Real-time Hand Tracking Using a Sum of Anisotropic Gaussians Model

Srinath Sridhar[1], Helge Rhodin[1], Hans-Peter Seidel[1], Antti Oulasvirta[2], Christian Theobalt[1]

[1]Max Planck Institute for Informatics
Saarbrücken, Germany
{ssridhar,hrhodin,hpseidel,theobalt}@mpi-inf.mpg.de

[2]Aalto University
Helsinki, Finland
antti.oulasvirta@aalto.fi

Supplementary Material, 3DV 2014

Introduction

In our paper, we introduce a new hand model consisting of 3D *Sum of Anisotropic Gaussians* (SAG), and a skeleton for motion tracking of hands in a multi-camera setup. In this supplementary document we explain in more detail, and generality, the following parts of our main paper.

Section 1	Projection of 3D SAG to 2D SAG (c.f. Section 4.2 in main paper)
Section 2	Computation of derivatives for our pose fitting energy (c.f. Section 4.3 in main paper)
Section 3	Experiments in addition to the ones in the paper (c.f. Section 5 in main paper)
Section 4	Qualitative comparison of our method with that of [5] (c.f. Section 5 in main paper)

Please refer to the main paper for important definitions and concepts such as SAGs, similarity measure, our energy formulation, and the relationship between Gaussians and ellipsoids.

1. Projection of Ellipsoids (c.f. Section 4.2 of main paper)

The perspective projection of an ellipsoid is an ellipse defined by the intersection of the elliptical cone, formed by the rays originating from camera center and tangential to the ellipsoid, with the image plane (see Figure 4 of the main paper). The projection equation is best explained in four separate steps.

World-to-Camera Transformation: The extrinsic camera parameters are the orientation $\mathbf{R}_{wc}$ and position $\mathbf{c}$ of the camera. They transform the ellipsoid $(\boldsymbol{\Sigma}_h, \boldsymbol{\mu}_h)$ to the camera coordinate system by

$$\begin{aligned} \boldsymbol{\Sigma}_c &= \mathbf{R}_{wc}\, \boldsymbol{\Sigma}_h\, \mathbf{R}_{wc}^T \\ \boldsymbol{\mu}_c &= \mathbf{R}_{wc}(\boldsymbol{\mu}_h - \mathbf{c}), \end{aligned} \tag{1}$$

such that the origin is at the camera center and the z direction is aligned with the camera view direction.

Construction of Elliptical Cone: We are interested in a mathematical expression for the elliptical cone that is formed by the rays originating at the camera center $\mathbf{c}$ and tangential to the ellipsoid (see Figure 4 of the main paper). According to [2] all points on this cone satisfy

$$\mathbf{x}^\top \mathbf{M} \mathbf{x} = 0, \tag{2}$$

where the *cone matrix* $\mathbf{M}$ is

$$\mathbf{M} = \boldsymbol{\Sigma}_c^{-1} \left(\boldsymbol{\mu}_c - \mathbf{c}\right) \boldsymbol{\mu}_c^\top \boldsymbol{\Sigma}_c^{-1} - \left(\boldsymbol{\mu}_c^\top \boldsymbol{\Sigma}_c^{-1} \boldsymbol{\mu}_c - 1\right) \boldsymbol{\Sigma}_c^{-1}. \tag{3}$$

Intersection of the Elliptical Cone with the Image Plane: The points that form the projected ellipsoid on the canonical image plane I are those points that satisfy both equation 2 and the image plane equation (see Figure 4 of the main paper). For a canonical image plane, the image plane equation is $z = 1$. We can derive an expression for the intersection of I and equation 2 as follows.

The second degree polynomial representation of a conic section is given as [1]

$$px^2 + qxy + ry^2 + sx + ty + u = 0,$$

where $\mathbf{x} = [x, y, 1]^T$. The above equation is equivalent to Equation 2 where $\mathbf{M}$ can be written as

$$\mathbf{M} = \begin{bmatrix} p & q/2 & s/2 \\ q/2 & r & t/2 \\ s/2 & t/2 & u \end{bmatrix} = \begin{bmatrix} m_1 & m_2 & m_3 \\ m_2 & m_4 & m_5 \\ m_3 & m_5 & m_6 \end{bmatrix}. \tag{4}$$

Here m_k represent the elements of the symmetric matrix $\mathbf{M}$. Let $\mathbf{M}_{33}$ represent the 2×2 submatrix of $\mathbf{M}$ excluding the 3rd row and 3rd column. The canonical parameters of the ellipse are given by

$$\widetilde{\boldsymbol{\mu}}_p = \frac{1}{(4pr - q^2)} \begin{bmatrix} (qt - 2rs) \\ (sq - 2pt) \end{bmatrix} = \frac{1}{|\mathbf{M}_{33}|} \begin{bmatrix} |\mathbf{M}_{31}| \\ -|\mathbf{M}_{23}| \end{bmatrix}, \tag{5}$$

$$\widetilde{\boldsymbol{\Sigma}}_p = -\frac{|\mathbf{M}|}{|\mathbf{M}_{33}|} \mathbf{M}_{33}^{-1}. \tag{6}$$

For a general camera with intrinsics matrix $\mathbf{K}$ (as defined in [3]), the projected ellipse $(\widetilde{\boldsymbol{\Sigma}}_p, \widetilde{\boldsymbol{\mu}}_p)$ from the canonical image plane is transformed to a general image plane. The transformed ellipse parameters are

$$\begin{aligned} \boldsymbol{\mu}_p &= \mathbf{K}_{33} \widetilde{\boldsymbol{\mu}}_p + \begin{bmatrix} k_{13} \\ k_{23} \end{bmatrix}, \\ \boldsymbol{\Sigma}_p &= \mathbf{K}_{33}\, \widetilde{\boldsymbol{\Sigma}}_p\, \mathbf{K}_{33}^T. \end{aligned} \tag{7}$$

We utilize this ellipsoid projection formulation to project the 3D SAG model to a 2D image SAG.

2. Derivatives of Pose Fitting Energy $E_{sim}(\Theta)$ (c.f. Section 4.3 of main paper)

The benefit of using the SAG representation for the image and hand is that the pose fitting energy (measured by Gaussian overlap or similarity) with respect to all skeleton parameters of the hand are analytical. The analytical derivatives allow efficient and accurate optimization of the hand model by gradient based optimization methods.

The similarity term in the objective function we want to maximize (see Section 4.4 in main paper) is given by

$$E_{sim}\left[\mathcal{C}_I, \Pi(\mathcal{C}_H)\right] = \sum_{q \in \mathcal{C}_I} \min\left(\left(\sum_{p \in \Pi(\mathcal{C}_H)} w_p^h\, E_{pq} \right), E_{qq} \right). \tag{8}$$

We are interested in the total derivative with respect to each degree of freedom of the skeleton, θ_j. We then have

$$\frac{dE_{sim}}{d\theta_j} = \sum_{q \in \mathcal{C}_I} \delta_{ij} \left(\sum_{p \in \Pi(\mathcal{C}_H)} w_p^h\, d(\mathbf{c}_p, \mathbf{c}_q)\, \frac{dD_{pq}}{d\theta_j} \right). \tag{9}$$

where $E_{pq} = d(\mathbf{c}_p, \mathbf{c}_q)\, D_{pq}$, and

$$\delta_{ij} = \begin{cases} 0 & \text{if } \sum_{p \in \Pi(\mathcal{C}_H)} w_p^h\, E_{pq} > E_{qq} \\ 1 & \text{otherwise.} \end{cases} \tag{10}$$

We obtain the analytical derivatives of E_{sim} with respect to the degrees of freedom Θ in three steps. For each Gaussian pair (h, q) and parameter θ_j we compute

$$\boxed{\left(\frac{\partial \boldsymbol{\Sigma}_h}{\partial \theta_j}, \frac{\partial \boldsymbol{\mu}_h}{\partial \theta_j}\right) \xrightarrow{a)} \left(\frac{d\boldsymbol{M}}{d\theta_j}\right) \xrightarrow{b)} \left(\frac{d\boldsymbol{\Sigma}_p}{d\theta_j}, \frac{d\boldsymbol{\mu}_p}{d\theta_j}\right) \xrightarrow{c)} \left(\frac{dD_{pq}}{d\theta_j}\right).} \tag{11}$$

In the following sections, we show how the derivatives are obtained in turn starting from the last.

2.1. Derivative of D_{pq} with respect to Θ, $\frac{dD_{pq}}{d\theta_j}$

The term D_{pq} can be written as (also see equation 7 in main paper)

$$D_{pq} = h\, e^{-0.5\, g} \text{ with } h = \frac{\sqrt{(2\pi)^k |\boldsymbol{\Sigma}_p\, \boldsymbol{\Sigma}_q|}}{\sqrt{|(\boldsymbol{\Sigma}_p + \boldsymbol{\Sigma}_q)|}} \text{ and} \tag{12}$$

$$g = (\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)^T (\boldsymbol{\Sigma}_p + \boldsymbol{\Sigma}_q)^{-1} (\boldsymbol{\mu}_p - \boldsymbol{\mu}_q). \tag{13}$$

The total derivative with respect to Θ is

$$\boxed{\frac{dD_{pq}}{d\theta_j} = \sum_i \frac{\partial D_{pq}}{\partial \boldsymbol{\mu}_{p_i}} \frac{d\boldsymbol{\mu}_{p_i}}{d\theta_j} + \sum_{k \in \{1,2,3\}} \frac{dD_{pq}}{d\boldsymbol{\Sigma}_{p_k}} \frac{d\boldsymbol{\Sigma}_{p_k}}{d\theta_j},} \tag{14}$$

where k iterates through the unique elements of the symmetric matrix $\boldsymbol{\Sigma}_p$ and

$$\begin{aligned} \frac{\partial D_{pq}}{\partial \boldsymbol{\mu}_p} &= -0.5\, h\, e^{-0.5g} \frac{\partial g}{\partial \boldsymbol{\mu}_p}, \\ \frac{dD_{pq}}{d\boldsymbol{\Sigma}_p} &= \frac{dh}{d\boldsymbol{\Sigma}_p} e^{-0.5g} - 0.5\, h\, e^{-0.5g} \frac{dg}{d\boldsymbol{\Sigma}_p} \\ &= \left(\frac{dh}{d\boldsymbol{\Sigma}_p} - 0.5\, h\, \frac{dg}{d\boldsymbol{\Sigma}_p}\right) e^{-0.5g}, \end{aligned} \tag{15}$$

and the partial derivative of subterm g with respect to $\boldsymbol{\mu}_p$ is

$$\begin{aligned} \frac{\partial g}{\partial \boldsymbol{\mu}_p} &= \frac{\partial (\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)^T (\boldsymbol{\Sigma}_p + \boldsymbol{\Sigma}_q)^{-1} (\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)}{\partial \boldsymbol{\mu}_p} \\ &= \left((\boldsymbol{\Sigma}_p + \boldsymbol{\Sigma}_q)^{-1} + (\boldsymbol{\Sigma}_p + \boldsymbol{\Sigma}_q)^{-T}\right) (\boldsymbol{\mu}_p - \boldsymbol{\mu}_q) \\ &= 2(\boldsymbol{\Sigma}_p + \boldsymbol{\Sigma}_q)^{-1} (\boldsymbol{\mu}_p - \boldsymbol{\mu}_q). \end{aligned} \tag{16}$$

By considering the symmetry of $\boldsymbol{\Sigma}_p$ we obtain

$$\begin{aligned} \frac{dg}{d\boldsymbol{\Sigma}_p} &= \frac{\partial g}{\partial \boldsymbol{\Sigma}_p} + \frac{\partial g}{\partial \boldsymbol{\Sigma}_p}^T - \operatorname{diag}\left(\frac{\partial g}{\partial \boldsymbol{\Sigma}_p}\right), \text{ where} \\ \frac{\partial g}{\partial \boldsymbol{\Sigma}_p} &= -(\boldsymbol{\Sigma}_p + \boldsymbol{\Sigma}_q)^{-1} (\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)\, (\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)^T (\boldsymbol{\Sigma}_p + \boldsymbol{\Sigma}_q)^{-1}, \end{aligned} \tag{17}$$

and

$$
\begin{aligned}
\frac{dh}{d\boldsymbol{\Sigma}_p} &= \frac{d}{d\boldsymbol{\Sigma}_p}\frac{\sqrt{(2\pi)^2|\boldsymbol{\Sigma}_p\,\boldsymbol{\Sigma}_q|}}{\sqrt{|(\boldsymbol{\Sigma}_p+\boldsymbol{\Sigma}_q)|}} \\
&= \sqrt{(2\pi)^2|\boldsymbol{\Sigma}_q|}\frac{\frac{d\sqrt{|\boldsymbol{\Sigma}_p|}}{d\boldsymbol{\Sigma}_p}\sqrt{|(\boldsymbol{\Sigma}_p+\boldsymbol{\Sigma}_q)|}-\frac{d\sqrt{|(\boldsymbol{\Sigma}_p+\boldsymbol{\Sigma}_q)|}}{d\boldsymbol{\Sigma}_p}\sqrt{|\boldsymbol{\Sigma}_p|}}{|(\boldsymbol{\Sigma}_p+\boldsymbol{\Sigma}_q)|} \\
&= 0.5\sqrt{(2\pi)^2|\boldsymbol{\Sigma}_q|}\frac{\frac{1}{\sqrt{|\boldsymbol{\Sigma}_p|}}\frac{d|\boldsymbol{\Sigma}_p|}{d\boldsymbol{\Sigma}_p}\sqrt{|(\boldsymbol{\Sigma}_p+\boldsymbol{\Sigma}_q)|}-\frac{1}{\sqrt{|(\boldsymbol{\Sigma}_p+\boldsymbol{\Sigma}_q)|}}\frac{d|(\boldsymbol{\Sigma}_p+\boldsymbol{\Sigma}_q)|}{d\boldsymbol{\Sigma}_p}\sqrt{|\boldsymbol{\Sigma}_p|}}{|(\boldsymbol{\Sigma}_p+\boldsymbol{\Sigma}_q)|}
\end{aligned}
\tag{18}
$$

where

$$
\begin{aligned}
\frac{d|\boldsymbol{\Sigma}_p|}{d\boldsymbol{\Sigma}_p} &= |\boldsymbol{\Sigma}_p|\left(2\boldsymbol{\Sigma}_p^{-1}-\operatorname{diag}\left(\boldsymbol{\Sigma}_p^{-1}\right)\right) \\
\frac{d|(\boldsymbol{\Sigma}_p+\boldsymbol{\Sigma}_q)|}{d\boldsymbol{\Sigma}_p} &= |\boldsymbol{\Sigma}_p+\boldsymbol{\Sigma}_q|\left(2(\boldsymbol{\Sigma}_p+\boldsymbol{\Sigma}_q)^{-1}-\operatorname{diag}\left((\boldsymbol{\Sigma}_p+\boldsymbol{\Sigma}_q)^{-1}\right)\right)
\end{aligned}
\tag{19}
$$

and therefore

$$
\frac{dh}{d\boldsymbol{\Sigma}_p} = 0.5\frac{\sqrt{(2\pi)^2|\boldsymbol{\Sigma}_q|}}{|(\boldsymbol{\Sigma}_p+\boldsymbol{\Sigma}_q)|} \tag{20}
$$

$$
\begin{pmatrix} \sqrt{|\boldsymbol{\Sigma}_p|}\left(2\boldsymbol{\Sigma}_p^{-1}-\operatorname{diag}\left(\boldsymbol{\Sigma}_p^{-1}\right)\right)\sqrt{|(\boldsymbol{\Sigma}_p+\boldsymbol{\Sigma}_q)|} \\ -\sqrt{|(\boldsymbol{\Sigma}_p+\boldsymbol{\Sigma}_q)|}\left(2(\boldsymbol{\Sigma}_p+\boldsymbol{\Sigma}_q)^{-1}-\operatorname{diag}\left((\boldsymbol{\Sigma}_p+\boldsymbol{\Sigma}_q)^{-1}\right)\sqrt{|\boldsymbol{\Sigma}_p|}\right. \end{pmatrix}
$$

$$
= 0.5\frac{\sqrt{(2\pi)^2|\boldsymbol{\Sigma}_p\,\boldsymbol{\Sigma}_q|}}{\sqrt{|(\boldsymbol{\Sigma}_p+\boldsymbol{\Sigma}_q)|}} \tag{21}
$$

$$
\left(\left(2\boldsymbol{\Sigma}_p^{-1}-\operatorname{diag}\left(\boldsymbol{\Sigma}_p^{-1}\right)\right)-\left(2(\boldsymbol{\Sigma}_p+\boldsymbol{\Sigma}_q)^{-1}-\operatorname{diag}\left((\boldsymbol{\Sigma}_p+\boldsymbol{\Sigma}_q)^{-1}\right)\right)\right.
$$

$$
= 0.5h\left(\left(2\boldsymbol{\Sigma}_p^{-1}-\operatorname{diag}\left(\boldsymbol{\Sigma}_p^{-1}\right)\right)-\left(2(\boldsymbol{\Sigma}_p+\boldsymbol{\Sigma}_q)^{-1}-\operatorname{diag}\left((\boldsymbol{\Sigma}_p+\boldsymbol{\Sigma}_q)^{-1}\right)\right)\right. \tag{22}
$$

2.2. Derivative of Projected Ellipse Parameters with respect to Θ, $\left(\frac{d\boldsymbol{\Sigma}_p}{d\theta_j}, \frac{d\boldsymbol{\mu}_p}{d\theta_j}\right)$

We formulate the parameters of the projected ellipse $\boldsymbol{\Sigma}_p, \boldsymbol{\mu}_p$ completely in terms of the cone matrix, $\mathbf{M}$. Thus, the total derivatives of $\boldsymbol{\Sigma}_p, \boldsymbol{\mu}_p$ with respect to Θ are given by

$$
\boxed{\begin{aligned}
\frac{d\boldsymbol{\mu}_p}{d\theta_j} &= \sum_k \frac{\partial\boldsymbol{\mu}_p}{\partial m_k}\frac{dm_k}{d\theta_j}, \\
\frac{d\boldsymbol{\Sigma}_p}{d\theta_j} &= \sum_k \frac{\partial\boldsymbol{\Sigma}_p}{\partial m_k}\frac{dm_k}{d\theta_j}.
\end{aligned}} \tag{23}
$$

In the next section we describe the partial derivatives.

2.3. Derivative of the Cone Matrix M with respect to Θ, $\frac{d\boldsymbol{M}}{d\theta_j}$

The partial derivatives from equation 23 are given by

$$\frac{\partial \boldsymbol{\mu}_p}{\partial m_{ij}} = \frac{1}{|\mathbf{M}_{33}|}\begin{bmatrix} |\mathbf{M}_{31}|\left[\mathrm{Tr}(\mathbf{M}_{31}^{-1}\mathbf{S}_{31}^{ij}) - \mathrm{Tr}(\mathbf{M}_{33}^{-1}\mathbf{S}_{33}^{ij})\right] \\ -|\mathbf{M}_{23}|\left[\mathrm{Tr}(\mathbf{M}_{23}^{-1}\mathbf{S}_{23}^{ij}) - \mathrm{Tr}(\mathbf{M}_{33}^{-1}\mathbf{S}_{33}^{ij})\right] \end{bmatrix},$$

$$\frac{\partial \boldsymbol{\Sigma}_p}{\partial m_{ij}} = \frac{|\mathbf{M}|}{|\mathbf{M}_{33}|}\left[\mathbf{M}_{33}^{-1}\mathbf{S}_{33}^{ij\,T}\mathbf{M}_{33}^{-1} - \mathbf{M}_{33}^{-1}\{\mathrm{Tr}(\mathbf{M}^{-1}\mathbf{S}^{ij}) + \mathrm{Tr}(\mathbf{M}_{33}^{-1}\mathbf{S}_{33}^{ij})\}\right]. \tag{24}$$

In the above expression, $\mathbf{S}^{ij}$ is the *structure matrix* for symmetric matrices that has elements i, j and j, i equal to one and is zero otherwise. Please see the Matrix Cookbook for more details [4].

The next step is to obtain the derivatives of the cone matrix M with respect to the elements $\boldsymbol{\mu}_{ci}$ and ${\boldsymbol{\Sigma}_c^{-1}}_{ij}$. Here $\boldsymbol{\mu}_c$ and $\boldsymbol{\Sigma}_c$ denote the parameters of an ellipsoid in the camera's local coordinate system. Please note that this is distinct from the world coordinate system. We obtain for the derivative with respect to the mean position

$$\begin{aligned}\frac{\partial \mathbf{M}}{\partial \boldsymbol{\mu}_{ci}} &= \boldsymbol{\Sigma}_c^{-1}\frac{\partial \boldsymbol{\mu}_c\boldsymbol{\mu}_c^T}{\partial \boldsymbol{\mu}_{ci}}\boldsymbol{\Sigma}_c^{-1} - \frac{\partial \boldsymbol{\mu}_c^T\boldsymbol{\Sigma}_c^{-1}\boldsymbol{\mu}_c}{\partial \boldsymbol{\mu}_{ci}}\boldsymbol{\Sigma}_c^{-1} \\ &= \boldsymbol{\Sigma}_c^{-1}(\mathbf{e}^i\boldsymbol{\mu}_c^T + \mathbf{e}^i\boldsymbol{\mu}_c^T)^T\boldsymbol{\Sigma}_c^{-1} - \left((\mathbf{e}^{i^T}\boldsymbol{\Sigma}_c^{-1}\boldsymbol{\mu}_c) + (\mathbf{e}^{i^T}\boldsymbol{\Sigma}_c^{-1}\boldsymbol{\mu}_c)^T\right)\boldsymbol{\Sigma}_c^{-1},\end{aligned} \tag{25}$$

where $\mathbf{e}^i$ is the standard basis vector with the ith entry set to one and zero otherwise. And similarly for the derivative with respect to $\boldsymbol{\Sigma}_c$

$$\begin{aligned}\frac{\partial \mathbf{M}}{\partial {\boldsymbol{\Sigma}_{c_k}}^{-1}} &= \frac{\boldsymbol{\Sigma}_c^{-1}\boldsymbol{\mu}_c\boldsymbol{\mu}_c^T\boldsymbol{\Sigma}_c^{-1}}{\partial {\boldsymbol{\Sigma}_{c_k}}^{-1}} - \left(\frac{\partial \boldsymbol{\mu}_c^T\boldsymbol{\Sigma}_c^{-1}\boldsymbol{\mu}_c}{\partial {\boldsymbol{\Sigma}_{c_k}}^{-1}}\boldsymbol{\Sigma}_c^{-1} + (\boldsymbol{\mu}_c^T\boldsymbol{\Sigma}_c^{-1}\boldsymbol{\mu}_c - 1)\frac{\partial \boldsymbol{\Sigma}_c^{-1}}{\partial {\boldsymbol{\Sigma}_{c_k}}^{-1}}\right) \\ &= \left(\mathbf{S}^k\boldsymbol{\mu}_c\boldsymbol{\mu}_c^T\boldsymbol{\Sigma}_c^{-T}\right) + \left(\mathbf{S}^k\boldsymbol{\mu}_c\boldsymbol{\mu}_c^T\boldsymbol{\Sigma}_c^{-T}\right)^T \qquad (26) \\ &\quad - {\boldsymbol{\mu}_c}^T\mathbf{S}^k\boldsymbol{\mu}_c\boldsymbol{\Sigma}_c^{-1} - (\boldsymbol{\mu}_c^T\boldsymbol{\Sigma}_c^{-1}\boldsymbol{\mu}_c - 1)\mathbf{S}^k, \qquad (27)\end{aligned}$$

where ${\boldsymbol{\Sigma}_c}^{-1}{}_k$ is the kth element of the upper half of ${\boldsymbol{\Sigma}_c}^{-1}$ and $\mathbf{S}^k$ is the corresponding structure matrix. The partial derivatives are

$$\begin{aligned}\frac{\partial \mathbf{M}}{\partial \boldsymbol{\mu}_{ci}} &= \boldsymbol{\Sigma}_c^{-1}(\mathbf{e}^i\boldsymbol{\mu}_c^T + \mathbf{e}^i\boldsymbol{\mu}_c^T)^T\boldsymbol{\Sigma}_c^{-1} - \left((\mathbf{e}^{i^T}\boldsymbol{\Sigma}_c^{-1}\boldsymbol{\mu}_c) + (\mathbf{e}^{i^T}\boldsymbol{\Sigma}_c^{-1}\boldsymbol{\mu}_c)^T\right)\boldsymbol{\Sigma}_c^{-1}, \\ \frac{\partial \mathbf{M}}{\partial {\boldsymbol{\Sigma}_{c_k}}^{-1}} &= \left(\mathbf{S}^k\boldsymbol{\mu}_c\boldsymbol{\mu}_c^T\boldsymbol{\Sigma}_c^{-T}\right) + \left(\mathbf{S}^k\boldsymbol{\mu}_c\boldsymbol{\mu}_c^T\boldsymbol{\Sigma}_c^{-T}\right)^T \qquad (28) \\ &\quad - {\boldsymbol{\mu}_c}^T\mathbf{S}^k\boldsymbol{\mu}_c\boldsymbol{\Sigma}_c^{-1} - (\boldsymbol{\mu}_c^T\boldsymbol{\Sigma}_c^{-1}\boldsymbol{\mu}_c - 1)\mathbf{S}^k \qquad (29)\end{aligned}$$

The total derivative is then

$$\boxed{\frac{d\mathbf{M}}{d\theta_j} = \sum_i \frac{\partial \mathbf{M}}{\partial \boldsymbol{\mu}_{ci}}\frac{d\boldsymbol{\mu}_{ci}}{d\theta_j} + \sum_{k\in\{1,2,3\}} \frac{\partial \mathbf{M}}{\partial {\boldsymbol{\Sigma}_{c_k}}^{-1}}\frac{d{\boldsymbol{\Sigma}_{c_k}}^{-1}}{d\theta_j}.} \tag{30}$$

For a general camera position and orientation we also need to take the extrinsic camera parameters into account. As the extrinsic parameters do not depend on the DOFs θ_j their derivatives are simply

$$\frac{d\boldsymbol{\mu}_c}{d\theta_j} = \mathbf{R}_{wc}\frac{\partial \boldsymbol{\mu}_h}{\partial \theta_j} \tag{31}$$

$$\frac{d\boldsymbol{\Sigma}_c^{-1}}{d\theta_j} = \mathbf{R}_{wc}\frac{\partial \boldsymbol{\Sigma}_h}{\partial \theta_j}\mathbf{R}_{wc}^T, \tag{32}$$

where $\mathbf{R}_{wc}^{-1}$ is the camera orientation, $\boldsymbol{\mu}_h$ the ellipsoid center, and $\boldsymbol{\Sigma}_h$ its covariance matrix in the world coordinate system.

2.4. Derivative of 3D Ellipsoid Parameters with respect to Joint Rotation α

The hand model is fitted to the observed images by modifying the degrees of freedom θ_j of the skeleton. In the case that θ_j specifies a rotational DOF *i.e.* the rotation around a joint axis $\mathbf{u}$ by α, its transformation according to Rodrigues formula is given by

$$\mathbf{R} = \mathbf{I}\cos(\alpha) + \sin(\alpha)[\mathbf{u}]_\times + (1-\cos(\alpha))\,\mathbf{u}\times\mathbf{u}, \tag{33}$$

where $[\mathbf{u}]_\times$ is the cross product matrix

$$[\mathbf{u}]_\times = \begin{bmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{bmatrix}. \tag{34}$$

The rotation $\mathbf{R}$ transforms the Ellipsoid covariance matrix $\boldsymbol{\Sigma}$ and its position $\boldsymbol{\mu}$ by

$$\begin{aligned} \boldsymbol{\Sigma}' &= \mathbf{R}\boldsymbol{\Sigma}\mathbf{R}^\top \\ \boldsymbol{\mu}' &= \boldsymbol{\mu}_j + \mathbf{R}(\boldsymbol{\mu}-\boldsymbol{\mu}_j) \end{aligned} \tag{35}$$

where $\boldsymbol{\mu}_j$ is the joint location. From equation 33 it follows that the derivative of $\mathbf{R}$ with respect to α at $\alpha = 0$ is

$$\left.\frac{\partial \mathbf{R}}{\partial \alpha}\right|_{\alpha=0} = [\mathbf{u}]_\times. \tag{36}$$

The partial derivatives of $\boldsymbol{\Sigma}'^{-1}, \boldsymbol{\mu}'$ with respect to the entries r_{ij} of $\mathbf{R}$ are

$$\frac{\partial \boldsymbol{\mu}'}{\partial r_{i,j}} = \mathbf{S}^{ij}(\boldsymbol{\mu}-\boldsymbol{\mu}_j) \tag{37}$$

$$\frac{\partial \boldsymbol{\Sigma}'^{-1}}{\partial r_{i,j}} = \left(RT(i,j) + RT(i,j)^\top\right), \text{ where} \tag{38}$$

$$RT(i,j) = \mathbf{R}\boldsymbol{\Sigma}'^{-1}\mathbf{S}^{ij} \tag{39}$$

and $\mathbf{S}^{ij}$ is the structure matrix with the i-,j-th entry equal to one and zero otherwise. Therefore the total derivative is

$$\frac{d\boldsymbol{\mu}'}{d\alpha} = \frac{\partial \mathbf{R}}{\partial \alpha}(\boldsymbol{\mu}-\boldsymbol{\mu}_j) \tag{40}$$

$$\frac{d\boldsymbol{\Sigma}'^{-1}}{d\alpha} = \sum_{i,j\in\{1,2,3\}} \left(RT(i,j) + RT(i,j)^\top\right)\frac{\partial r_{ij}}{\partial \alpha}. \tag{41}$$

In the case that θ_j specifies the translation of the whole hand model by $\delta\boldsymbol{\mu}_j$, θ only affects the ellipsoid position by

$$\boldsymbol{\mu}' = \boldsymbol{\mu} + \delta\boldsymbol{\mu}_j, \tag{42}$$

and the covariance matrix remains unchanged. Therefore, the total derivatives with respect to $\delta\boldsymbol{\mu}_j$ are

$$\boxed{\begin{aligned} \frac{d\boldsymbol{\mu}'}{d\delta\theta_j} &= \frac{d\delta\boldsymbol{\mu}_j}{d\delta\theta_j} \\ \frac{d\boldsymbol{\Sigma}'}{d\delta\theta_j} &= 0. \end{aligned}} \tag{43}$$

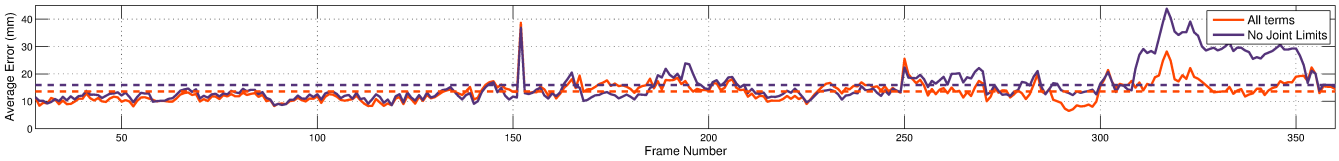

Figure 1. Average error with our SAG-based full pose fitting energy (orange) and with the joint limits term disabled (purple). The full pose fitting energy with all terms produces the least error. Best viewed in color.

3. Additional Experiments

Effect of Joint Limits Term: Our SAG-based pose fitting energy consists of two terms as given in Equation 9 in the main paper. We did an additional experiment to assess the effect of disabling the joint limits term. Figure 1 shows the average error over the `adbadd` sequence.

Dataset: For all experiments and results, we used the publicly available Dexter 1 dataset as well as motions that we recorded. These sequences consist of the following types of motions.

- Abduction and adduction (`adbadd`)
- Finger flexion and extension (`flexex1`)
- Global hand motion (all sequences)
- Global hand rotation (all sequences)
- Fast, slow and subtle movements (real-time sequences)

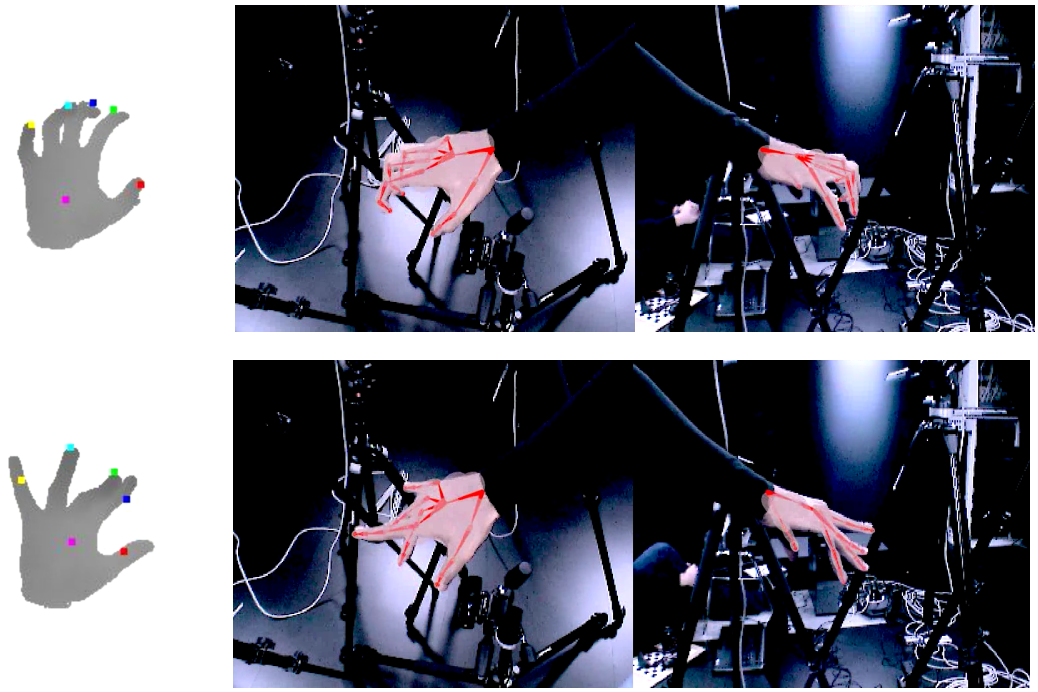

Figure 2. Qualitative comparison of our method (right) with that of [5] (left, fingertip positions only). Our fingertip localization is much better because we use multiple RGB images which helps resolve ambiguities. Note that [5] perform frame-by-frame pose estimation on **only depth data** without any temporal tracking.

4. Qualitative Comparisons

Figure 2 shows qualitative comparison of our method with that of [5]. Note that their method works using only depth while ours uses multi-view RGB. However, this comparison is valid because we use a publicly available calibrated dataset consisting of both depth and RGB data.

References

[1] A. B. Ayoub. The central conic sections revisited. *Mathematics Magazine*, 66(5):322–325, Dec. 1993. 2

[2] D. Eberly. Perspective projection of an ellipsoid. http://www.geometrictools.com/, 1999. 1

[3] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*, volume 2. Cambridge Univ Press, 2000. 2

[4] K. B. Petersen and M. S. Pedersen. *The matrix cookbook*. Technical University of Denmark, 2006. 5

[5] D. Tang, H. J. Chang, A. Tejani, and T.-K. Kim. Latent regression forest: Structured estimation of 3d articulated hand posture. In *Proc. of CVPR 2014*. 1, 8