# Efficient and Precise Interactive Hand Tracking Through Joint, Continuous Optimization of Pose and Correspondences

## Supplementary Materials


Jonathan Taylor Lucas Bordeaux* Thomas Cashman* Bob Corish* Cem Keskin* Toby Sharp*
Eduardo Soto*† David Sweeney* Julien Valentin*‡ Benjamin Luff§ Arran Topalian§ Erroll Wood¶
Sameh Khamis Pushmeet Kohli Shahram Izadi Richard Banks Andrew Fitzgibbon Jamie Shotton

Microsoft Research‖ McMaster University† University of Oxford‡ University of Abertay§ University of Cambridge¶


In this supplementary material we include additional details about our specific implementation of the described method.

## 1 Computational Breakdown

Fig. 1 gives an indication of the relative cost of different parts of our tracking algorithm. These timings are naturally variable and depend heavily on the machine being tested. Fig. 1 shows averages derived from running our tracker on a Dell Precision T7610 workstation with dual Intel® Xeon® E5-2680 CPUs. See our paper for evidence that the tracker is robust to reductions in the number of iterations or the number of starting points.

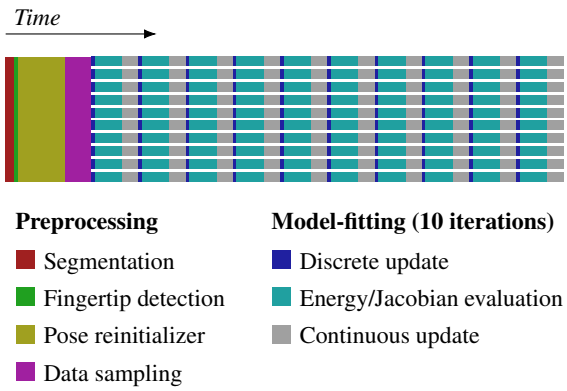


**Figure 1:** *Approximate proportion of wall-clock time spent on each part of our algorithm to process a single frame. This example shows 10 starting points running in parallel for 10 iterations. Proportions given here are averages; in practice the optimizer processes each starting point at slightly different rates.*

## 2 Background Term Positions

Our background term $E_{\text{bg}}$ requires the definition of a set of surface coordinates $\mathcal{U}^{\text{bg}} = \{u_h^{\text{bg}}\}_{h=1}^{H} \subseteq \Omega$ which, when evaluated on the surface of our model, are penalized for projecting outside the hand segmentation. These points were chosen heuristically to roughly cover the model, with a denser sampling on the fingers (see Fig. 2).

## 3 Pose Prior Training Data

The pose prior was obtained from Tan et al. [2016]. It was trained by fitting a multivariate Gaussian distribution to

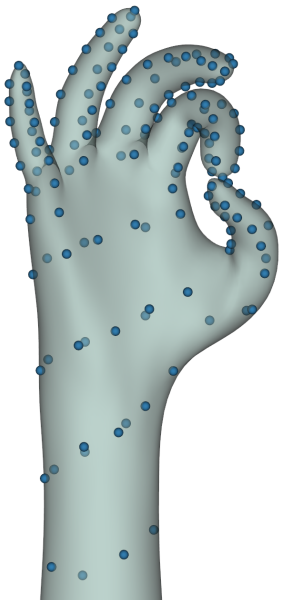

**Figure 2:** *The $H = 146$ positions $\mathcal{U}^{bg}$, shown as the positions $\{S(u_h^{bg}; \theta)\}_{h=1}^{H}$ evaluated on the model surface.*

- 104 hand poses obtained by interactively selecting frames from a live camera, for which the tracker of Sharp et al. [2015] inferred poses that the authors deemed sufficiently plausible and diverse.
- 38 hand-crafted poses representing specific gestures such as pinching, closed fists and pointing, for which the tracker of Sharp et al. [2015] was less accurate.

The same set of poses was used to determine the joint limit constraints $\psi^{\text{min}}$ and $\psi^{\text{max}}$ used in the $E_{\text{limit}}$ term of our energy.

## 4 Sphere Definitions for $E_{\text{int}}$

Our self-intersection term requires the definition of a set of spheres that approximate the volume of the fingers of our model, so that their intersection can be penalized. The spheres we used were positioned and sized heuristically (see Fig. 3). For spheres located in the same rigid portion of a finger (same color and shade in Fig. 3), any intersection does not indicate self-intersection of the model surface, so these pairs of spheres are excluded from the tested set of pairs $\mathcal{P}$. Similarly, our term $E_{\text{limit}}$ already penalizes highly unlikely poses where adjacent rigid portions of a finger self-intersect, so pairs of spheres in this configuration (same color and adjacent shade in Fig. 3) are also excluded from $\mathcal{P}$.

*Roughly equal contributions ordered alphabetically by last name.
‖All authors were at Microsoft for the majority of the work.

## 5 Discrete Update Proposal Positions

Our discrete update step requires a set of proposal coordinates $\mathcal{U}_i^{\text{prop}}$ to be considered in each iteration $i$. For each such set, we simply randomly sampled $Q = 259$ such coordinates (see Fig. 4).

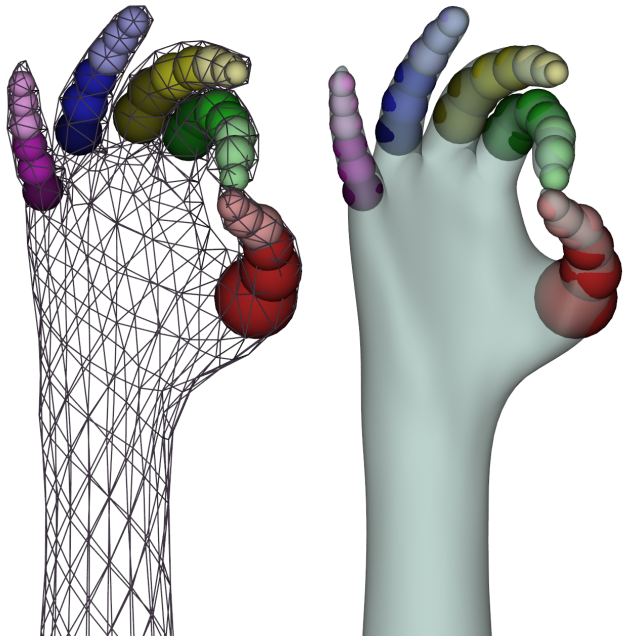

**Figure 3:** *The $S = 32$ spheres with radii $\{r_s\}_{s=1}^S$ and locations $\{c_s(\theta)\}_{s=1}^S$. $\mathcal{P}$ consists of the $|\mathcal{P}| = 427$ pairs where the spheres are shown above in different colours or non-adjacent shades of the same colour.*

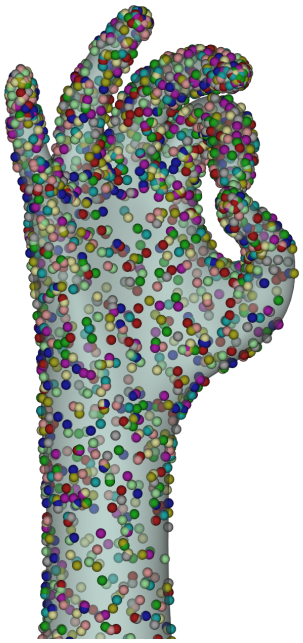

**Figure 4:** *Discrete update proposals $\mathcal{U}_i^{prop}$ for $i = 1, .., 10$. The $Q = 259$ proposals in each set are displayed as the points $\{S(u_{iq}^{prop}; \theta)\}_{q=1}^Q$, using a different colour for each $i$.*

## References


SHARP, T., KESKIN, C., ROBERTSON, D., TAYLOR, J., SHOTTON, J., KIM, D., RHEMANN, C., LEICHTER, I., VINNIKOV, A., WEI, Y., FREEDMAN, D., KOHLI, P., KRUPKA, E., FITZGIBBON, A., AND IZADI, S. 2015. Accurate, robust, and flexible realtime hand tracking. In *Proc. CHI*, 3633–3642.

TAN, D. J., CASHMAN, T., TAYLOR, J., FITZGIBBON, A., TARLOW, D., KHAMIS, S., IZADI, S., AND SHOTTON, J. 2016. Fits like a glove: Rapid and reliable hand shape personalization. In *Proc. CVPR*.