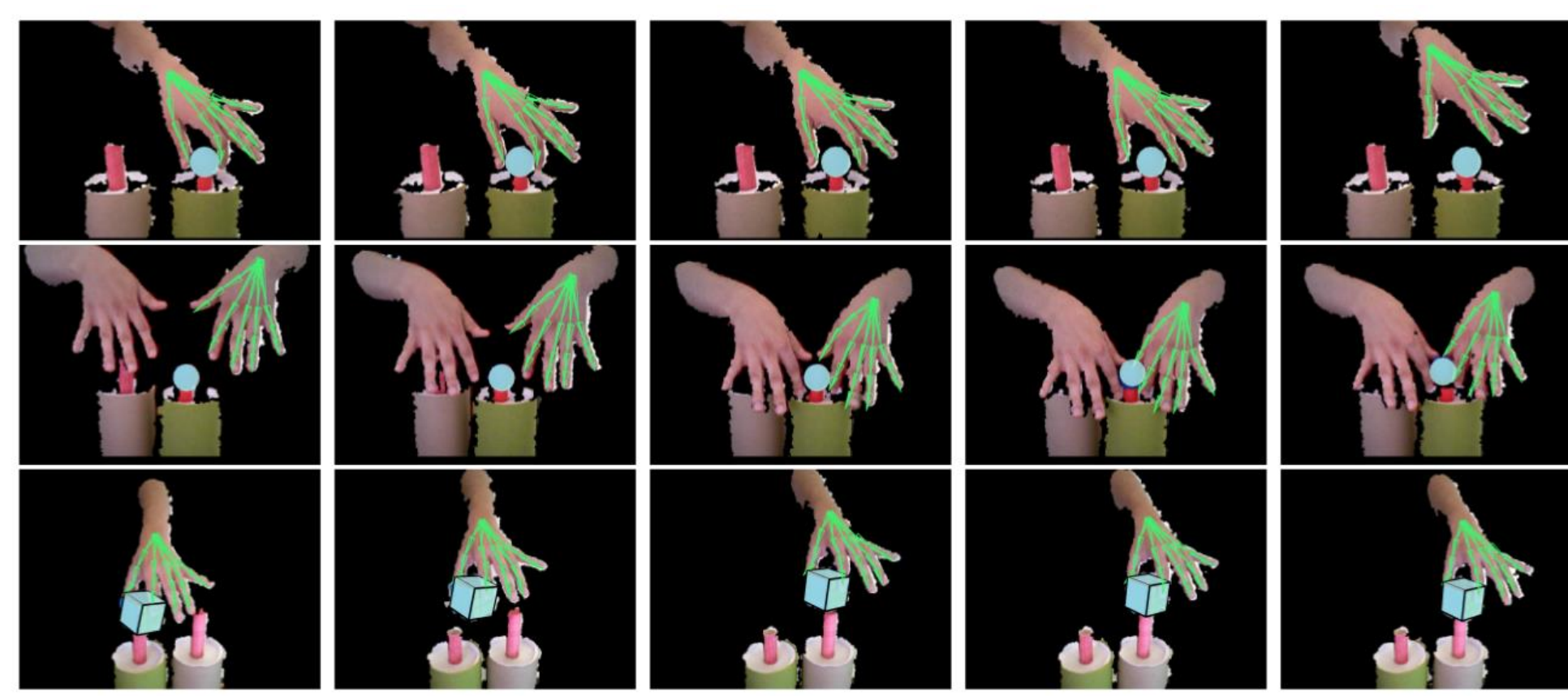


2 Segmentation and Part Labelling

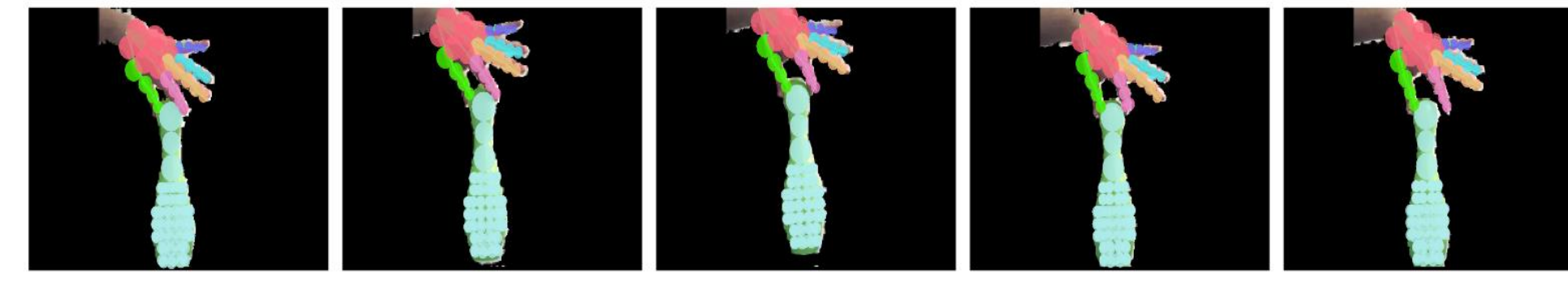


Results

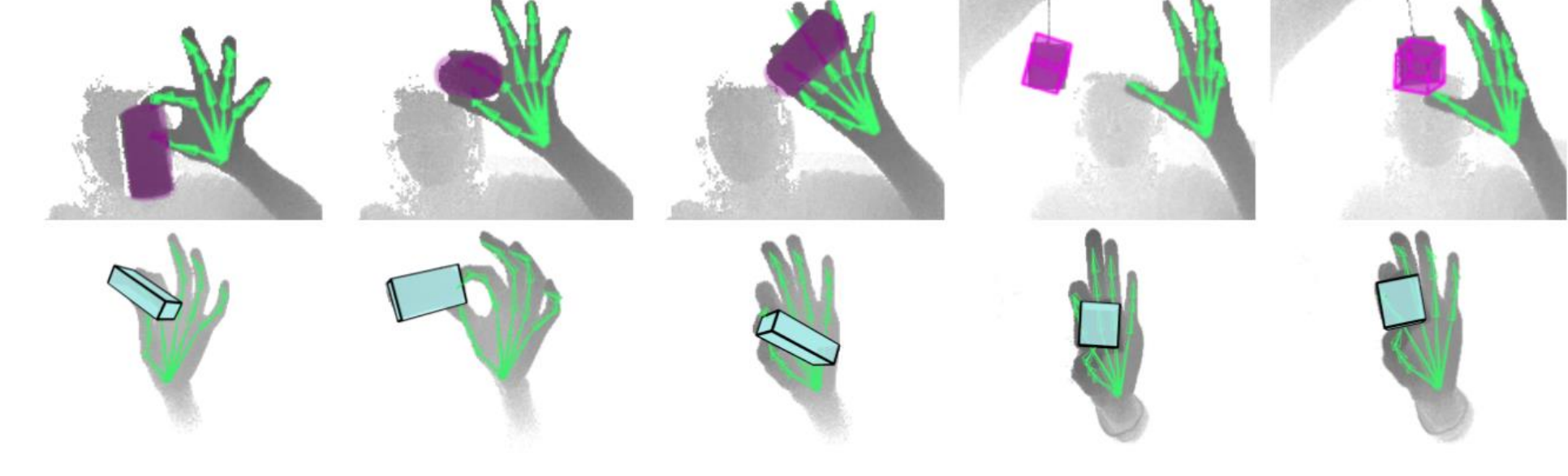
RESULTS ON DATA FROM TZIONAS ET AL. (IJCV 2016)



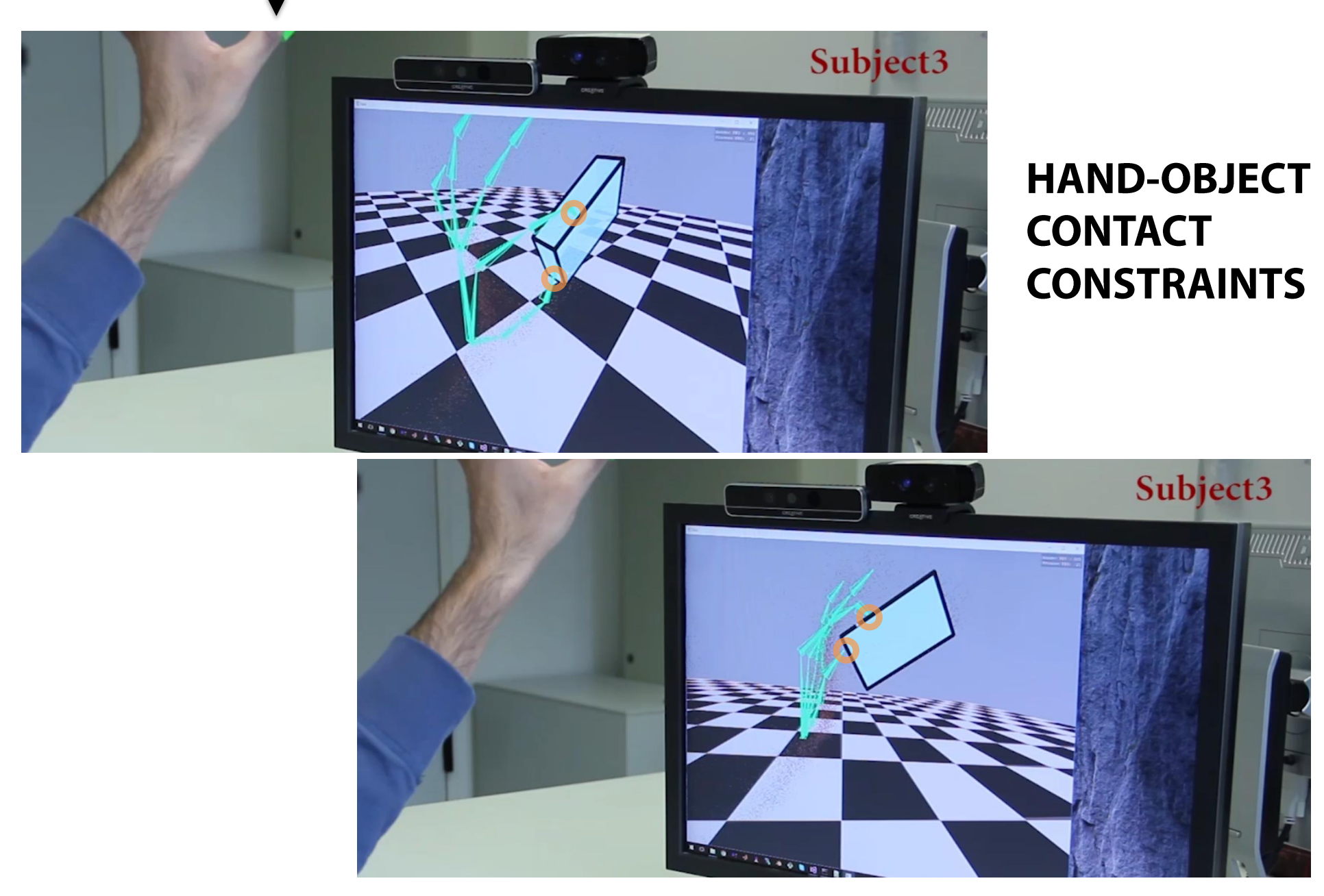
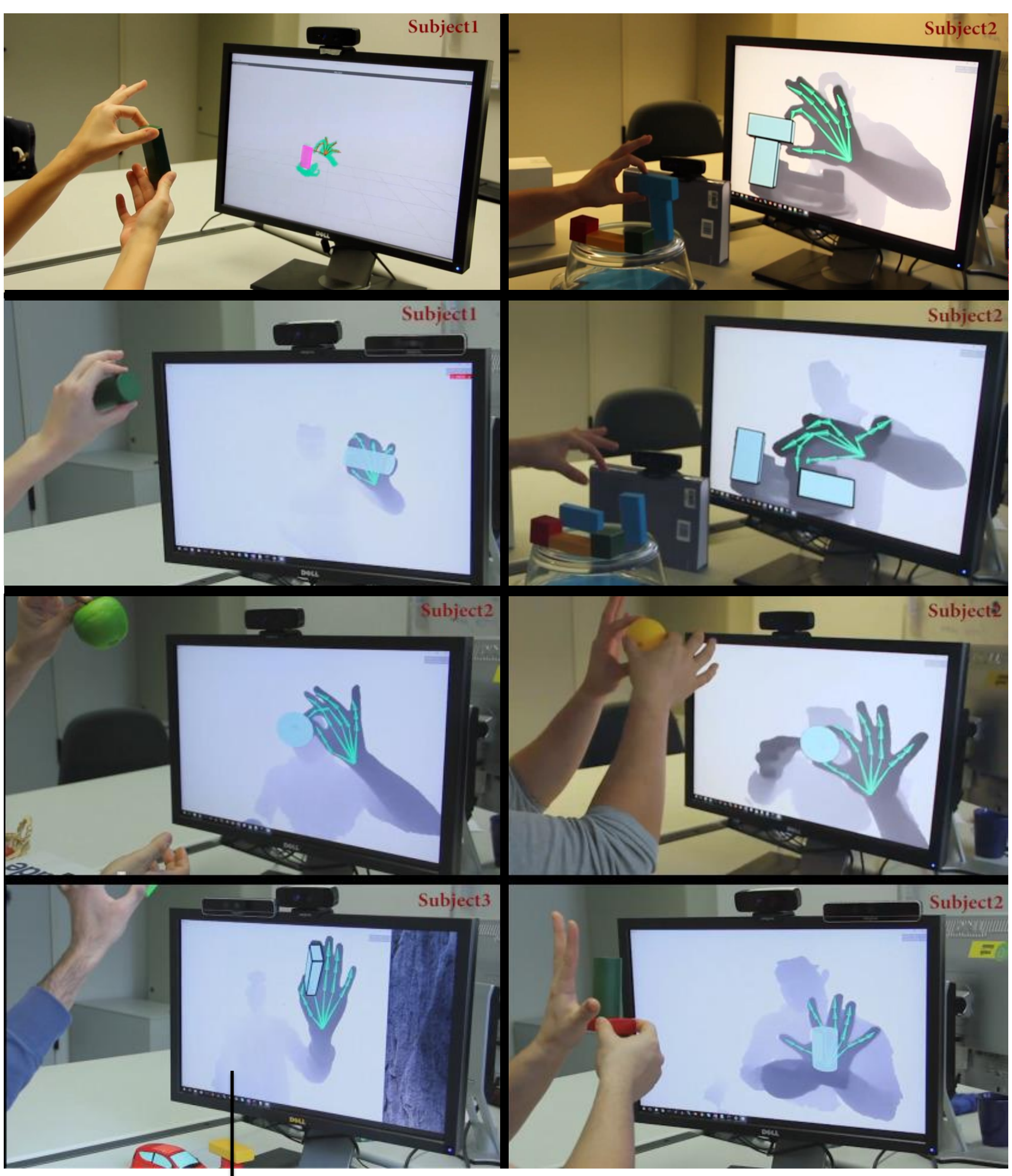
RESULTS ON DATA FROM TZIONAS AND GALL (ICCV 2015)



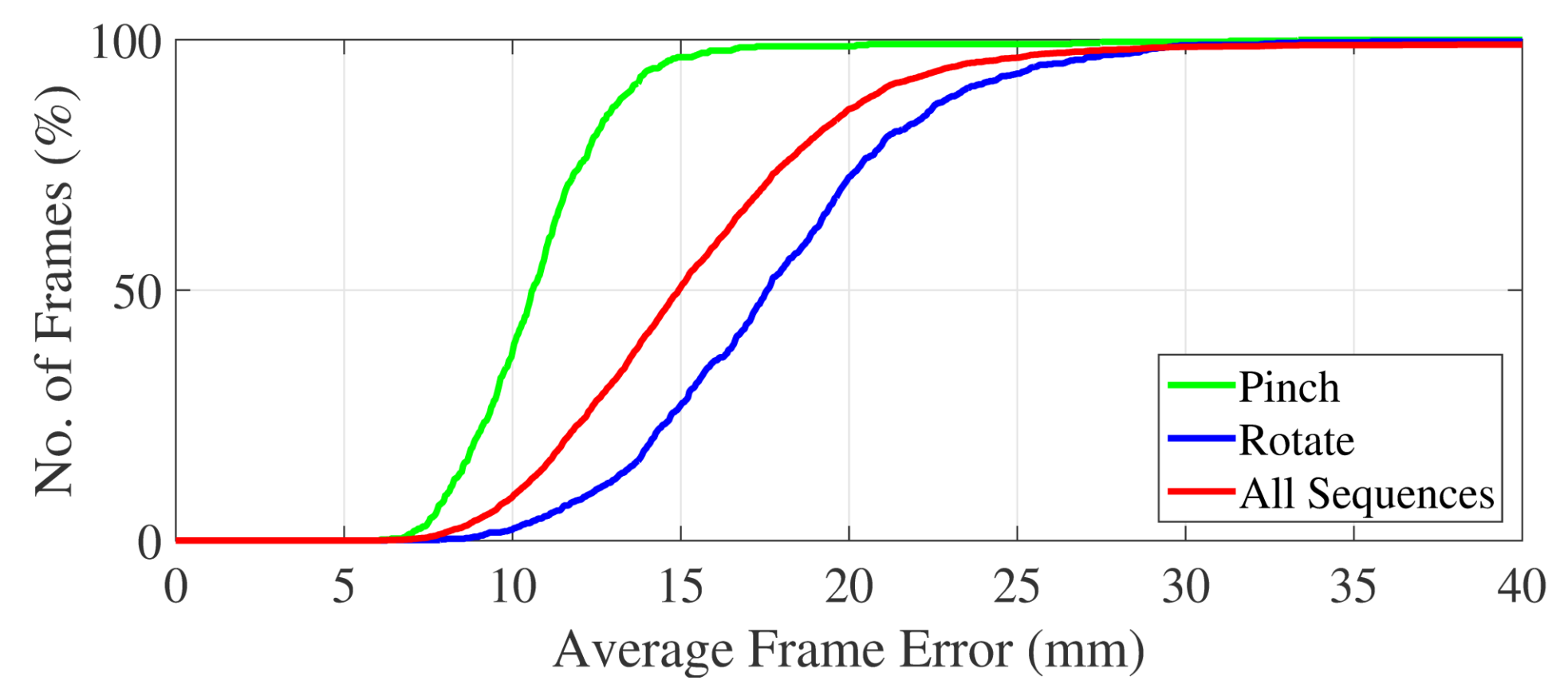
REAL-TIME TRACKING



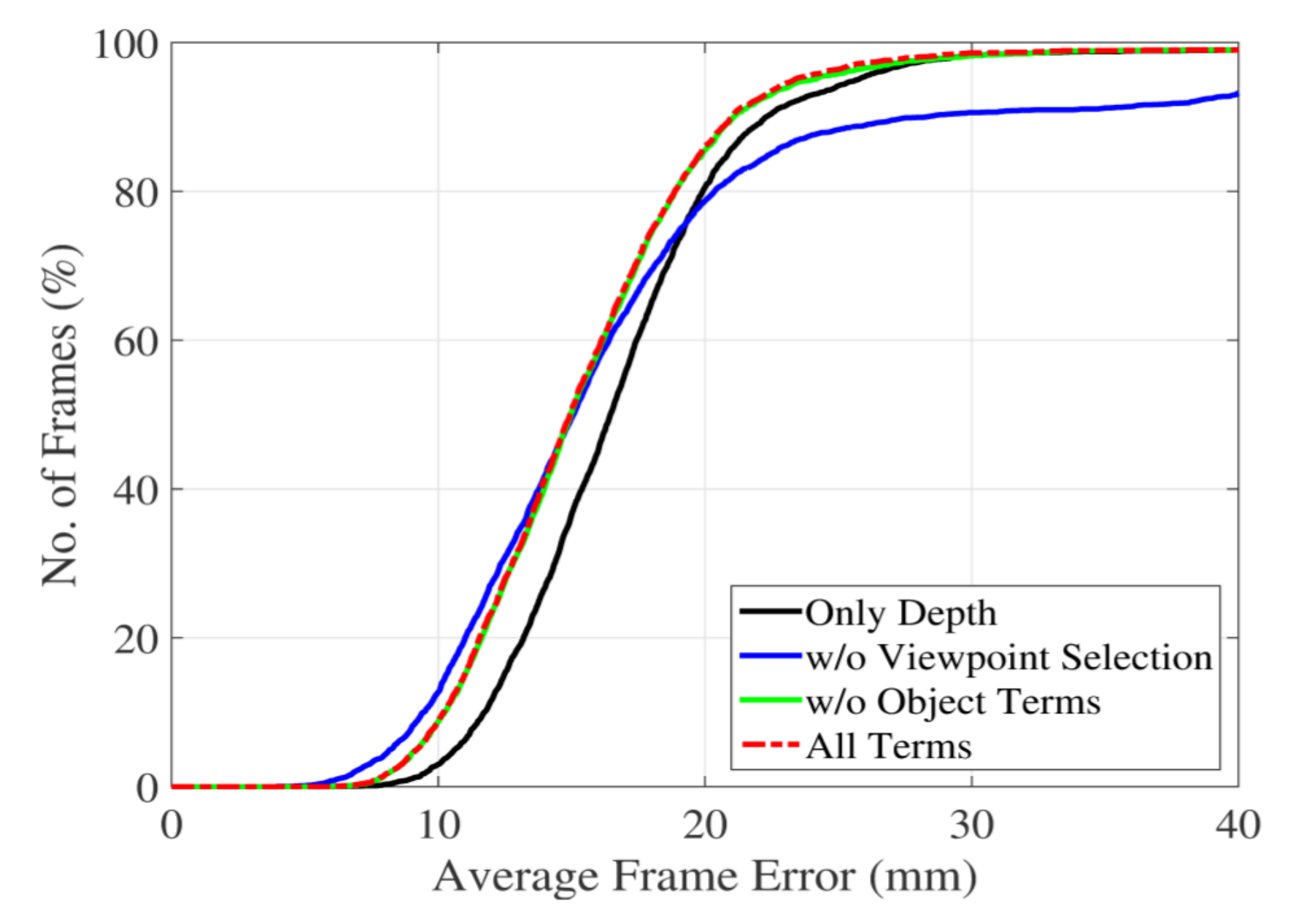
LIVE CAPTURE SETUP



Tablet with video goes here



Tracking consistency for the best, worst and average case in Dexter+Object. We are consistently below 30 mm.



Ablative analysis with different terms disabled shows robustness of our method.

- Runs at **30 FPS**
- Objects of **various shapes and sizes**
- Works with **different users**
- Average joint position error of **15.7 mm** on Dexter+Object
- 3D Gaussian mixture alignment improves hand-only tracking results on Dexter by **2 mm**



Dexter+Object dataset available!

handtracker.mpi-inf.mpg.de



For access to a wide range of human shape and performance capture datasets, please visit:

gvvperfcapeva.mpi-inf.mpg.de



Supported by the ERC Starting Grant projects CapReal (335545) and COMPUTED (637991), and the Academy of Finland.

We would like to thank Christian Richardt.