Motion Capture of Hands in Action using Discriminative Salient Points — Supplementary Material —

1 Implementation details (the distance field $\Psi_{f_s \to f_t}$):

The distance field $\Psi_{f_s \to f_t}$ is defined as a scalar filed $\mathbb{R}^3 \to \mathbb{R}^+$ shaped like a cone circumscribing face f_s in one of its circular sections, see Figure 2(left) in the paper. The further a vertex v is inside this cone, the more it gets penalized. The specific formulation of $\Psi_{f_s \to f_t}$, used in our implementation, is the following:

$$\Psi_{f_s \to f_t}(v) = \begin{cases} \left|(1 - \Phi(v)) \;\; \Upsilon(n \cdot (v - o))\right|^2 & \Phi(v) < 1 \\ 0 & \Phi(v) \geqslant 1 \end{cases} \tag{1}$$

where $n \in \mathbb{R}^3$ and $o \in \mathbb{R}^3$ are respectively the normal and the baricenter of the face f_s and the function $\Phi : \mathbb{R}^3 \to \mathbb{R}$ is defined as follows

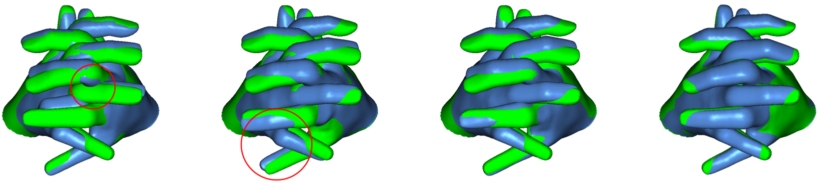
$$\Phi(v) = \frac{\|(v - o) - (n \cdot (v - o))\, n\|}{-\frac{r}{\sigma}(n \cdot (v - o)) + r} \tag{2}$$

where r is the radius of the circle centered at o and circumscribing f_s, and σ is a parameter defining the FOV of the cone. In all our experiments σ was set to 0.5. The scalar function $\Upsilon(x)$ defines how much the penalty increases the further a vertex is inside the cone and it is defined as follows

$$\Upsilon(x) = \begin{cases} -x + 1 - \sigma & x \leqslant -\sigma \\ -\frac{1-2\sigma}{4\sigma^2}x^2 - \frac{1}{2\sigma}x + \frac{1}{4}(3 - 2\sigma) & -\sigma \leqslant x \leqslant \sigma \\ 0 & x \geqslant \sigma \end{cases} \tag{3}$$

2 Experimental details:

In Figure 1 we present a visual explanation for the numbers reported on Table 1 in the paper. The aim was to evaluate the algorithm performance with respect to the used visual information, and this was done on synthetic data so that a groundtruth is available for comparison.



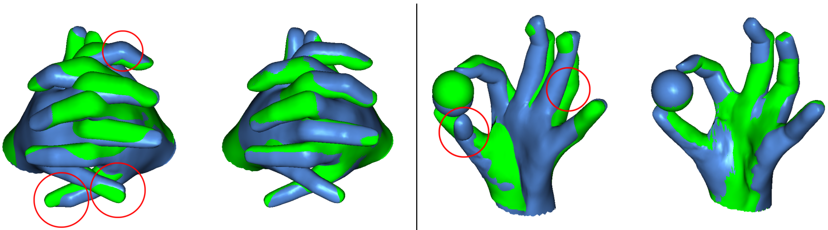
LO + $\mathcal{E}$ *Synthetic Seq. #1* (*Frame #49*)

LO + $\mathcal{EC}$ *Synthetic Seq. #1* (*Frame #49*)

LO + $\mathcal{ECO}$ *Synthetic Seq. #1* (*Frame #49*)

LO + $\mathcal{ECOS}$ *Synthetic Seq. #1* (*Frame #49*)

Fig. 1. Evaluation of the algorithm performance with respect to the used visual information (refer to the syntectic data experiment reported on Table 1 in the paper). The estimated model (blue) is overlaid on top of the groundtruth model (green). Larger discrepancies indicate larger errors. It is noticeable that, edges alone are very informative, but they allow self intersections (LO + $\mathcal{E}$). The collision term copes for these issues (LO + $\mathcal{EC}$). Optical flow improves a little on the algorithm accuracy, however it implicity contributes towards temporal smoothness of the motion (LO + $\mathcal{ECO}$). In the end, the salient points significantly increase the accuracy of the estimated pose. (LO + $\mathcal{ECOS}$).



HOPE128 + $\mathcal{ECOS}$ *Synthetic Seq. #1* (*Frame #34*) | LO + $\mathcal{ECOS}$ *Synthetic Seq. #1* (*Frame #34*) | HOPE128 + $\mathcal{ECOS}$ *Synthetic Seq. #2* (*Frame #84*) | LO + $\mathcal{ECOS}$ *Synthetic Seq. #2* (*Frame #84*)

Fig. 2. Comparison of the algorithm performance with respect to our implementation of HOPE + $\mathcal{ECOS}$. HOPE performs decently, but its results are typically noisier than ours due to the sampling/evolutionary nature of its optimization technique (i.e., the PSO).

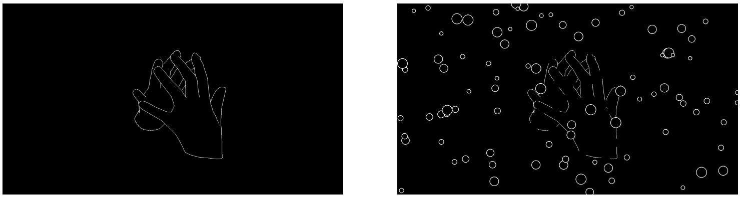


Fig. 3. Example of structured noise added and subtracted to the edge images to simulate edge detection errors. This is done in order to make the experiments on synthetic data as similar as possible to a realistic scenario.