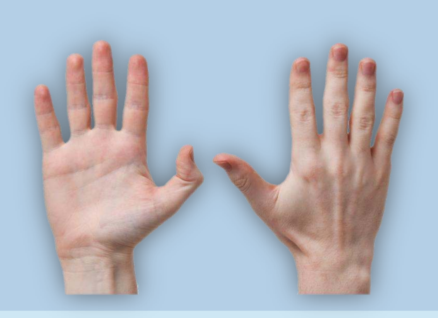
Interactive Markerless Articulated Hand Motion Tracking Using RGB and Depth Data

Srinath Sridhar, Antti Oulasvirta, Christian Theobalt

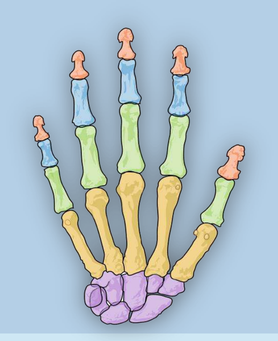
MPI Informatik and Saarland University

{ssridhar,oantti,theobalt}@mpi-inf.mpg.de

PROBLEM



Markerless tracking of hand motion



Model the hand as a **skeleton**

CHALLENGES

- Many **DoFs**
- **Fast finger motion**
- **Self-occlusions**
- **Homogeneous** skin **color** distribution

CONTRIBUTIONS

1. **Part-based Hybrid Approach**: We combine generative and discriminative methods using a part-based pose retrieval strategy.
2. **Interactive**: Our method captures hand motions with a level of precision and speed necessary for interactive applications.
3. **Annotated Dataset**: We provide an extensive, annotated benchmark dataset consisting of general hand motion sequences.

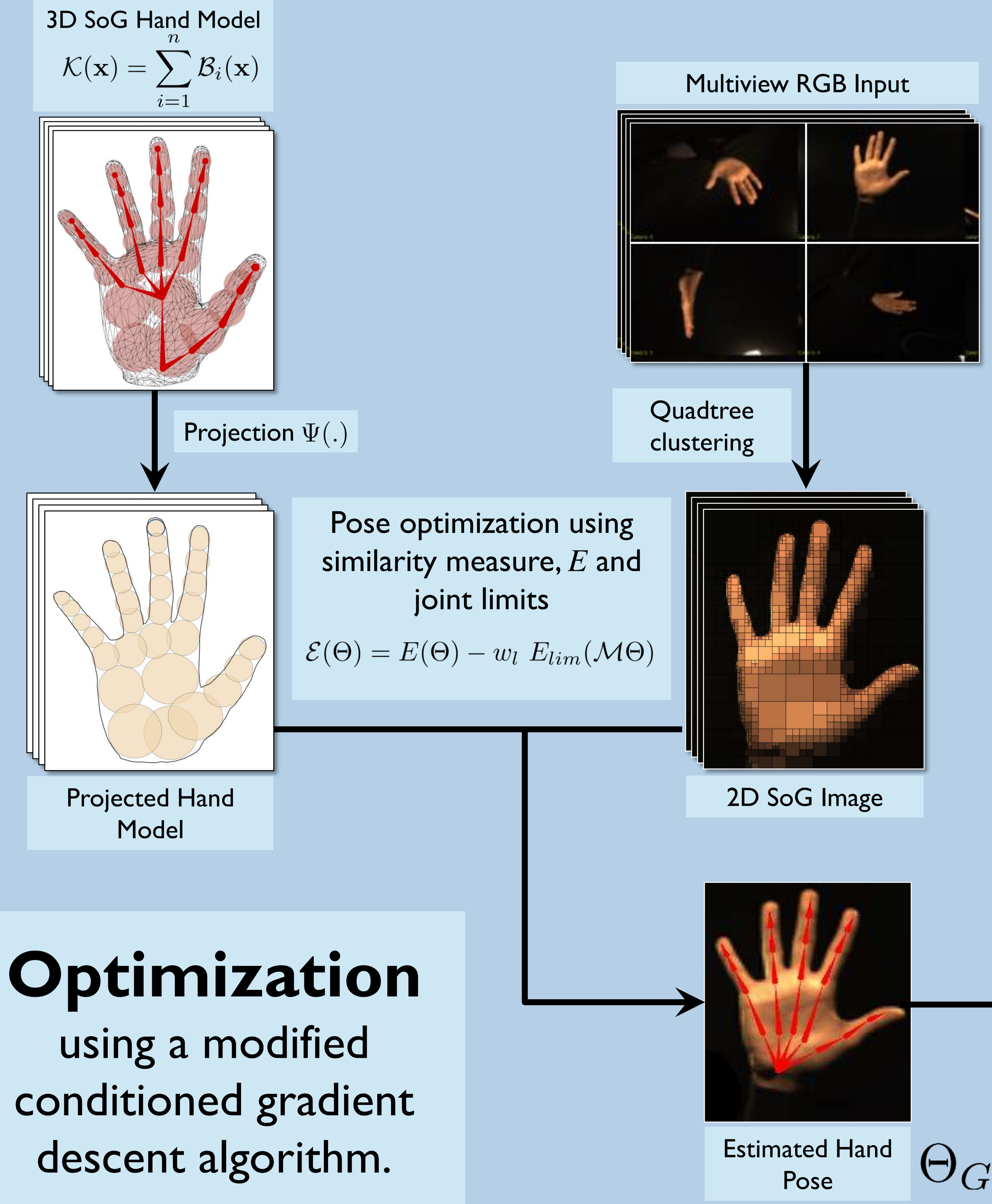
1 Generative Hand Pose Estimation

Sum of Gaussians (SoG) Representation [1]

+ Results in mathematically smooth similarity, E
+ Yields analytical formulation of energy function and its derivative

$$\mathcal{B}(\mathbf{x};\mu,\sigma) = \exp\left(-\frac{\|\mathbf{x}-\mu\|^2}{2\sigma^2}\right) \qquad E_{ij} = d(\mathbf{c}_i,\mathbf{c}_j)\int_\Omega \mathcal{B}_i(\mathbf{x})\mathcal{B}_j(\mathbf{x})\,\mathrm{d}\mathbf{x}$$

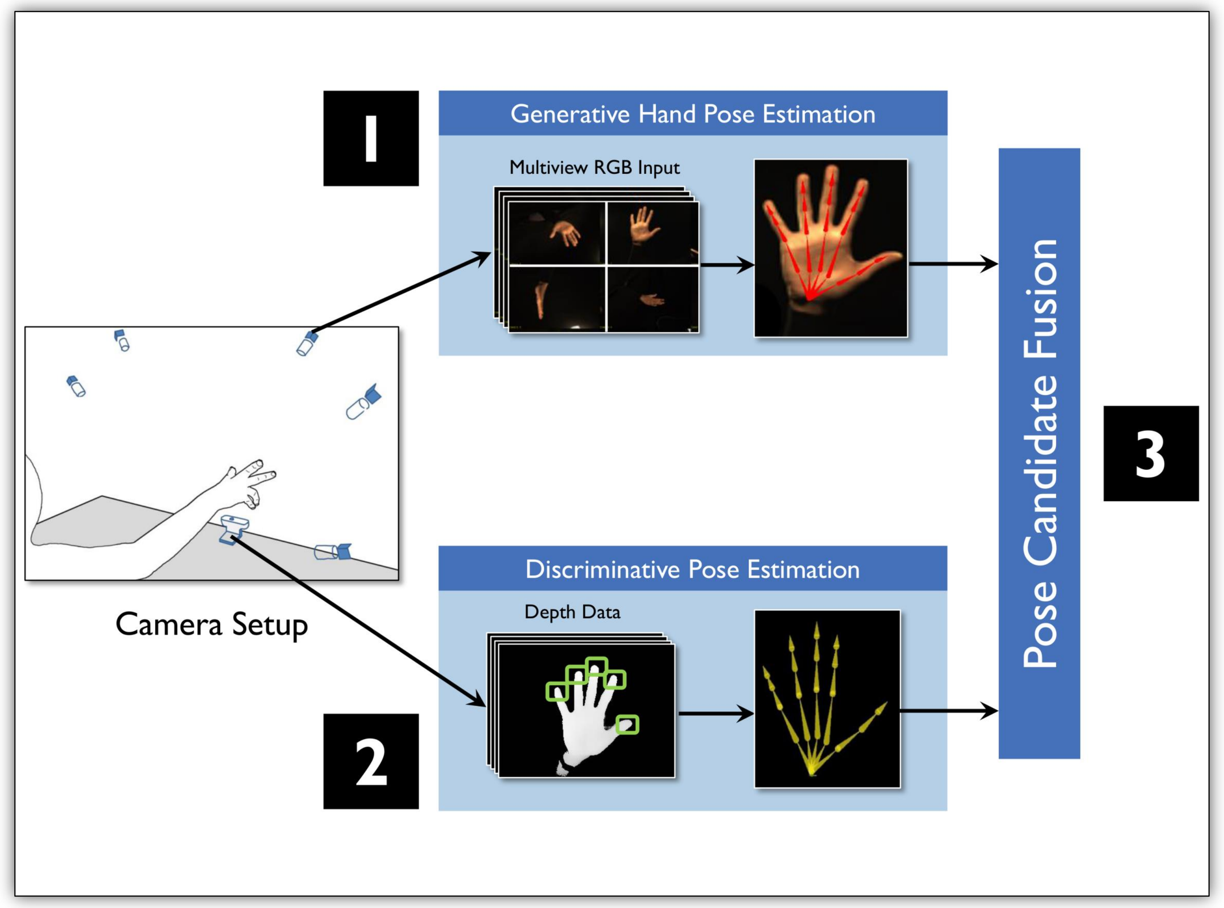
$d(\mathbf{c}_i,\mathbf{c}_j)$ is the Wendland function



Optimization

using a modified conditioned gradient descent algorithm.

[1] C. Stoll, N. Hasler, J. Gall, H. Seidel, and C. Theobalt, *Fast articulated motion tracking using a sums of Gaussians body model*, in Proc. of ICCV 2011.

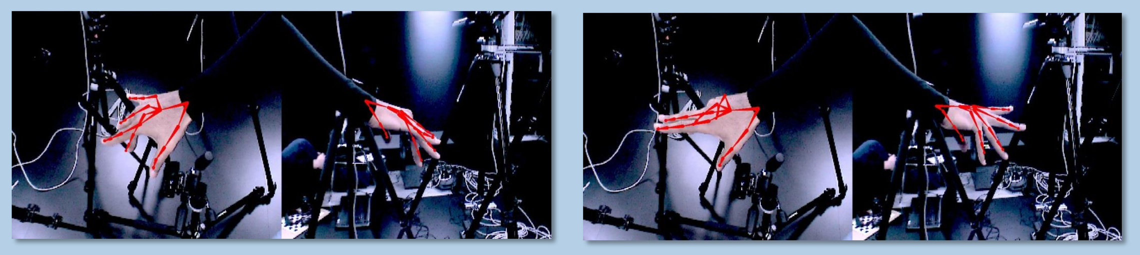


3 Pose Candidate Fusion and Results

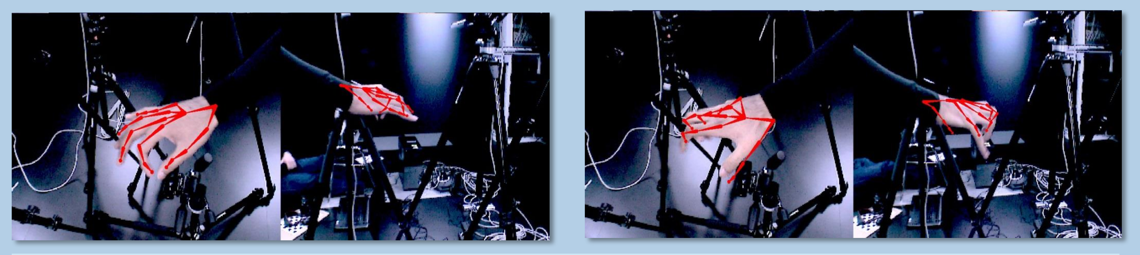
Hybrid Pose Fusion

+ Helps recover from local optima (see 4) that occur in generative pose estimation
+ Enables automatic pose initialization

$$\Theta_F = \underset{\Theta \in \{\Theta_G, \Theta_D\}}{\arg\max} \; \{\mathcal{E}(\Theta_G), \mathcal{E}(\Theta_D)\}$$



Tracking results for finger articulations



Tracking under challenging motions and image blur

2 Discriminative Pose Estimation

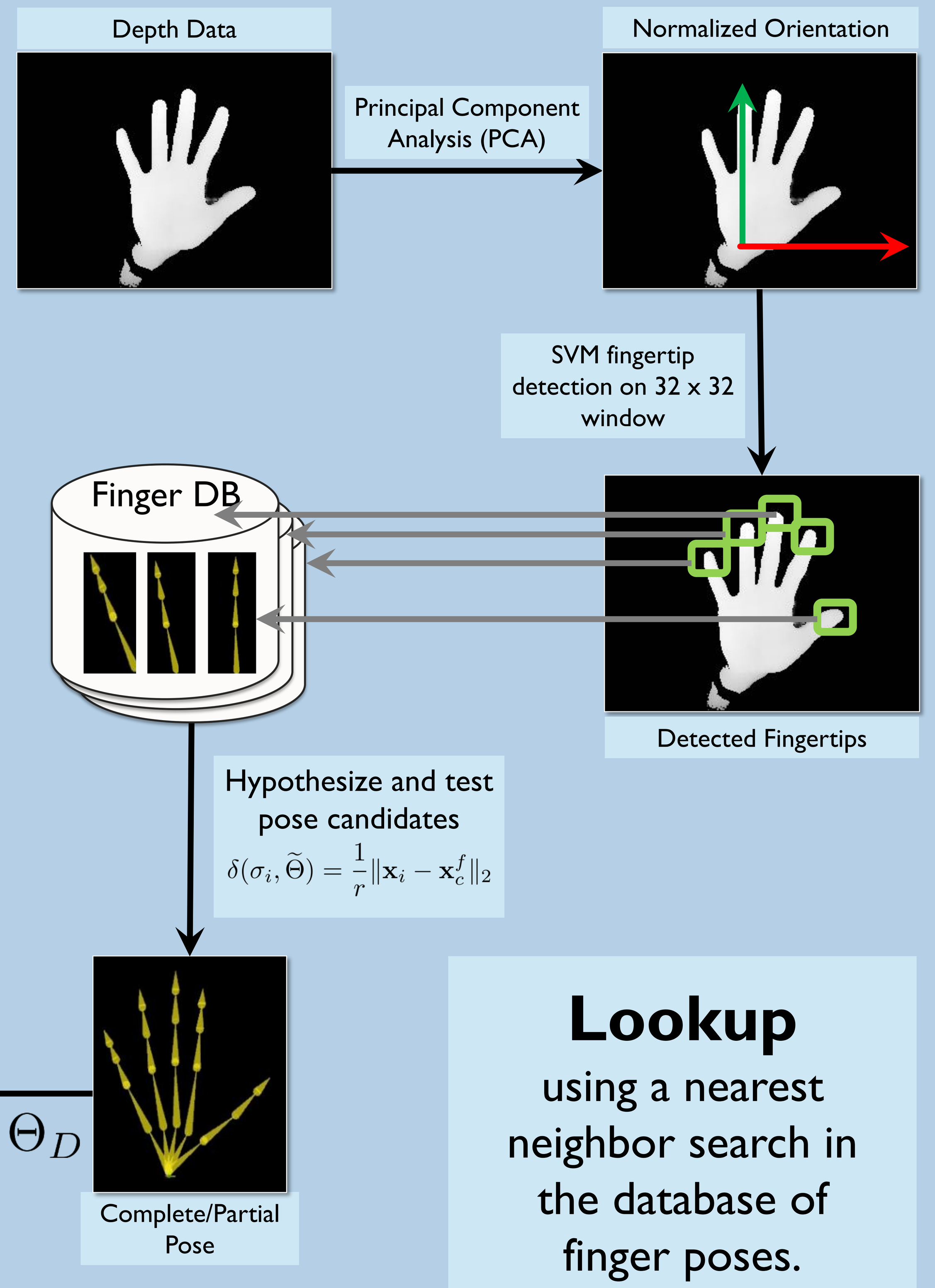
Part-based Pose Retrieval

+ Allows creation of a small pose database for each finger
+ Enables recovery of partial hand pose

Example: Part-based database: 5 x 81 poses

Global pose database: 10^{10} poses

Assuming 3 discretizations per DoF

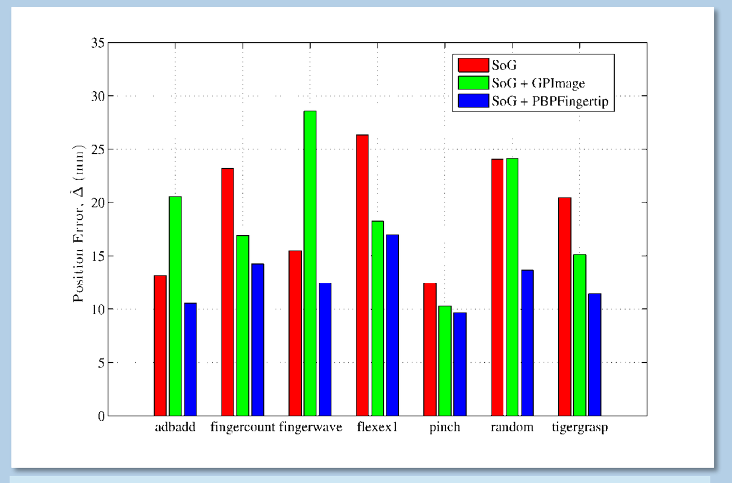


Lookup

using a nearest neighbor search in the database of finger poses.

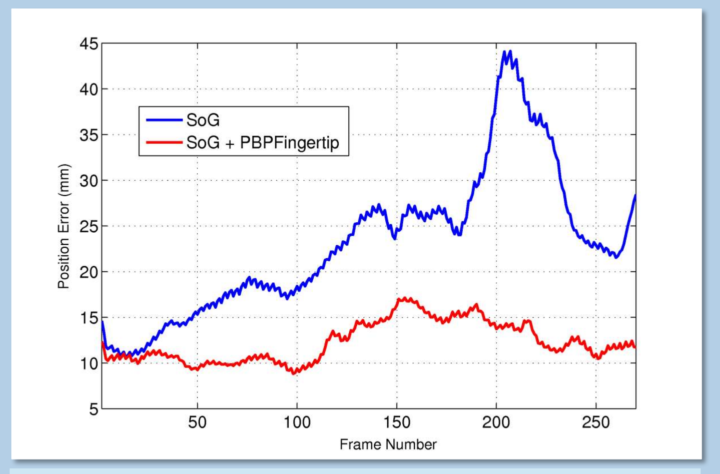
4 Quantitative Results and Dataset

10 FPS Interactive framerate



The average position error over the entire dataset. Our approach (blue) results in smaller errors.

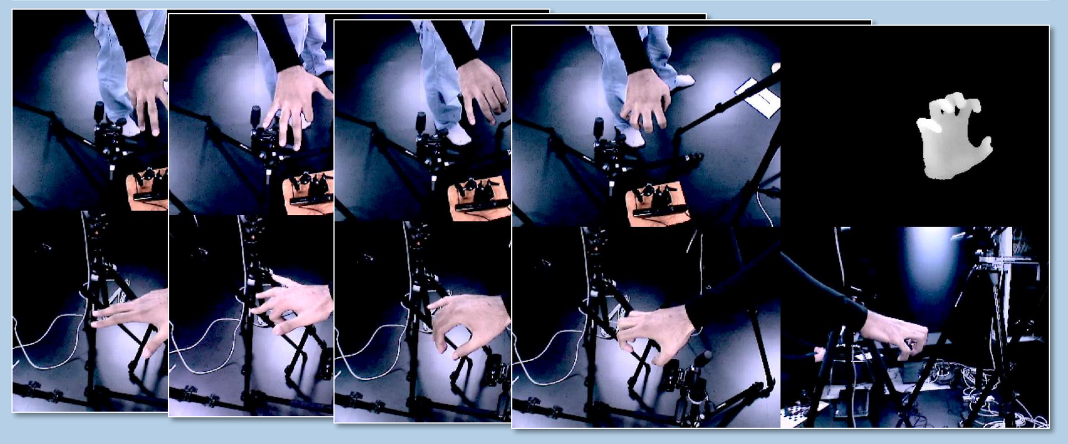
13 mm average error over all datasets



The average position error over the fingerwave dataset. Our hybrid approach (red) does not drift.

Hybrid approach avoids error accumulation

adbadd | fingerwave | pinch | tigergrasp | fingercount | flexex1 | random



Dexter 1

7 RGB+depth **sequences** | **Slow/fast** finger motion with **self-occlusions** | **Fingertip annotations** for all frames

Available for download!

handtracker.mpi-inf.mpg.de

Supported by the Max Planck Center for Visual Computing and Communication (MPC-VCC) and the ERC Starting Grant CapReal. Thanks to Han Du, James Tompkin and Thomas Helten.