Supplementary Material for Self-Calibrated Multi-Sensor Wearable for Hand Tracking and Modeling

Nikhil Gosala*, Fangjinhua Wang*, Zhaopeng Cui, Hanxue Liang, Oliver Glauser, Shihao Wu, Olga Sorkine-Hornung, *Member, IEEE,*

1 Phalange angle computation

The alignment of the $\mathcal{H}_D$ with the $\mathcal{H}_G$ allows the use of the axes defined in [1] to compute the phalange angles from the regressed 3D joints. The flexion between phalanges, defined about the X-axis, is thus computed by projecting the phalanges onto the YZ plane, e.g., the green plane in Fig. 6 in the main paper, normalising them, and then computing the cosine inverse of their dot product. Likewise, abduction, defined about the Z-axis, is computed by first projecting the phalanges onto the XY plane, before computing the angle. Note that the thumb possesses more degrees of freedom than the other fingers and dynamically estimate more than one common plane.

$\mathcal{H}_G$ Pose reconstruction. Once the merged base joint 3D positions and the merged angles represented in $\mathcal{H}_D$ have been computed, they can be incorporated to reconstruct the final merged hand pose estimate represented in $\mathcal{H}_G$. The incorporation of the merged angles into the base pose requires the manipulation of the phalange vector connecting the base joint to the first joint - the base phalange - in each finger.

The component of the base phalange parallel to the plane of interest is rotated by an angle equal to the corresponding merged angle to obtain the plane-parallel component of the next phalange vector. This new plane-parallel component is then added to the common plane-perpendicular component to obtain the phalange vector of the middle phalange. This rotation of the base vector, followed by the addition of the common plane-perpendicular component to compute the updated phalange vector, is iteratively applied to all phalanges in the finger to obtain the updated phalange vectors.

These phalange vectors are then scaled to lengths that are either computed on the fly from the depth pose or estimated during a pre-calibration phase, and the 3D positions of the updated joints are computed using trivial vector algebra to obtain the merged pose estimate of the finger.

The pose reconstruction steps can be summarised as follows:

$$\mathbf{v}_{\text{rot}}^{\parallel} = \mathbf{v}_{\text{base}}^{\parallel}\cos\theta + (\mathbf{n}\times\mathbf{v}_{\text{base}}^{\parallel})\sin\theta + \mathbf{n}(\mathbf{n}\cdot\mathbf{v}_{\text{base}}^{\parallel})(1-\cos\theta) \tag{1}$$

$$\mathbf{v}_{\text{rot}} = \mathbf{v}_{\text{rot}}^{\parallel} + \mathbf{v}_{\text{merged}}^{\perp} \tag{2}$$

$$c_{\text{merged}}^{i+1} = c_{\text{merged}}^{i} + l * \frac{\mathbf{v}_{\text{rot}}}{\|\mathbf{v}_{\text{rot}}\|} \tag{3}$$

Here, $\mathbf{v}_{\text{rot}}^{\parallel}$ is the plane-parallel component of the rotated vector computed using the Rodrigues' rotation formula, $\mathbf{v}_{\text{base}}^{\parallel}$ is the plane-parallel component of the base vector which is rotated by the merged angle θ to compute the rotated vector, $\mathbf{n}$ is the plane normal about which the phalange is rotated, $\mathbf{v}_{\text{merged}}^{\perp}$ is the common plane-perpendicular component l is the length of the phalange vector, and c_{merged}^{i+1} is the new centre located at a distance l from the previous joint c_{merged}^{i} along the unit phalange vector $\hat{v}_{\text{rot}}$.

2 Angle computation from depth pose

The angle between two 3D vectors is usually estimated by computing the cosine inverse of the normalised dot product between the two vectors, i.e., $\theta = \cos^{-1}(\frac{\mathbf{v}_1\cdot\mathbf{v}_2}{\|\mathbf{v}_1\|\|\mathbf{v}_2\|})$. This angle θ, computed using the dot product, is estimated along the plane containing both vectors $\mathbf{v}_1$ and $\mathbf{v}_2$. However, such an approach cannot be employed in practice because the pose estimation network from [2] independently regresses each 3D joint without accounting for the collinearity of joints in each finger. Specifically, when the finger joints are non-collinear, the use of such an approach causes the angle between phalanges to be estimated along a plane different from the true plane, which introduces significant error in the computed angle estimate. This approach is henceforth referred to as the *dynamic-plane* approach to honor the dynamic estimation of the angle computation plane, and the approach proposed by us is henceforth called *fixed-plane* approach.

Fig. 6 in the main paper illustrates the plane selected by the dynamic-plane approach as compared to the true plane when the collinearity constraint of the finger joints is not satisfied. A large disparity between the true and

* *The two authors contributed equally to this paper.*

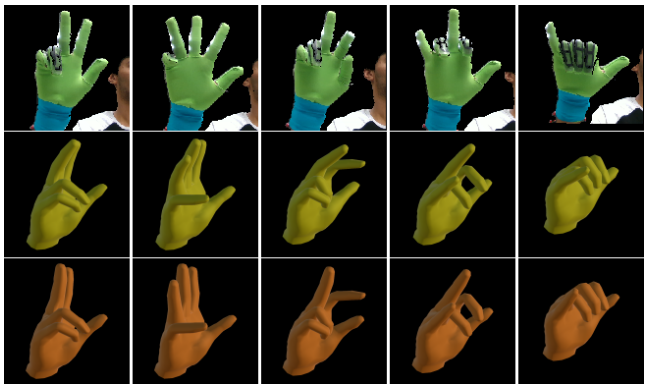

Fig. 1: Qualitative evaluation of the phalange angles estimated by the dynamic-plane (Row 2) and the fixed-plane (Row 3) approaches. The hand is colored using yellow and orange. The angles are obtained from the dynamic-plane and fixed-plane approaches, respectively. It can be observed that the straight fingers in the true pose are incorrectly predicted to be bent by the dynamic-plane approach but are accurate and straight in our fixed-plane approach. This error can be attributed to the incorrect plane selection of the dynamic-plane approach and is a direct consequence of ignoring the finger-joint collinearity constraint violation and directly estimating the rotation normal using the dot product.

Joint		Fixed Plane	Dynamic Plane
Index Top	(rad)	0.162	0.267
	(rad)	0.059	0.204
	(rad^2)	0.089	0.083
Middle Bottom	(rad)	0.112	0.219
	(rad)	0.073	0.179
	(rad^2)	0.023	0.032
Ring Top	(rad)	0.209	0.231
	(rad)	0.180	0.208
	(rad^2)	0.026	0.026
Pinky Bottom	(rad)	0.176	0.213
	(rad)	0.123	0.198
	(rad^2)	0.033	0.024
Thumb Bottom	(rad)	0.133	0.133
	(rad)	0.113	0.113
	(rad^2)	0.009	0.009

TABLE 1: Mean (μ), median (), and variance (σ^2) of the absolute difference between the phalange angles estimated by the fixed-plane and dynamic-plane approaches as compared to the ground truth from [1]. The fixed-plane approach has a very small error with the maximum median deviation being only $0.180\,\mathrm{rad}$ for the *Ring Top* joint, whereas the dynamic-plane approach consistently has a higher error with a maximum median deviation of $0.208\,\mathrm{rad}$ for the *Index Top* joint.

estimated planes can be observed, which makes computing the phalange angles using the dynamic-plane approach very unreliable. Fig. 1 qualitatively evaluates the phalange angles computed by the dynamic-plane approach (yellow) as compared to those estimated by the fixed-plane (orange) approach. It can be observed that the straight and extended fingers in the true hand pose are incorrectly predicted and bent in the dynamic-plane approach but are straight in the fixed-plane approach. This error can be attributed to the incorrect plane selection of the dynamic-plane approach and is a direct consequence of ignoring the finger-joint collinearity constraint violation.

This error in the dynamic-plane approach is also reflected in Table 1 which records the mean, median, and variance of the absolute deviation in the phalange angles as compared to the ground truth from [1]. The fixed-plane approach performs significantly better than its dynamic-plane counterpart, with the former having a maximum deviation of only $0.180\,\mathrm{rad}$ as opposed to $0.204\,\mathrm{rad}$ of the latter. Furthermore, the fixed-plane approach also results in smaller mean and median deviations for all finger joints, thus validating our decision of preferring the fixed-plane approach to the dynamic-plane approach.

3 More detail on hand model adaptation

Since the representation used in the $\mathcal{H}_G$ is largely a superset of that used in the $\mathcal{H}_D$, the centres in the $\mathcal{H}_G$ having a corresponding centre in the $\mathcal{H}_D$ are retained while the others are discarded. For the only centre in the $\mathcal{H}_D$ that does not have a corresponding centre in the glove model i.e., the centre representing the middle of the palm, simple geometry is used to compute its estimated position. Mathematically,

$$
\begin{aligned}
c^x_{\text{palm\_centre}} &= \frac{1}{4}(c^x_{\text{palm\_pinky}} + c^x_{\text{palm\_index}} + c^x_{\text{palm\_left}} + c^x_{\text{palm\_right}}) \\
c^y_{\text{palm\_centre}} &= \frac{1}{4}(c^y_{\text{palm\_pinky}} + c^y_{\text{palm\_index}} + c^y_{\text{palm\_left}} + c^y_{\text{palm\_right}}) \\
c^z_{\text{palm\_centre}} &= \frac{1}{2}(c^z_{\text{pinky\_base}} + c^z_{\text{index\_base}})
\end{aligned}
\tag{4}
$$

where c^x, c^y and c^z denote the X, Y, and Z coordinates of the centre in consideration respectively. The mapping of the centres between the depth and glove representations is depicted in Fig. 5 in the main paper, where the retained centers are depicted in blue, the discarded centers in red, and the center whose position is estimated using Eq. (4) is shown in green.

The collinearity constraint in the fingers is enforced in the final pose estimate by using a common plane-perpendicular component for all phalange vectors belonging to a finger. This common plane-perpendicular vector is estimated by first weighting the constituent plane-perpendicular components of the phalanges relative to each other and then computing the sum of the weighted plane-perpendicular vectors. Mathematically,

$$
K = \sum_i \alpha_{\mathbf{v}_i} \tag{5}
$$

$$
\mathbf{v}^{\perp}_{\text{merged}} = \sum_i \frac{\alpha_{\mathbf{v}_i}}{K} * \mathbf{v}^{\perp}_i \tag{6}
$$

where K is the cumulative sum of confidences of all phalange vectors in a finger, $\alpha_{\mathbf{v}}$ is the confidence of a phalange vector, $\mathbf{v}^{\perp}$ is the component of phalange vector $\mathbf{v}$ perpendicular to the plane of interest, and $\mathbf{v}^{\perp}_{\text{merged}}$ is the common merged plane-perpendicular component used to straighten the finger.

Phalange angle computation evaluation. Fig. 3 evaluates the performance of our phalange-angle computation