Realtime and Robust Hand Tracking from Depth

Chen Qian ; Xiao Sun ; Yichen Wei ; Xiaoou Tang ; Jian Sun, "Realtime and Robust Hand Tracking from Depth, 2014 IEEE

2019. 04. 01

Minji Jo, Hanyang Univ.

Hand tracking

- traditional approach :
  - Glove-based technologies
  - PlayStation Eye
- 2009~2011 :
  - :Microsoft Kinect
- 2016 :
  - LEAP MOTION

Hand tracking step

- hand detection
- finger identification
- (gesture recognition)

Hand model

- we adopt the commonly used 26 degrees of freedom (DOF) hand motion model
- We preserve the kinematic constraints
- the polygonal mesh model

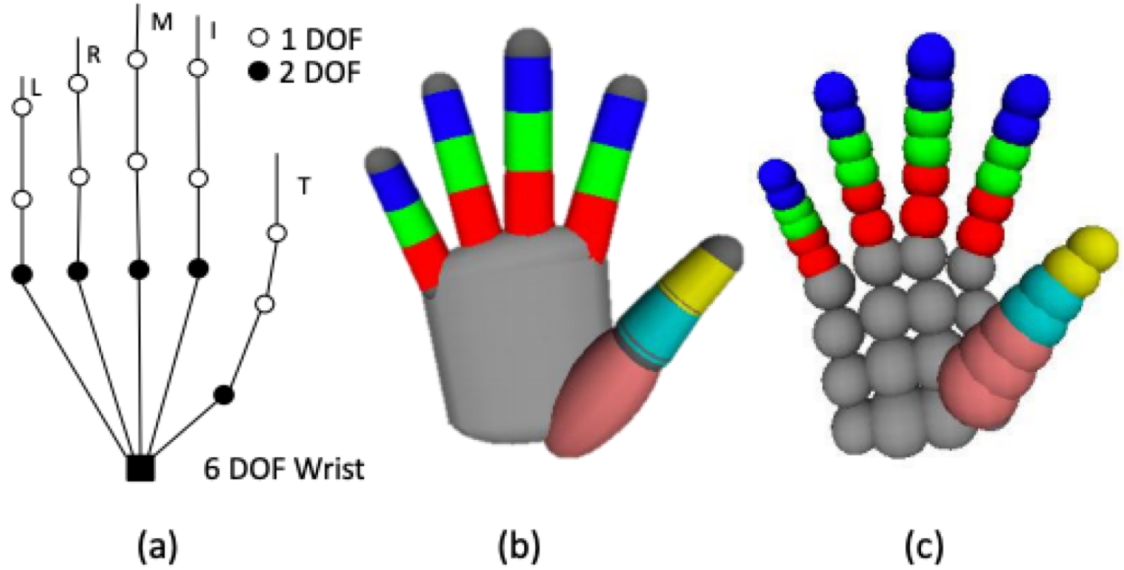


Figure 1. (a) 26 DOFs hand motion model; (b) the hand model in [8]; (c) our hand model that approximates (b) using 48 spheres.

Cost function

- the discrepancy between the hand model and input depth

$$\lambda \cdot \sum_{p \in sub(\mathcal{P})} D(p, s_{x(p)})^2 + \sum_i B(c_i, \mathcal{D})^2 + \sum_{i,j} L(s_i, s_j)^2. \quad (1)$$

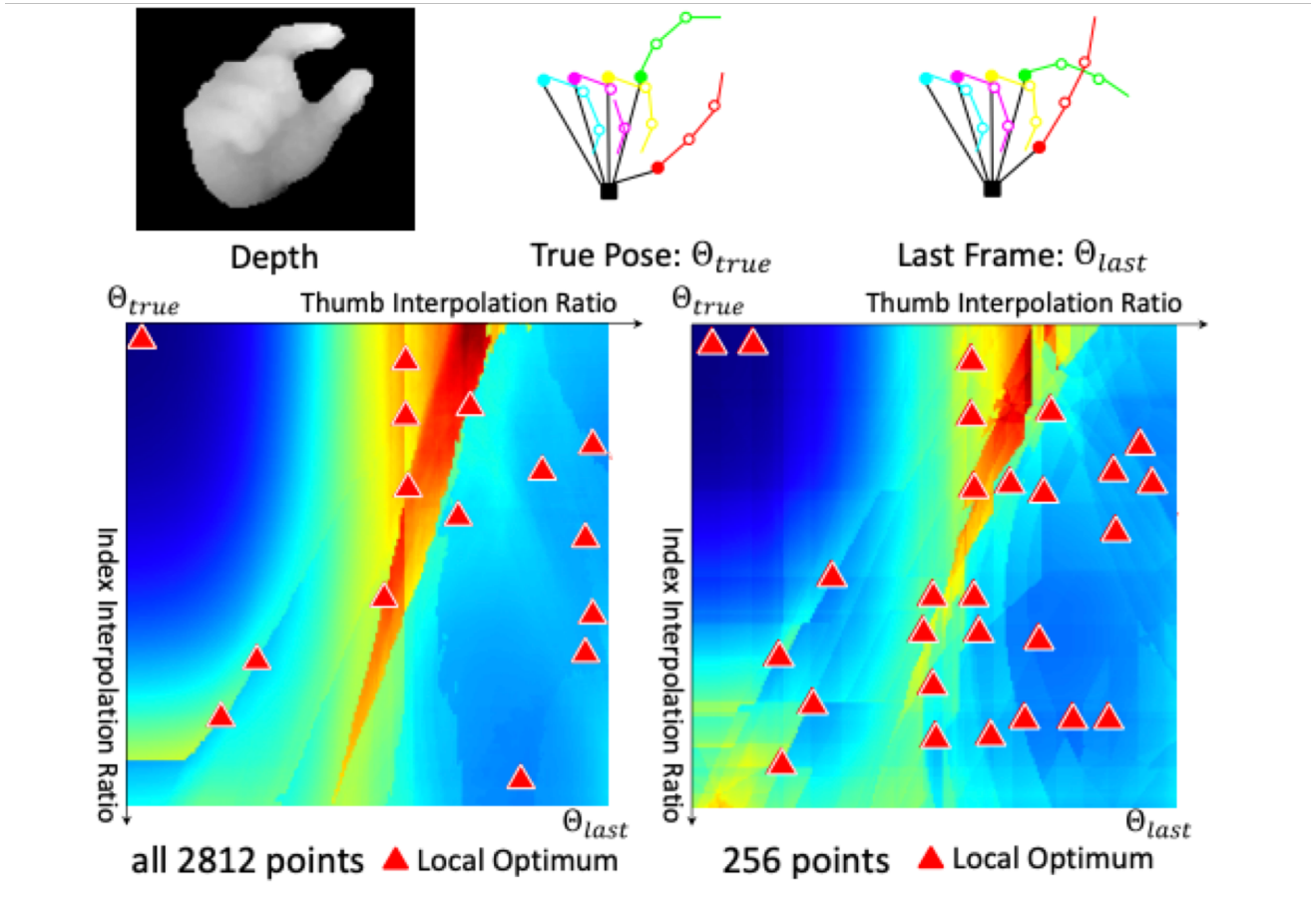
$$D(p, s) = abs(||p - c||_2 - r).$$

$$B(c, \mathcal{D}) = \begin{cases} \max(0, \mathcal{D}(j(c)) - c_z) & \text{if } \mathcal{D}(j(c)) \text{ is valid} \\ dist(j(c), \text{silhouette of } \mathcal{D}) & \text{otherwise} \end{cases} \quad (3)$$

$$L(s_i, s_j) = \max(r_i + r_j - ||c_i - c_j||_2, 0).$$

ICP,PSO Local Optimization

- Iterated Closest Point (ICP) method
  - cannot handle non-rigid object
  - ICP quickly reaches local optima
- Particle Swarm Optimization (PSO)
  - it does not work well in our case
  - suffers from premature convergence



ICP+PSO Local Optimization

- each particle takes an additional ICP like gradient descent step before the random particle movement in each PSO generation

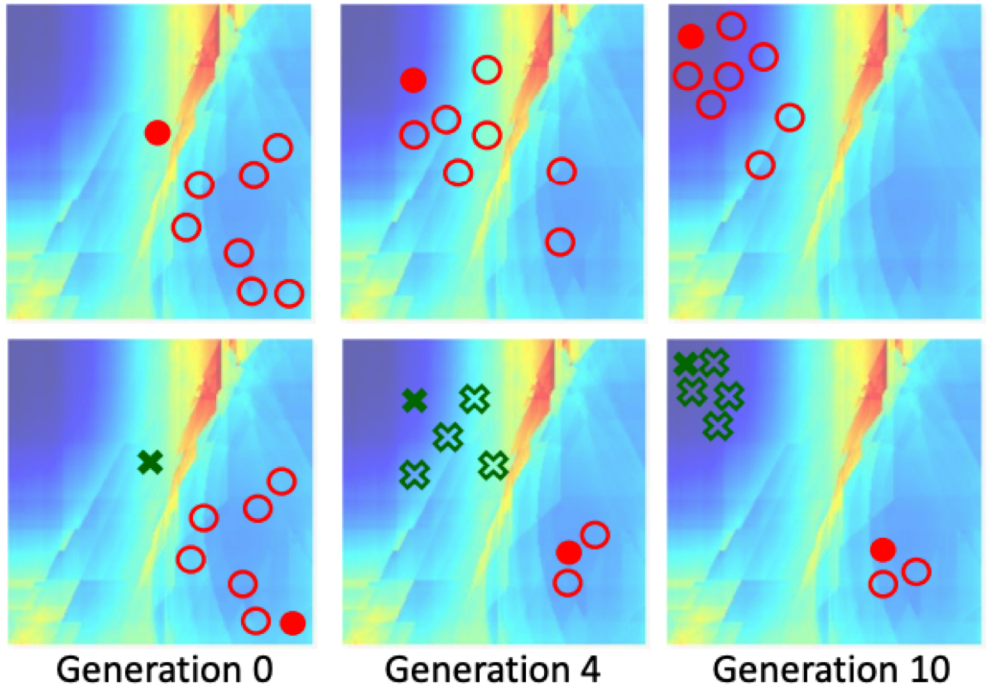


Figure 3. (better viewed in color) Illustration of the ICP-PSO optimization processes using the example in Figure 2 for $k = 1$ (top) and $k = 2$ (bottom). See text for details.

Finger Detection for Hand Initialization

- detect the extreme points on 2D XY plane and 1D Z direction NOT 3D

- tip/non-tip : finger segment geometric checking
  - XY-Fingers
  - Z-Fingers

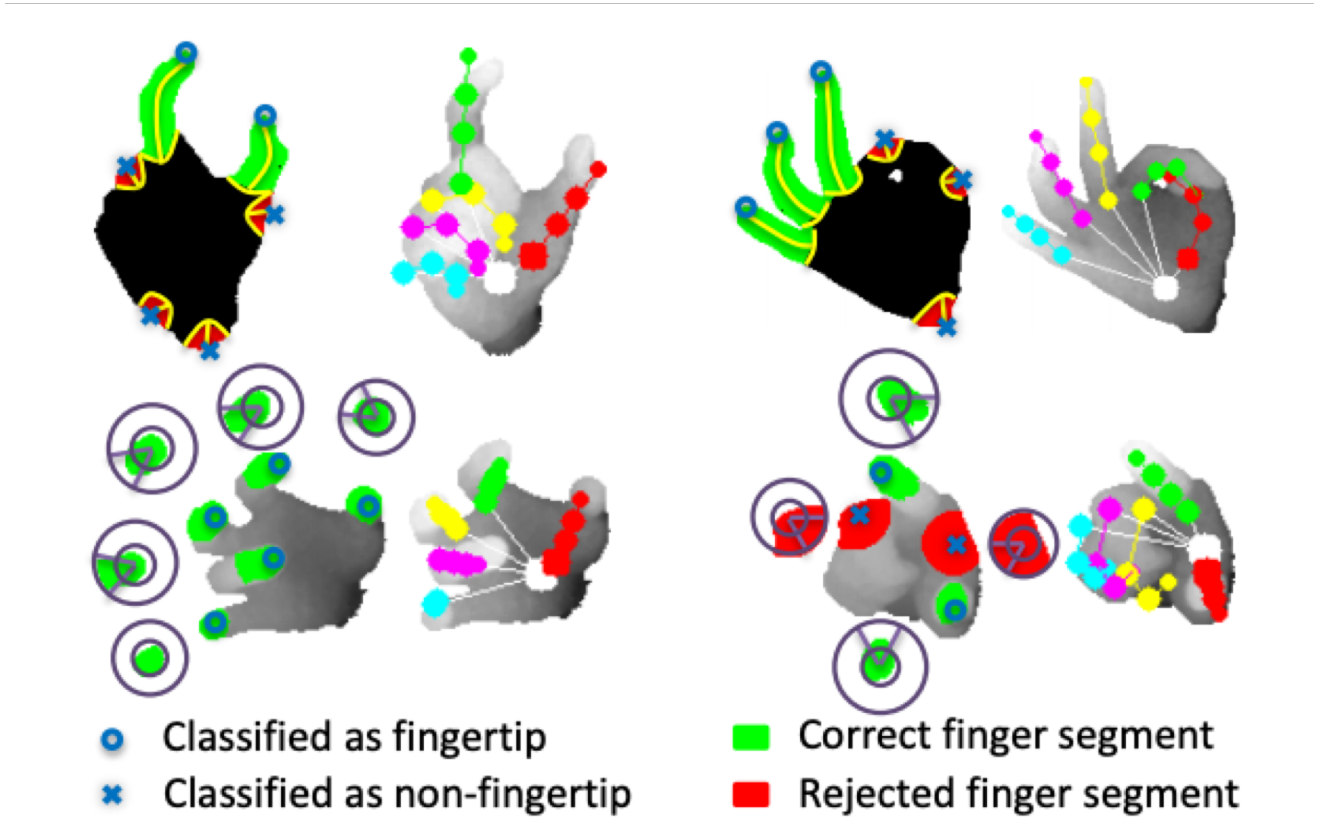


Figure 4. (better viewed in color) Illustration of finger detection

Experiments

Subject	1	2	3	4	5	6
Scale	1.1	1.0	0.9	0.95	1.1	1.0
FORTH	35.4	19.8	27.3	26.3	16.6	46.2
FORTH*	19.8	15.8	19.8	15.4	16.0	21.0
PSO	26.7	14.8	44.7	18.1	15.0	24.3
PSO*	18.6	12.1	21.2	14.4	13.7	22.4
ICP	27.3	20.7	34.4	25.1	17.00	32.8
ICP*	17.9	15.9	19.2	15.6	10.8	25.9
ICP-PSO	9.3	24.1	14.4	13.4	11.0	20.0
ICP-PSO*	**8.0**	**7.4**	**10.8**	**10.9**	**7.3**	**11.7**

Table 2. Model scales and average joint tracking errors (in mm) of 6 subjects. Methods with $*$ use initialization.

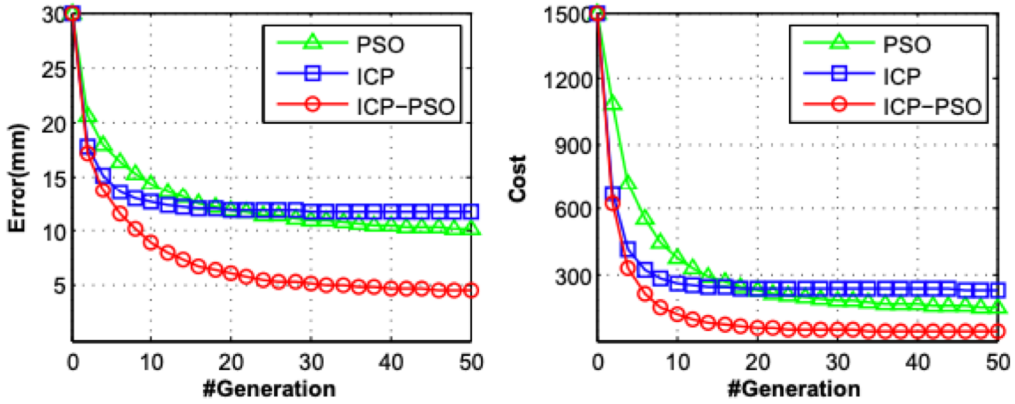


Figure 5. Average error and cost function values decrease as optimization methods run through generations.

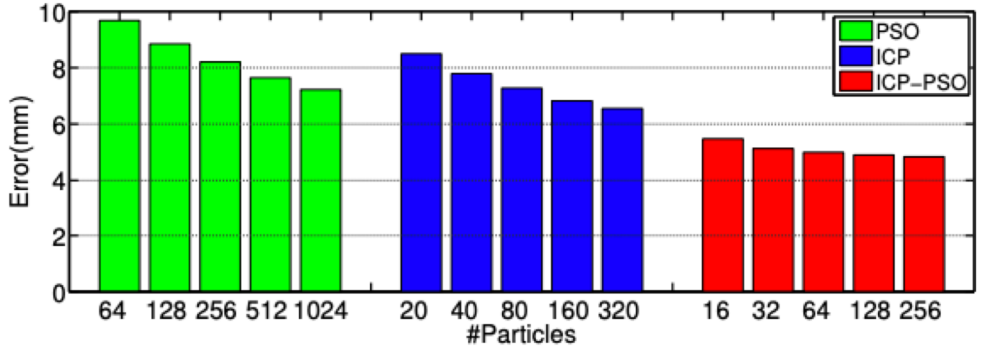


Figure 6. Average errors using different number of particles of the three methods.

E_{init}	15-25 mm				25-35 mm				35-45 mm			
Metric	E_{real}	E_{syn}	S_{real}	S_{syn}	E_{real}	E_{syn}	S_{real}	S_{syn}	E_{real}	E_{syn}	S_{real}	S_{syn}
PSO	9.80	7.79	61.4%	83.4%	12.27	10.16	44.2%	67.5%	16.25	14.25	31.3%	50.0%
ICP	10.52	7.71	52.5%	76.9%	14.50	11.75	27.5%	48.8%	19.80	17.71	13.8%	27.3%
ICP-PSO, $k = 1$	5.98	3.37	87.7%	95.7%	8.93	5.79	72.2%	85.4%	13.95	10.81	54.2%	68.4%
ICP-PSO, $k = 2$	5.65	3.07	89.7%	96.6%	8.39	5.00	74.4%	88.6%	12.75	9.23	58.0%	73.6%
ICP-PSO, $k = 4$	**5.53**	**2.91**	**90.8%**	**97.9%**	**7.93**	**4.53**	**76.8%**	**90.2%**	**12.28**	**8.99**	**60.4%**	**74.2%**

Table 1. Averaged performance metrics of three methods on three difficulty levels, using real and synthetic depth maps. Note that k is the k-means parameter in ICP-PSO.

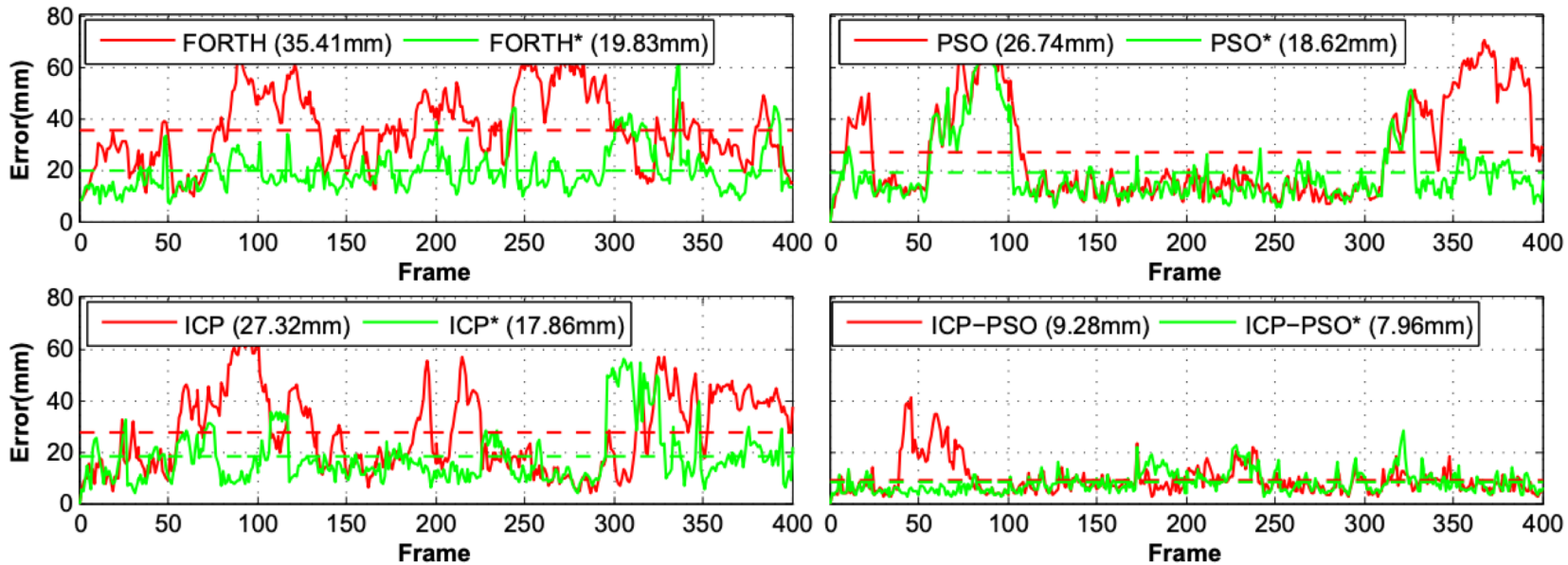


Figure 7. Average joint error in all frames of first subject. Each plot shows the results of one method using $(*)$ and not using initialization. The horizontal dotted lines are mean errors of each method over all frames.

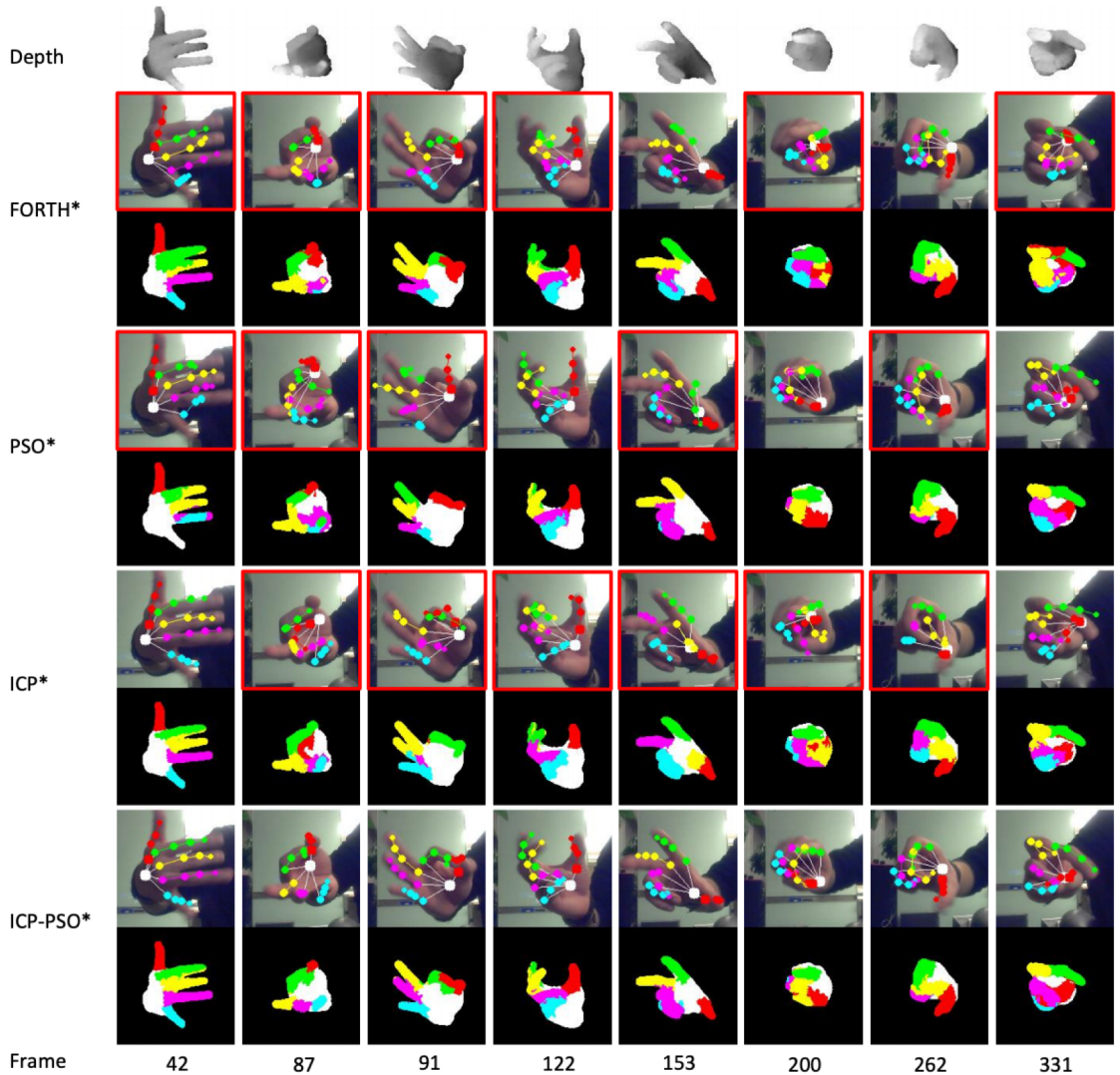


Figure 8. (better viewed in color) Example tracking results of the first subject. Those with red frames contain large errors. Color coded correspondence map of each result is also shown for better visualization.