3D Finger Rotation Estimation from Fingerprint Images

YONGJIE DUAN, Department of Automation, BNRist, Tsinghua University, China
JINYANG YU, Department of Automation, BNRist, Tsinghua University, China
JIANJIANG FENG, Department of Automation, BNRist, Tsinghua University, China
KE HE, Department of Automation, BNRist, Tsinghua University, China
JIWEN LU, Department of Automation, BNRist, Tsinghua University, China
JIE ZHOU, Department of Automation, BNRist, Tsinghua University, China

Various touch-based interaction techniques have been developed to make interactions on mobile devices more effective, efficient, and intuitive. Finger orientation, especially, has attracted a lot of attentions since it intuitively brings three additional degrees of freedom (DOF) compared with two-dimensional (2D) touching points. The mapping of finger orientation can be classified as being either absolute or relative, suitable for different interaction applications. However, only absolute orientation has been explored in prior works. The relative angles can be calculated based on two estimated absolute orientations, although, a higher accuracy is expected by predicting relative rotation from input images directly. Consequently, in this paper, we propose to estimate complete 3D relative finger angles based on two fingerprint images, which incorporate more information with a higher image resolution than capacitive images. For algorithm training and evaluation, we constructed a dataset consisting of fingerprint images and their corresponding ground truth 3D relative finger rotation angles. Experimental results on this dataset revealed that our method outperforms previous approaches with absolute finger angle models. Further, extensive experiments were conducted to explore the impact of image resolutions, finger types, and rotation ranges on performance. A user study was also conducted to examine the efficiency and precision using 3D relative finger orientation in 3D object rotation task.

CCS Concepts: • **Human-centered computing** → **Interaction techniques**; • **Computing methodologies** → **Artificial intelligence**.

Additional Key Words and Phrases: 3D relative finger orientation, fingerprint image, deep learning, 3D manipulation

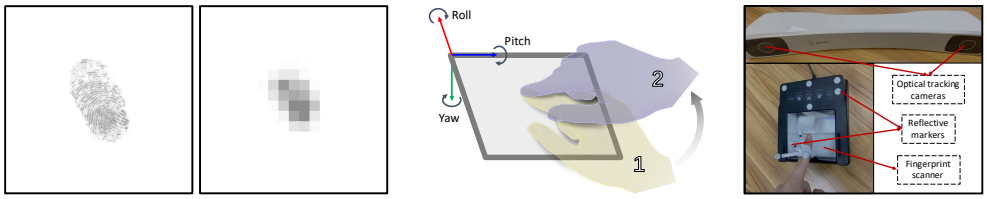
ACM Reference Format:
Yongjie Duan, Jinyang Yu, Jianjiang Feng, Ke He, Jiwen Lu, and Jie Zhou. 2023. 3D Finger Rotation Estimation from Fingerprint Images. *Proc. ACM Hum.-Comput. Interact.* 7, ISS, Article 431 (December 2023), 23 pages. https://doi.org/10.1145/3626467

1 INTRODUCTION

Over the past decades, touchscreen has been widely used as the main interaction technology of smartphones, tablets, and other mobile devices. However, due to the low resolution of capacitive

Authors' addresses: Yongjie Duan, dyj17@mails.tsinghua.edu.cn, Department of Automation, BNRist, Tsinghua University, Beijing, China; Jinyang Yu, jy-yu20@mails.tsinghua.edu.cn, Department of Automation, BNRist, Tsinghua University, Beijing, China; Jianjiang Feng, jfeng@tsinghua.edu.cn, Department of Automation, BNRist, Tsinghua University, Beijing, China; Ke He, he-k18@mails.tsinghua.edu.cn, Department of Automation, BNRist, Tsinghua University, Beijing, China; Jiwen Lu, lujiwen@tsinghua.edu.cn, Department of Automation, BNRist, Tsinghua University, Beijing, China; Jie Zhou, jzhou@tsinghua.edu.cn, Department of Automation, BNRist, Tsinghua University, Beijing, China.

2573-0142/2023/12-ART431
https://doi.org/10.1145/3626467



(a) (b) (c)

Fig. 1. Estimating 3D relative finger orientation based on 2D images and the data acquisition system. (a) A fingerprint image (500 ppi) and its simulated capacitive image (10 ppi). (b) Definition of 3D relative finger rotation (from pose 1 to pose 2), represented by three Euler angles: yaw (green), pitch (blue), and roll (red). (c) The whole data acquisition system, including optical tracking cameras, a fingerprint scanner, and several reflective markers attached on a small stent and the fingerprint scanner respectively.

images captured by measuring capacitive changes caused by fingers touching on the touchscreens [32], only two-dimensional (2D) touching point is acquired from the touchscreen driver despite the flexibility of human fingers. Recently, researchers have proposed various novel interaction techniques, including touching area shape [33], size [4], fingernails [22], part of finger [15], finger identification [25], shear force [14, 17], finger angles [12, 24, 28, 37, 38, 47], hand part recognition [15], and hand pose [1, 6, 21], to extend the richness of input vocabulary in various HCI applications.

Among these interaction techniques, finger angles present notable manipulation potential because of the precise and intuitive interaction using fingers for additional degrees of freedom (DOF) compared to 2D touching points. For interactions based on shear force, interactions using finger orientation present superiority since finger angle is easier to be controlled across different users. Compared to size and shape of touching area, finger orientation provides more DOFs, and one of the main factors for changes in touching area shape and size is diverse finger angles. Intuitively, introducing 3D finger orientation allows natural mapping for 3D object manipulation [24, 46] and interactions on small screens, e.g. slider selection on smartwatches [47]. Besides, pencil width modification, menu selection, and joysticks substitute can also be implemented using finger angles [44]. Besides, 3D finger orientation is also utilized to refine 2D touching points [20]. In addition to accurate finger orientation estimation, identifying extreme finger angles, e.g. touching with large pitch angle, provides valuable information for rejecting unintended touching.

It is still a challenging task to estimate 3D finger angles accurately based on 2D images. Recently, capacitive images are utilized to predict 3D finger angles [28, 47], while presenting limited accuracy due to quite low resolution and lack of information. Some auxiliary sensors were employed to alleviate the information gap, e.g. the photosensitive device attached to fingernails [46], RGB cameras [7], and depth cameras [24, 29, 31]. Obviously, utilizing these extra devices is inconvenient for mobile applications. Besides, performing roll rotation is comparatively easy, but roll angle is rarely explored in prior works probably because it is hard to estimate based on capacitive images. With the development of under-screen fingerprint sensing techniques [34, 36, 48], capturing fingerprint images of fingers interacting with screens becomes feasible. Some recent works also proposed to utilize fingerprint images to estimate 3D finger angles for interaction [8, 16]. Figure 1(a) shows an example of comparison between fingerprint and capacitive image (simulated for an approximate resolution). Compared to capacitive image, fingerprint image with a higher resolution contains more useful information, e.g. ridge patterns and fine boundaries, making it feasible to predict all 3D finger angles.

Vogelsang et al. [44] classified the mapping from input to object into two types, namely absolute and relative. Absolute mapping utilizes the angles between fingers and the display for interaction, resulting in fast and intuitive manipulations such as fast reorientation. However, ergonomic constraints limit the input space in absolute mapping, e.g. large pitch angles are not feasible to perform with long fingernails, thus requiring counterintuitive scaling for complete 3D orientation control [47]. Consequently, experts advocate for the investigation of relative input interactions [18, 27] which focuses on the difference between initial finger angles (while touching the display) and current finger angles. The definition of 3D relative finger angles while touching on a display refers to the concepts in flight control (as shown in Figure 1(b)). Relative finger orientation-based interaction is similar to mouse input, whereby manipulation is terminated when lifting fingers off the screen and resumes after moving down [44]. Intuitively, manipulation range is enlarged using relative mapping (e.g. $[-90°, 90°]$ for relative pitch angle and $[-180°, 180°]$ for relative yaw angle), which enables knob, wheel menu picker, and other interactions with large or circular value space [44].

Various applications using finger orientation have been proposed and fully explored in previous research [44], while we believe the unsatisfactory precision is one of the most important factors preventing its widespread adoption in daily applications. Absolute finger orientation is explored in previous studies without exception. Although relative finger angles can be obtained by calculating the transformation between two absolute finger angles, predicting 3D absolute finger angles itself may be inaccurate since zero finger orientation definitions are difficult to determine and not consistent across fingers with diverse finger shapes and sizes, especially for thumbs. In contrast, estimating relative finger angles directly advances the estimation accuracy as it focuses on the difference between two inputs and mitigates the problem of defining zero finger orientation.

Therefore, in this paper, we focus on improving the precision and stability of 3D relative finger orientation estimation. Different from prior works which estimate absolute finger angles from a single capacitive or fingerprint image, we first propose to estimate 3D relative finger angle based on two fingerprint images directly and utilize the relative rotation angles for interaction. To further improve the generalization ability across different fingers with various shapes, sizes, identities, and other characteristics, we decompose the latent feature extracted from fingerprint into pose-relevant and pose-irrelevant, then estimate relative 3D finger angles based on the pose-relevant features only.

All three relative finger angles (yaw, pitch, and roll) that fully describe 3D finger rotation are predicted simultaneously, thanks to the incorporation of fingerprint images with higher resolution and more information than capacitive images. Mean absolute error (MAE) of three relative finger angles in experiments are $9.14°$ for yaw, $6.01°$ for pitch, and $8.41°$ for roll, respectively. Given the accurately estimated 3D relative finger orientations, rotations of human fingers in 3D space could be described more precisely, thus promoting the richness of input vocabulary for touch based interface.

2 RELATED WORKS

Existing finger orientation estimation techniques can be classified into three categories based on sensing techniques, including touch sensor based, auxiliary device based, and fingerprint sensor based.

2.1 Touch Sensor

Capacitive images are widely used in various human interaction applications and commercial devices. Generally, the shape of silhouette and distribution of capacitance around the touching area vary with different finger angles. This property is applied in various 3D finger orientation estimation

algorithms directly. Wang et al. [45] attempted to estimate yaw using the shape of touching area. Rogers et al. [37] attempted to use a capacitive sensor matrix to track 2D touching points and finger orientation (yaw and pitch) simultaneously. Zaliva et al. [50] made a step forward by extracting additional features, such as capacitive value and asymmetry of touching area, to estimate 3D finger angles for gesture recognition. However, no quantitative experiment was conducted to evaluate finger angles estimation accuracy in these studies. Xiao et al. [47] extracted 42 manually defined features from touching area to predict yaw and pitch angles, but the estimation performance is evaluated on several discrete angles (15° increments). Mayer et al. [28] first adopted a convolution neural network (CNN) model for pitch and yaw estimation. Using the ground truth finger angles recorded by an optical tracking system, they reported a state-of-the-art (SOTA) performance based on capacitive images. However, due to the low resolution of capacitive images, the deep network is very shallow and the estimation errors are not sufficiently small. Meanwhile, estimating roll angle is rarely explored despite the fact that rolling is easy to perform for most fingers and the range of roll is also larger than pitch.

2.2 Auxiliary Device

Some auxiliary devices were also employed for finger orientation estimation. Watanabe et al. [45] placed a photosensitive sensor on fingernails to identify luminance attenuation with different finger angles, while the attached sensor constrained the natural range of finger rotations. Furthermore, Kratz et al. [24], Mayer et al. [29], and Murugappan et al. [31] utilized depth cameras to capture point clouds of fingers while touching, based on which pitch and yaw angles were estimated. However, application scenarios are restricted due to the difficulties and costs of introducing additional sensors for sensing a finger interacting on mobile devices.

2.3 Fingerprint Sensor

Fingerprint based finger orientation estimation has received little attention in HCI community. In [13], 2D translation and rotation (namely, relative yaw angle) between two fingerprint images were calculated by analyzing optical flow. Holz and Baudisch [20] proposed to rectify 2D touching points using 3D finger orientation estimated from fingerprint images by k-nearest neighbor searching, while estimation accuracy for 3D finger angles was not reported since they focused on 2D touching position rectification. Major weaknesses using this method for finger angle estimation are: (1) it requires enrollment of a number of fingerprints with ground truth angles from the specific finger and (2) estimation performance conflicts with time consumption in estimation stage, i.e. higher performance requires more enrolled fingerprints while consuming more time in searching. Duan et al. [8] attempted to reduce searching time in enrollment by reconstructing 3D surface from sequential fingerprint images, and all three fingerprint angles were estimated by point matching and projection parameters estimation, which requires higher image resolution for robust key-points extraction. He et al. [16] proposed to predict 3D finger angles from 2D fingerprint images directly, and demonstrated the superiority of utilizing fingerprints for 3D finger angles estimation compared with capacitive images. Besides, both absolute and relative finger angle play an important role in fingerprint recognition, in which only yaw angle is considered since fingerprint is usually viewed as a 2D image [26]. Registration of two fingerprints, involving estimation of relative yaw angle, is a routine step in fingerprint recognition. For absolute yaw angle estimation, [49] reported very low error (1.45° for yaw angle). However, fingerprint images adopted in [49] are rolled fingerprints captured using forensic techniques, which contains more information and higher quality compared with fingerprints used in daily applications. And the range of yaw angle is also limited since these fingerprints are collected in controlled environment for identity recognition purpose.

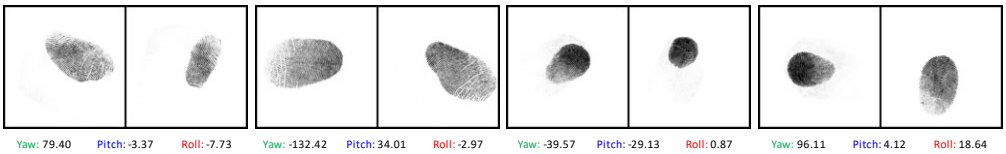


Fig. 2. Several collected examples. The 3D relative orientation is described by three Euler angles based on the definition in Figure 1(b).

In summary, several recent research have been conducted to explore the estimation of absolute 3D finger angles from 2D capacitive or fingerprint images. However, despite benefits of relative input interaction, to the best of our knowledge, there has been no research to estimate 3D relative finger angles based on 2D fingerprint images up to date.

3 DATASET

To train and evaluate the proposed algorithm, we constructed a dataset consisting of fingerprint pairs and their corresponding 3D relative finger rotation based on the dataset collected by He et al. [16]. Note that we cannot directly use the dataset in [16], since such research considers only absolute finger angles, while our task aims to estimate 3D relative finger rotation. For the completeness of the paper, a brief summary of the data collection method is provided here, and more details can be found in [16].

The acquisition system, shown in Figure 1(c), contains a frustrated total internal reflection (FTIR) fingerprint scanner DF500[1] and an optical tracking system PST Iris[2]. A small stent with 4 reflective markers was fixed on the back of fingers, and additional 5 reflective markers were also attached on the fingerprint scanner. Specifically, sequential fingerprint images (frames) are captured by the fingerprint scanner, and the corresponding 3D finger angles are determined by calculating the transformation between the small stent and fingerprint scanner.

In total, the dataset was collected from 22 participants, including 12 male and 10 female with ages from 20 to 48 [16]. Six fingers including thumbs, index, and middle from both left and right hands for each subject were captured (132 fingers in total). Participants were asked to press fingers on the fingerprint scanner with arbitrary angles to ensure a wide range of all three angles in a comfortable way, during which the fingerprint images and the corresponding 3D finger absolute angles were captured simultaneously (the acquisition sample rate was set as 20 Hz). And then we excluded those data in which no fingerprint is captured due to extreme finger angles.

The dataset was then regrouped into pairs randomly, where each data pair consists of two fingerprint images as well as their corresponding 3D relative finger rotation angles. Note that only data pairs from the same finger are considered, since the relative rotation between different fingers is meaningless for interaction purpose. Considering the similarity between adjacent frames, up to 3,000 pairs were selected from each finger. Finally, a total of 332,418 pairs consisting of fingerprint images and their corresponding 3D relative finger rotation ground truth were collected. The fingerprint scanner used in this study offers a 1.6" × 1.5" effective touch area, and the size of captured fingerprint images is 800 × 750 pixels with resolution of 500 ppi originally. We down-sampled them to 180 ppi and cropped to 256 × 256 pixels for computation efficiency. Several examples are shown in Figure 2.

[1] see details in http://www.dotutech.com/en/pro_d.php?id=3
[2] see details in https://www.ps-tech.com/products-pst-iris

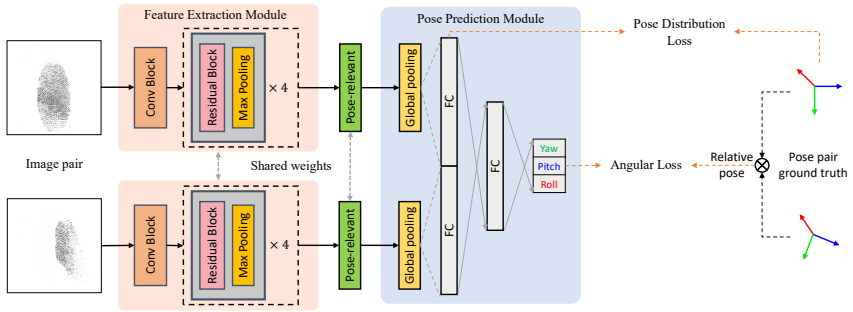


Fig. 3. Pose-relevant features are extracted and concatenated to predict 3D relative finger rotation.

4 METHODS

We propose to estimate 3D finger rotation using a siamese framework that utilizes two neural networks with the same architecture and shared weights to extract features from a pair of 2D fingerprint images. The extracted features are then concatenated to predict 3D relative finger rotation. Considering the different characteristics of 3D relative finger rotation caused by diverse initial finger angles and finger types, we also utilize metric learning to constrain the extracted features. As mentioned in Section 3, data pairs, consisting of two fingerprint images and their corresponding 3D relative rotation ground truth, are used for algorithm training and evaluation. Figure 3 illustrates the schematic representation of our approach.

4.1 Network Architecture

Siamese network has been widely used in relative camera pose estimation [9]. Inspired by this idea, we propose a pose-aware siamese network to extract pose-relevant features, which are then concatenated to estimate 3D relative rotation represented by three Euler angles: roll α, pitch β, and yaw γ. The proposed network comprises three modules: (a) a feature extraction module to derive features from input images; (b) a reconstruction module to restore input images based on the extracted features; and (c) a prediction module to estimate 3D relative finger angles. The feature extraction module consists of a convolution block with a kernel size of 7 and a stride of 2, alongside several residual-pooling blocks. The resultant features maps are flattened using global average pooling and followed by a concatenation layer. Finally, the 3D relative finger rotation is predicted via two fully connected (FC) layers. And the decoder module is introduced to restore the input images, thus preserving sufficient information in the latent feature space.

4.2 Feature Decomposition

Characteristics of 3D relative finger orientation between two input images are different with diverse initial finger angles, e.g., the difference between input images is different when rolling with initial absolute pitch angle at 10° and 70°. Thus, to extract discriminative features containing more essential information and exclude pose-irrelevant information like finger size, identity, and other irrelevant poses, it is helpful to decompose to latent interpretable features, and recent studies have demonstrated its effectiveness in various tasks [2, 5, 42]. In this paper, an auto-encoder (AE) network is employed for discriminative feature extraction and latent space disentanglement as shown in Figure 4. With the reconstruction branch, sufficient information is reserved within latent feature space and only pose-relevant part is utilized for following 3D relative angles estimation.

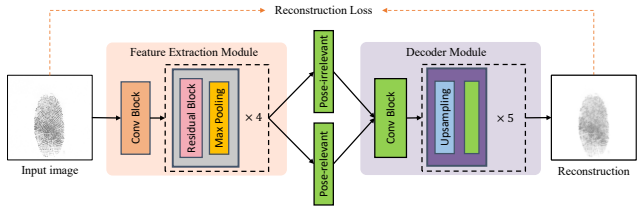


Fig. 4. Feature extraction module (decompose latent features into pose-relevant and pose-irrelevant) and decoder module.

4.3 Pose Distribution Learning

Inspired by the idea of metric learning [23, 40, 51], which aims to learn a reliable distance function for measuring image similarities, in this paper, latent feature disentanglement is achieved by constraining similarities between the extracted pose-relevant features. Considering that the definition of zero finger orientation is inconsistent across different fingers, we propose to approximate the distance between extracted pose-relevant features $D(\boldsymbol{f}_i, \boldsymbol{f}_j)$ to the relative rotation distance between 3D absolute finger angles $D(p_i, p_j)$. Specifically, distance $D(\boldsymbol{f}_i, \boldsymbol{f}_j)$ is defined as Euclidean distance between the extracted embeddings $\boldsymbol{f}$ from two input images:

$$D(\boldsymbol{f}_i, \boldsymbol{f}_j) = \left\| \boldsymbol{f}_i - \boldsymbol{f}_j \right\|_2, \tag{1}$$

The 3D finger orientation distance $D(p_i, p_j)$ is calculated as angular distance $D(p_i, p_j)$

$$D(p_i, p_j) = \arccos\left(\frac{\mathrm{trace}(\boldsymbol{R}_j \boldsymbol{R}_i^{-1}) - 1}{2}\right), \tag{2}$$

where $\boldsymbol{R}_i$ and $\boldsymbol{R}_j$ are the rotation matrix of 3D finger orientation p_i and p_j, respectively, and $\mathrm{trace}(\cdot)$ denotes the trace of matrix. Note that the angular distance $D(p_i, p_j)$ is the rotation angle in axis-angle representation of 3D relative rotation from pose p_i to pose p_j. Then the distance of log distance ratios based on triplets [23, 51] can be minimized:

$$\mathcal{L}(\boldsymbol{f}_i, \boldsymbol{f}_j, \boldsymbol{f}_k) = \left(\log \frac{D(\boldsymbol{f}_i, \boldsymbol{f}_j)}{D(p_i, p_j)} - \log \frac{D(\boldsymbol{f}_i, \boldsymbol{f}_k)}{D(p_i, p_k)}\right)^2, \tag{3}$$

which is modified to improve the computation efficiency and take full advantage of samples within a mini-batch [51]:

$$\mathcal{L}(\boldsymbol{B}) = b(b-1) \sum_{i \neq j \in \boldsymbol{B}} \left(\log \frac{D(\boldsymbol{f}_i, \boldsymbol{f}_j)}{D(p_i, p_j)}\right)^2 - \left(\sum_{i \neq j \in \boldsymbol{B}} \log \frac{D(\boldsymbol{f}_i, \boldsymbol{f}_j)}{D(p_i, p_j)}\right)^2, \tag{4}$$

where $\boldsymbol{B}$ denotes the indices within a mini-batch and b is the batch size.

In this way, information from fingerprint images of various 3D finger angles and different fingers is incorporated, thus improving the features discrimination and generalization ability of the proposed approach.

4.4 Objective Functions

4.4.1 Angular loss. Three Euler angles (namely yaw, pitch, and roll) are predicted and used to represent relative 3D finger rotation. The widely used mean-squared error loss function is utilized for optimization:

$$\mathcal{L}_{\text{ang}} = \frac{1}{3b} \sum_{i \in \boldsymbol{B}} \left(d(\hat{\alpha}_i, \alpha_i)^2 + d(\hat{\beta}_i, \beta_i)^2 + d(\hat{\gamma}_i, \gamma_i)^2\right), \tag{5}$$

where $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ are predictions, and α, β, and γ are the corresponding ground truth value respectively. Considering the periodicity of orientation angle, which means -180° is consistent with 180° while completely different for neural network, we normalize the difference between two angles $d(\theta_1, \theta_2)$ by determining the minimum of $\Delta\theta$ and $360^\circ - \Delta\theta$.

4.4.2 Reconstruction Loss. By constraining the distance between original and reconstructed image in pixels, extracted latent features retain sufficient information for accurate image recovery:

$$\mathcal{L}_{\text{rec}} = \frac{1}{b\omega} \sum_{i \in \boldsymbol{B}} \sum_{j \in \boldsymbol{\Omega}} |\hat{\boldsymbol{I}}_{i,j} - \boldsymbol{I}_{i,j}|, \tag{6}$$

where $\hat{\boldsymbol{I}}$ is the reconstruction of input $\boldsymbol{I}$, $\boldsymbol{\Omega}$ denotes the pixel indices across input image, and ω is the size of $\boldsymbol{\Omega}$.

4.4.3 Pose distribution loss. As mentioned in Section 4.3, we utilize the dense loss within a mini-batch to constrain the distribution of extracted pose-relevant features:

$$\mathcal{L}_{\text{pose}} = \frac{1}{b(b-1)} \sum_{i \neq j \in \boldsymbol{B}} D_{\log}^2 - \left(\frac{1}{b(b-1)} \sum_{i \neq j \in \boldsymbol{B}} D_{\log}\right)^2, \tag{7}$$

where

$$D_{\log} = \log \frac{D(\boldsymbol{f}_i, \boldsymbol{f}_j)}{D(p_i, p_j)} = \log D(\boldsymbol{f}_i, \boldsymbol{f}_j) - \log D(p_i, p_j). \tag{8}$$

4.4.4 Overall loss. The proposed network is optimized by minimizing the following overall objective function

$$\mathcal{L} = \mathcal{L}_{\text{ang}} + \lambda_{\text{rec}}\mathcal{L}_{\text{rec}} + \lambda_{\text{pose}}\mathcal{L}_{\text{pose}} + \lambda_{\text{reg}}\mathcal{L}_{\text{reg}}, \tag{9}$$

where λ_{rec}, λ_{pose}, and λ_{reg} denotes trade-off parameter of reconstruction loss, pose distribution loss, and regularization loss, respectively. In this paper, we utilize $L2$ norm of trainable network parameters for regularization. We set $\lambda_{\text{rec}} = 1.0$, $\lambda_{\text{pose}} = 1.0$, and $\lambda_{\text{reg}} = 0.01$ such that these loss functions have similar decreasing rates during training.

4.5 Implementation Details

We implement the proposed network in PyTorch and train it on a single NVIDIA GeForce 2080Ti. During training, data is augmented via random image translations up to 20% of the image width and height, as well as random rotations within the range of $[-180^\circ, 180^\circ]$ to increase the diversity of yaw angles. AdamW optimizer with initial learning rate of 0.00035 is utilized to update network parameters. The learning rate decays by 0.1 when performance on validation subset does not increase for 10 epochs, and training procedure stops after learning rate decays for three times. The entire network is trained from scratch.

5 EXPERIMENTS

In this section, we first introduce the dataset used for training and evaluation. Afterwards, the compared baseline methods and a new performance metric are described. Subsequently, experimental results are reported, and the effects of various factors on the estimation performance are also explored such as input image resolution and finger type.

5.1 Dataset

Experiments were conducted on the dataset described previously, consisting of a total of 332,418 pairs from 132 fingers (22 subjects) with fingerprint images and their corresponding 3D relative finger angles ground truth. We randomly split the dataset into three subsets: 235,569 pairs from 92

Table 1. Distributions of three angles in the test subset. The definition of "AD" is shown in Equation 10. "AD" of "Absolute" denotes the angular rotation from zero finger orientation to current absolute finger orientation. "SD" denotes standard deviation.

	Absolute				Relative			
	Yaw	**Pitch**	**Roll**	**AD**	**Yaw**	**Pitch**	**Roll**	**AD**
Min	-88.0	-87.5	-85.2	13.0	-180.0	-88.9	-101.3	2.0
Max	90.0	-9.1	89.5	153.0	179.9	73.5	102.5	180.0
Mean	2.8	-41.3	-0.2	81.8	-2.0	-13.7	-0.8	100.9
SD	60.5	16.6	37.3	18.9	103.7	29.8	35.1	49.2

fingers for training, 36,968 pairs from 13 fingers for validation, and 59,881 pairs from 27 fingers for testing. Note that three subsets may contain fingers from the same participants, but all images from the same finger were assigned to single subset. The use of fingerprint images with an original resolution of 500 ppi is neither practical nor essential for mobile devices in daily applications due to their limited volume and computational resources. Therefore, we down-sample the fingerprint images to 180 ppi (256×256 pixels). Before down-sampling, low-pass filtering was applied to provide spatial anti-aliasing. The statistical details of the test subset can be found in Table 1, which highlights the considerable enlargement of the ranges for all three relative angles compared to absolute angles, particularly for yaw and pitch angles.

5.2 Baseline Methods

We reimplemented the method in Mayer et al. [28] as a baseline since it first introduced deep learning to estimate finger angles based on capacitive images. Considering that the network was designed for capacitive images with low resolution originally, images with higher resolution, e.g., fingerprint images, might not be compatible due to the limited network receptive field. Consequently, on the one hand, we simulated capacitive images by down-sampling the original fingerprint images with spatial anti-aliasing to match the general resolution of capacitive images (~4 mm pitch, ~10 ppi), as shown in Figure 1(a). On the other hand, we also developed a deeper multi-task CNN model in [16] to predict 3D finger angles based on fingerprint images. Note that we added an adaptive global average pooling layer with output size of 3 in the CNN model in Mayer et al. [28] to estimate on images with arbitrary resolutions. The aforementioned networks predict absolute 3D finger angles directly from the single input image, then the relative 3D finger orientation can be determined from two separate inputs. Besides, we also proposed a naïve siamese network that uses the same feature extraction module and FC layers as our proposed model, while removing the latent feature disentanglement module. This network also predicts the 3D finger relative orientations directly.

To ensure fairness and comprehensiveness of comparison with baseline methods, the same dataset is utilized for training and evaluation for all methods, and we also apply the same data augmentation as mentioned above. All networks are trained from scratch using the same optimizer and learning rate decay schedule.

5.3 Performance Metrics

Apart from the widely used Euler angles, we also evaluate the estimation performance in 3D space directly. Similar with Equation (2), 3D rotation error, named 3D angular distance (AD), is proposed for a better measurement of distance between two 3D finger rotations, which can be calculated by:

$$\mathrm{AD} = \arccos\left(\frac{\mathrm{trace}(\boldsymbol{R}\hat{\boldsymbol{R}}^{-1}) - 1}{2}\right), \tag{10}$$

Table 2. Quantitative results for 3D relative finger angles estimation. Default image resolution is 180 ppi. Errors are reported in degrees.

Algorithm	Yaw			Pitch			Roll			AD		
	MAE	RMSE	SD	MAE	RMSE	SD	MAE	RMSE	SD	MAE	RMSE	SD
CNN in Mayer et al. [28][1]	36.50	57.37	44.27	15.99	21.84	14.87	19.04	26.97	19.10	44.42	60.67	41.32
CNN in Mayer et al. [28]	22.45	34.39	26.03	11.03	14.44	9.33	15.04	22.24	16.38	28.07	35.47	21.69
Multi-task CNN [16]	13.91	22.45	17.63	8.74	11.65	7.70	11.72	16.89	12.17	19.69	24.96	15.34
Naïve Siamese	10.20	14.91	**10.87**	7.56	9.39	5.56	10.12	13.92	9.55	14.99	17.16	**8.35**
Pose-aware Siamese	**9.14**	**14.83**	11.67	**6.01**	**7.84**	**5.04**	**8.41**	**12.13**	**8.75**	**13.46**	**16.96**	10.31

[1] down-sample the input fingerprint image to 10 ppi for capacitive image simulation.

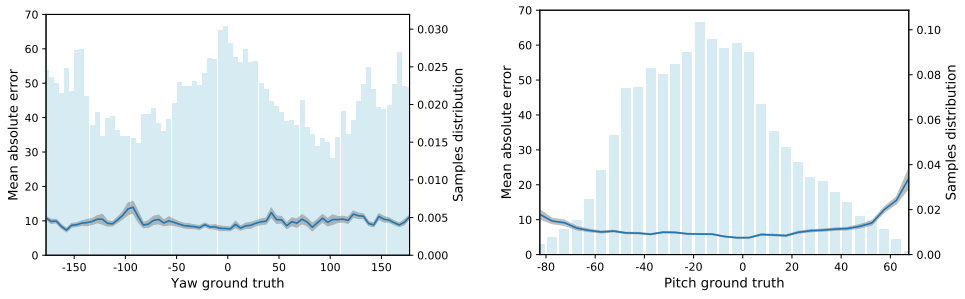
where $\boldsymbol{R}$ and $\hat{\boldsymbol{R}}$ are rotation matrixes of ground truth 3D relative finger orientation and prediction, respectively. Mean absolute error (MAE), root mean squared error (RMSE) and standard deviation (SD), are applied to perform quantitative performance comparison between different approaches.

5.4 Comparison with Baselines

In Table 2, we show the experimental results of the proposed method and baseline methods on the test subset. As shown in the table, the proposed approach performs better than other baseline methods. The models predicting 3D relative finger angles, i.e. naïve siamese and pose-aware siamese networks, perform better than those models based on absolute finger orientation, which demonstrates the superiority of estimating 3D relative finger rotation directly. Note that the metric SD is decreased when estimating relative finger angles directly, which means relative estimations are more stable and robust compared to calculating relative rotation angles between two absolute finger angle estimations. Besides, we observe that with a higher resolution of input image, i.e., fingerprint rather than capacitive images, estimation performance of [28] also increases. Figure 5 also shows the error distribution of our method under different ground truth values of three Euler angles and AD.

Figure 6 shows several results estimated by baseline methods and our proposed relative finger orientation estimation model. As shown in the figure, inaccurate and opposite rotations are observed in those approaches that calculate relative finger angles based on absolute finger angle estimations, which itself is not reliable.

5.4.1 Analysis of image resolution. Additional experiments are conducted to investigate the impact of input resolution. We down-sampled the original fingerprint images to several different resolutions, including 180 ppi (256×256 pixels), 120 ppi (171×171 pixels), 60 ppi (85×85 pixels), 30 ppi (43×43 pixels originally, padded to 48×48 pixels), and 10 ppi (14×14 pixels originally, padded to 32×32 pixels). Examples with different resolutions are shown in Figure 7, and experimental results are shown in Table 3. The down-sampled images with resolution of 10 ppi may not be consistent with real capacitive images (e.g., capacitive sensing detects the proximity rather than contact [41]). Hence, these simulated 10 ppi images are solely utilized for the shallow CNN model [28] as a performance reference. As shown in the table, it is observed that estimation performance increases with higher resolution, since additional features, such as ridge patterns, help to recognize 3D finger rotation angles accurately. Superior performance is consistently achieved by predicting 3D relative finger angles directly, especially for low image resolution, which further demonstrates the effectiveness of focusing on difference between inputs in 3D relative finger orientation estimation.



(a) Error distribution for yaw angle.

(b) Error distribution for pitch angle.

(c) Error distribution for roll angle.

(d) Error distribution for AD.

Fig. 5. Error distribution of the proposed method for 3D relative finger orientation represented by all three angles and angular distance (AD) in the test subset. 95% confidence interval (CI) is shown as gray area.

Note that this experiment does not intend to verify whether our method can be applied on capacitive images. We are actually trying to explore the estimation performance given different resolutions of input images. Such experiments are valuable and comprehensive since lower resolution always means higher operation speed, allowing for a better trade-off between estimation error and frame rate in practical HCI scenarios.

5.4.2 Analysis of finger type. As mentioned above, zero orientation definitions are not consistent across different fingers, which decreases performance of absolute finger orientation estimation in prior works. Therefore, estimation accuracy of our proposed method on different fingers were evaluated. Specifically, performance on three finger types is shown in Table 4. We found that the estimation accuracy on thumb is lower than the other two fingers when calculating 3D relative finger angles based on absolute finger angles. The main reason maybe that the shape and size of thumbs are more diverse compared to other fingers, making the definition of zero finger orientation across different fingers less reliable. While for our proposed method, difference between input images is focused and impacts like finger shape, size, and identities are also alleviated by latent feature disentanglement, thus achieving better and more robust performance.

Besides, we also conduct more experiments to explore the effects of various factors on the estimation performance, and more details can be seen in supplementary materials.

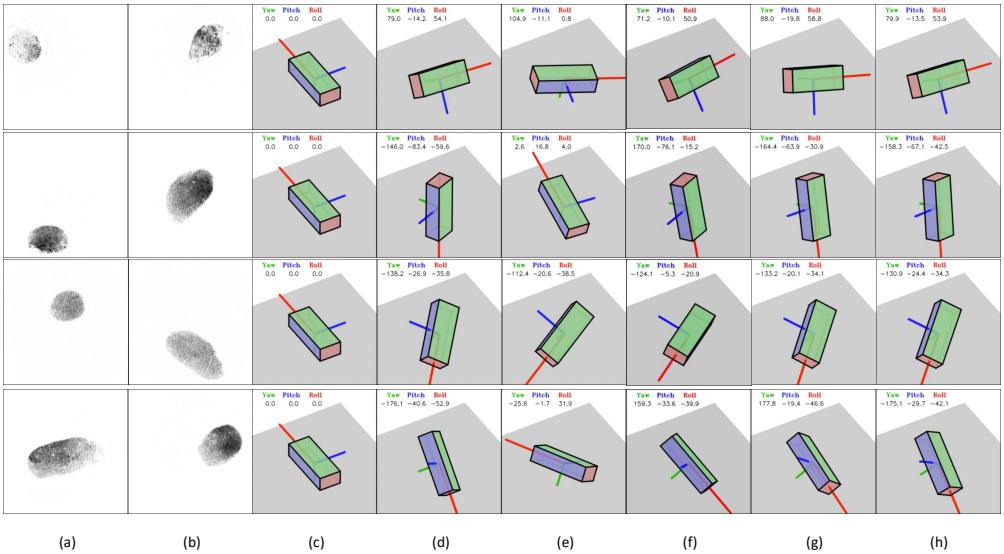


Fig. 6. Examples of 3D finger rotation estimation. From left to right, they are (a) the first frame, (b) the second frame, (c) zero initial pose, (d) 3D rotation ground truth from the first one to the second, (e) results of CNN in Mayer et al. [28], (f) results of multi-task CNN [16], (g) results of naïve siamese, and (h) results of pose-aware siamese network.

Fig. 7. Examples of fingerprint images with different resolutions.

6 USER STUDY

Finger orientation has been demonstrated to be intuitive in a number of HCI applications [44], and 3D manipulation task is very suitable for evaluating input technology rigorously since it requires more DOFs, accurate as well as stable control, and quick response [3]. Therefore, apart from quantitative evaluation experiments, a user study is also conducted to investigate relative finger orientation based 3D interaction in a realistic scenario, i.e. a representative 3D manipulation task in which a virtual object (teapot) is rotated to the target orientation.

6.1 Device

Due to the restriction of acquiring images in real time from fingerprint sensors embedded in mobile devices, we implemented using a fingerprint scanner with a 1.6" × 1.5" touch area (the same device shown in Fig. 1(c)). The fingerprint scanner was connected to a PC with GeForce 1080 GPU, and the same operation described in Section 3 was adopted, i.e., down-sampling the original fingerprint images to 180 ppi and cropped to 256 × 256 pixels.

Table 3. Quantitative results for 3D relative finger orientation estimation with different resolutions. Errors are reported in degrees.

Algorithm	PPI	Yaw			Pitch			Roll			AD		
		MAE	RMSE	SD	MAE	RMSE	SD	MAE	RMSE	SD	MAE	RMSE	S
CNN in Mayer et al. [28]	10	36.50	57.37	44.27	15.99	21.84	14.87	19.04	26.97	19.10	44.42	60.67	4
	30	31.20	49.82	38.84	13.72	18.89	13.00	17.10	24.25	17.19	38.15	52.58	3
	180	22.45	34.39	26.03	11.03	14.44	9.33	15.04	22.24	16.38	28.07	35.47	2
Multi-task CNN [16]	30	24.91	43.68	35.88	12.59	17.73	12.49	15.71	22.78	16.50	32.36	47.33	3
	180	13.91	22.45	17.63	8.74	11.65	7.70	11.72	16.89	12.17	19.69	24.96	1
Naïve Siamese	30	17.40	28.86	23.03	10.70	14.58	9.90	12.64	17.98	12.78	24.58	33.47	2
	180	10.20	14.91	10.87	7.56	9.39	5.56	10.12	13.92	9.55	14.99	17.16	8
Pose-aware Siamese	30	17.35	33.01	28.08	9.52	13.63	9.76	12.00	17.45	12.71	23.75	36.12	2
	60	15.22	29.04	24.73	8.65	12.46	8.97	10.68	15.77	11.61	20.78	31.89	2
	120	10.49	16.07	12.17	7.03	9.40	6.23	9.61	13.55	9.56	15.40	18.74	1
	180	9.14	14.83	11.67	6.01	7.84	5.04	8.41	12.13	8.75	13.46	16.96	1

Table 4. Quantitative results of 3D relative finger orientation estimation models on different fingers. Errors are reported in degrees.

	Algorithm	Yaw			Pitch			Roll			AD		
		MAE	RMSE	SD	MAE	RMSE	SD	MAE	RMSE	SD	MAE	RMSE	SD
Thumb	Multi-task CNN [16]	19.56	33.09	26.69	9.94	13.70	9.42	14.55	22.89	17.67	24.99	34.33	23.5
	Pose-aware Siamese	10.34	17.47	14.08	5.97	7.86	5.11	9.95	15.45	11.86	14.32	18.35	11.4
Index	Multi-task CNN [16]	9.94	14.62	10.72	7.59	9.40	5.54	9.36	13.16	9.26	15.19	16.87	7.3
	Pose-aware Siamese	7.33	14.02	11.95	5.68	7.25	4.50	7.11	10.34	7.50	11.56	15.77	10.7
Middle	Multi-task CNN [16]	13.24	18.81	13.36	8.83	11.78	7.80	12.01	16.54	11.36	19.49	22.90	12.0
	Pose-aware Siamese	10.06	15.00	11.12	6.26	8.21	5.30	8.68	12.27	8.67	14.25	16.99	9.2

6.2 Tasks

Each trial is started when users begin to manipulate the virtual object (the first time fingers touch the scanner), and two types of termination are explored in our study: (1) task is terminated on a key press by participants when the teapot is believed to reach the target orientation; (2) task is terminated after keeping absolute errors of all three angles less than 3° for 1 second. Then task completion time (type 1&2 task) and rotation error (type 1 task) are utilized to evaluate the efficiency and precision of 3D interaction techniques using different relative finger orientation estimation models, including (1) multi-task CNN [16] which calculates relative finger angles based on two absolute finger orientations, and (2) the proposed relative finger orientation estimation approach. We mapped the estimated relative finger angles to control 3D rotation of virtual object.

Inspired by the PRISM (precise and rapid interaction through scaled manipulation) techniques proposed by Frees et al. [10, 11], which scales down input movements to improve accuracy of direct manipulation, we implemented a naïve scaling method in which finger rotation angle around each axis is scaled down when users move their fingers slower than a pre-defined threshold around the corresponding axis (less than 3° between adjacent frames in this study).

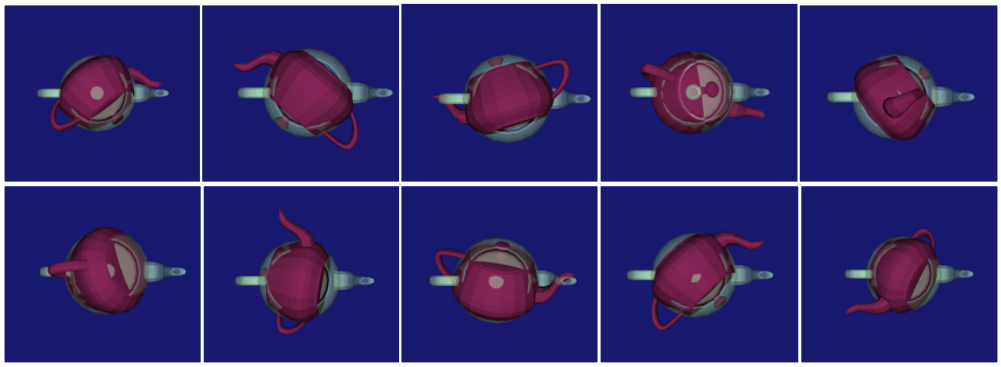


Fig. 8. Rotate teapot from initial orientation (red) to the target orientation (green).

(a) Type 1 task (b) Type 2 task

Fig. 9. Task completion times (in seconds) of (a) type 1 task and (b) type 2 task respectively, where error bars denote 95% CI.

Table 5. Average rotation errors with 95% CI in 3D manipulation task (type 1). Errors are reported in degrees.

Algorithm	Yaw	Pitch	Roll	AD
Multi-task CNN [16]	4.27±0.84	2.30±0.37	3.28±0.59	6.63±0.88
Pose-aware Siamese	2.98±0.44	1.36±0.18	2.61±0.39	4.82±0.41

6.3 Participants

Ten unpaid participants (all male with a university degree) took part in our comparative study, and none of participants were involved in the training dataset. They were intentionally given the minimal guidance on using the device to complete the manipulation task and asked to balance accuracy and speed in type 1 task. A video demo is also provided in supplementary materials.

Per relative finger orientation estimation model, we asked our participants to carry out 10 repetitions. The initial object orientation was randomly selected from remaining orientations within a pool, and we used the same pool for all methods (as shown in Figure 8). Besides, the order of employing models was also counter-balanced to reduce the bias caused by learning effects.

6.4 Quantitative Results

A total of 400 trials were collected from 10 participants (200 trials for each type of task). Quantitative results are shown in Figure 9 and Table 5. The CNN model in Mayer et al. [28] was also evaluated based on the simulated capacitive images (10 ppi). However, the tasks were rarely completed (each task is limited to maximum of 100 seconds) in most cases since the object orientation is hard to control and wobbles around the target orientation which is caused by inaccurate, unstable, or even contrary estimation of 3D finger angles.

6.4.1 Task Completion Time. It is observed that less completion time is achieved for rotating objects to the desired orientation using the proposed model, since keeping an object close to the target orientation requires a more accurate and stable response to the changes of finger orientation. Besides, paired t-test revealed that statistically significant differences for interaction based on the proposed model and the multi-task CNN [16] model with $p = 0.0003$ for type 1 task and $p = 0.0003$ for type 2 task. Note that log-normalizing was utilized as it is standard in such cases [3, 39]. And before that a Shapiro-Wilk test was applied to ensure that data normality assumption was met ($W = 0.987, p = 0.470$ for type 1 task and $W = 0.985, p = 0.331$ for type 2 task).

6.4.2 Rotation Error. For type 1 task, higher rotation precision is achieved based on the proposed model. Different from task completion time, we found that paired t-test is not appropriate since the data is not normally distributed according to Shapiro-Wilk test. Therefore, paired Wilcoxon test (also known as Wilcoxon signed-rank test) was utilized and we also found statistically significant differences regarding rotation error after completing type 1 task: $W = 1684, Z = -2.892, p = 0.0038$ for yaw, $W = 1815, Z = -2.441, p = 0.0147$ for pitch, $W = 1131, Z = -4.793, p < 0.0001$ for roll, and $W = 954, Z = -5.402, p < 0.0001$ for AD.

6.4.3 Experience. Three of the participants ranked themselves as skilled at 3D object manipulation task since they use 3D software or 3D video-games frequently. For these participants with relevant experience, the average task completion time using the proposed method is reduced to 13.67s and 14.58s for type 1&2 task respectively, which indicates a better control can be achieved with experience.

This 3D object rotation study shows the importance of finger angle estimation accuracy in human computer interaction applications. Note that this manipulation speed and accuracy are better than or at least comparable to the four input techniques for 3D object rotation reported in [19], although this is not a fair comparison due to different experimental settings.

6.5 Qualitative Results

Based on our observations during manipulation and subjective feedbacks from participants in this user study, several qualitative insights and discussions can be obtained.

6.5.1 Perferences. Subjective preferences were collected after participants finished the required 3D manipulation tasks. We asked the participants how they felt about the tested models and ranked them, including sensitivity (sensitivity of response to finger rotation angles), stability (consistency of response to finger rotation angles), and overall preference. Results are depicted in Figure 10. We found that most participants prefer the proposed 3D relative finger orientation estimation model, and followed by the absolute finger angle based model, multi-task CNN [16]. Stability seems to play an important role when participants rank all tested models. Due to the limited information provided by a single input, the absolute multi-task CNN [16] model presents lower reliability for manipulation. While our method does not suffer from this since more effective information can be reserved from two input images. For the CNN model inspired by Mayer et al. [28], which was designed for estimating pitch and yaw based on capacitive images originally, all participants agree that it is not easy to use, with the main problem being the large estimation error or the deviation from the actual rotation direction, since the lower image resolution makes it difficult to predict all three finger angles. And that is also the reason only pitch and yaw angle are concerned in prior works based on capacitive images.

6.5.2 Control Strategy. We did not impose a constraint on which finger the participants used during manipulation. An interesting finding was that people prefer to use their index fingers for manipulation, especially for those absolute finger angle estimation models. This can be backed

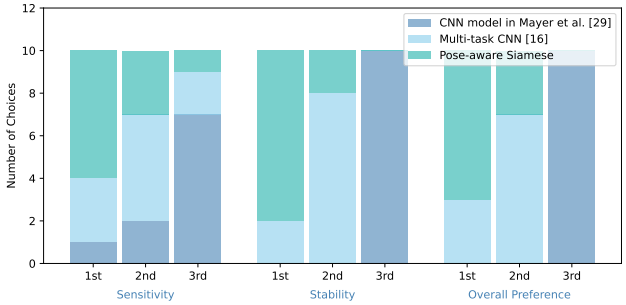


Fig. 10. Participants' preference for each model regarding different criteria, i.e., the number of times of each model was ranked as 1st/2nd/3rd.

by the observation that estimation performance on index fingers is higher than other fingers (see Table 4). Besides, some users tended not to change pitch angle drastically, but to control it using roll angle by first rotating 90° around yaw axis. The main reason is the limited range of pitch angle according to the feedbacks, which also fits the picture of roll angle being valuable in finger based 3D manipulation tasks.

6.5.3 Subjective Feedbacks. Some participants pointed out that if the fingerprint scanner can be made as small as a smartphone, it would not have to be placed on a table and could be easily held by one hand and operated by the other hand, making it more convenient for manipulation. Some users believe that if the refresh rate were higher, the manipulation experience would be better. There are also several people claimed that the reason they prefer a certain method is that the estimated finger angles are more consistent with their self-perception, even though we did not record the ground truth of 3D finger angles in user study.

7 DISCUSSION

Finger orientation based interactions have been explored in prior works, but have not yet been widely adopted in real applications. We believe that one of the main obstacles for real applications is the relatively lower estimation precision. Except for the low resolution of touch sensors, predicting absolute finger angles but utilizing relative finger rotation in real applications also limits the improvement of estimation performance. In this paper, we first proposed to predict 3D relative finger rotation angles based on two input fingerprint images directly. Experimental results show the superiority of the proposed 3D relative model compared to previous absolute models. And the user study also demonstrates the effectiveness and efficiency of the proposed approach.

Higher input image resolution matters for finger angle estimation. Compared to capacitive images which were widely used in previous research, fingerprints contain more useful information related to finger orientation, thus helping to promote the estimation accuracy and robustness. The resolution difference between fingerprint and capacitive images is so large (180 ppi versus 10 ppi in our study), and it is unlikely to reconstruct the original ridge pattern in fingerprint images by up-sampling capacitive images (such as [35]), which is necessary for estimating complete 3D finger angles. Intuitively, computational cost will increase with higher image resolution, resulting in longer latency, which is concerned in interactions. Compared to predict finger angles based on single input, more information is considered in the proposed relative approach, thus achieving a better performance under the same sensor conditions.

Estimating absolute finger angles presents different performance on different fingers. Zero finger orientation is difficult to define (especially for pitch) and not consistent across different fingers with diverse finger shapes, sizes, and habits of movements. Therefore, for those absolute finger angle models, estimation performance is different among different fingers due to the uncertainty of absolute finger angle ground truth, especially for thumbs whose shapes and sizes present large variance. While the proposed relative finger angle model concerns about the difference between two inputs (from the same finger) and does not suffer from the problem of zero orientation definition seriously.

Higher estimation consistency on different rotation ranges is achieved using the proposed relative finger angle model since relative information is involved. While for prior absolute finger angle models, estimation performance decreased when the actual rotation range is large. Due to the large estimation error on large rotation range, users tend to perform rotation in a small range for a more precise estimation, thus requiring more times of lifting fingers.

Roll angle is valuable in finger orientation based 3D manipulation tasks. Roll angle was rarely explored in previous works since it is difficult to infer in capacitive images. However, compared to pitch angle, roll angle presents larger controllable range and is easy to control. During manipulations in user study, we also found that some participants preferred to convert pitch rotation to roll rotation by first performing 90° yaw rotation.

Although the proposed method estimates 3D relative finger rotation based on fingerprint data, which is widely used in biometric recognition and contains sufficient identity information, privacy concerns may not be a major issue. Firstly, as demonstrated in Section 5.4, the proposed method can still perform well on fingerprint images with slightly lower resolution, from which the minutiae are difficult to be observed and extracted. Additionally, it is also possible to, like existing fingerprint-based identity recognition systems, prevent external access to fingerprint data and only allow access to the estimated 3D finger angles. This can help to avoid privacy issues caused by leakage of fingerprint data

8 LIMITATIONS

Although promising performance is achieved using fingerprint images for 3D relative finger orientation estimation, there are several limitations to be tackled.

Different from capacitive images, which are collected based on electrical capacitance changes when fingers touch on the display, fingerprint images are hard to capture for extreme finger angles, e.g. large pitch angle whose absolute value is over 80°. Therefore, estimating 3D extreme finger angles is less accurate, which limits the controllable range during interaction. This can be alleviated by relative finger orientation proposed in this paper. A fusion of fingerprint and capacitive images is likely to improve estimation accuracy for extreme finger angles further. Meanwhile, such fusion scheme can provide additional valuable information, e.g. capacitance changes while fingerprint not captured, for rejecting unintended touching, where interactions with extreme finger angles are performed in most cases.

As shown in Figure 5, the estimation performance is still unsatisfactory when the relative rotation angle is large, especially for roll angle which is inherently more difficult. This can be alleviated by selecting keyframes during interaction to split the 3D rotation procedure into several small discrete rotations, and combining them to obtain the final 3D finger rotation. Collecting more samples with large relative finger angles and designing more robust and efficient data augmentation strategies may also help.

Although simulated images with various resolutions were utilized as input in our experiments, we did not verify whether the proposed method can be applied to capacitive images. And we think that it is not feasible to estimate all three 3D finger angles simultaneously due to the limited

resolution of capacitive images. Besides, fingerprint images in our study were captured using FTIR-based fingerprint scanner, instead of under-screen fingerprint sensors in mobile devices. Due to the limitation of the Trusted Execution Environment (TEE)[3] in smartphones, fetching fingerprint images from mobile devices directly is not feasible without special support of fingerprint sensor and smartphone manufacturers. Therefore, considering the promising development of fingerprint based interaction techniques, these pioneer manufacturers are expected to make a step forward to facilitate the related research. Although the performance may drop on under-screen images due to lower quality, the reduction should not be more severe than the reduction observed in reducing image resolution, which means privacy concern may not be an issue since satisfactory precision can be achieved based on low resolution images erasing identity information already (Table 3).

Various simulated image resolutions are explored in our experiments, nevertheless, other factors are not explored systematically yet, such as wetness, dryness, and complicated skin distortion of fingers. And only males participated in our user study. We also noticed that our participants have various skin quality and the proposed model seems to be robust to skin quality.

Besides, the refresh rate of our method relies on the frame rate of capturing fingerprint images and processing time of network inference. The inference time of the proposed approach is about 0.0011 seconds per image pair on a PC with GeForce 1080 GPU. So the current bottleneck is the frame rate of fingerprint sensor (30 Hz), which was designed for person identification rather than interaction. For applications requiring very small latency, fingerprint sensors with higher frame rate but lower spatial resolution should be utilized. However, currently, such sensors are not available in the market. In addition, considering that the computing resources on mobile devices are relatively limited, apart from reducing the resolution of image acquisition, methods such as model quantization can also be employed to reduce the computational demand during inference on actual mobile devices.

Similar with capacitive images, both relative translation and rotation can be obtained based on fingerprint images. However, in this study, only finger angles were utilized in 3D object manipulation tasks since we focus on 3D relative finger rotation estimation in this paper. Six DOFs interactions based on fingerprint images can be further investigated in the future research. In addition, DOF separation and more advanced mapping strategies are not fully investigated in our user study, which have been demonstrated to be effective in previous studies [3, 10, 11, 30, 43]. Besides, similar with [3], elaborately comparing with previous 3D manipulation techniques, such as mouse+keyboard, tactile, and tangible inputs, to analyze strengths or weaknesses and explore suitable potential applications of different input techniques is also a promising research topic in the future.

9 CONCLUSION

Compared with absolute finger orientation, relative finger angles present superiority in several HCI applications. In this paper, we propose a 3D finger rotation estimation framework via 2D fingerprint images. Relative finger orientation, represented by three Euler angles, is directly estimated based on two input images, rather than calculating the relative transformation between two estimated absolute finger angles. Considering the characteristic of relative pose transformation varies with different initial finger angles and different fingers, metric learning is further utilized to constrain the distribution of extracted features to incorporate pose-relevant information. To explore the performance of the proposed approach, we collected a dataset consisting of fingerprint images with their corresponding ground truth 3D finger angles. The experimental results demonstrate the effectiveness and efficiency of our approach, and the proposed method is robust to different initial finger angles, finger types, and ranges of 3D rotation. Furthermore, a user study also revealed the

[3] https://source.android.com/security/trusty

superiority using the proposed 3D relative finger orientation in 3D object rotating task. We have also discussed the limitations of the current study. With accurate 3D finger rotation estimation, more innovative finger orientation based HCI interaction techniques would be possible in the future.

ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China under Grant 61976121 and 62376132.

REFERENCES

[1] Karan Ahuja, Paul Streli, and Christian Holz. 2021. TouchPose: hand pose prediction, depth estimation, and touch classification from capacitive images. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. 997–1009. https://doi.org/10.1145/3472749.3474801

[2] Avi Ben-Cohen, Roey Mechrez, Noa Yedidia, and Hayit Greenspan. 2019. Improving CNN training using disentanglement for liver lesion classification in CT. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 886–889. https://doi.org/10.1109/EMBC.2019.8857465

[3] Lonni Besançon, Paul Issartel, Mehdi Ammi, and Tobias Isenberg. 2017. Mouse, tactile, and tangible input for 3D manipulation. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 4727–4740. https://doi.org/10.1145/3025453.3025863

[4] Sebastian Boring, David Ledo, Xiang'Anthony' Chen, Nicolai Marquardt, Anthony Tang, and Saul Greenberg. 2012. The fat thumb: using the thumb's contact size for single-handed mobile interaction. In *Proceedings of the 14th International Conference on Human-computer Interaction with Mobile Devices and Services*. 39–48. https://doi.org/10.1145/2371574.2371582

[5] Cheng Chen, Qi Dou, Yueming Jin, Hao Chen, Jing Qin, and Pheng-Ann Heng. 2019. Robust multimodal brain tumor segmentation via feature disentanglement and gated fusion. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 447–456. https://doi.org/10.1007/978-3-030-32248-9_50

[6] Frederick Choi, Sven Mayer, and Chris Harrison. 2021. 3D hand pose estimation on conventional capacitive touchscreens. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction*. 1–13. https://doi.org/10.1145/3447526.3472045

[7] James Crowley, François Berard, Joelle Coutaz, et al. 1995. Finger tracking as an input device for augmented reality. In *International Workshop on Gesture and Face Recognition*. Citeseer, 195–200.

[8] Yongjie Duan, Ke He, Jianjiang Feng, Jiwen Lu, and Jie Zhou. 2022. Estimating 3D finger pose via 2D-3D fingerprint matching. In *27th International Conference on Intelligent User Interfaces*. 459–469. https://doi.org/10.1145/3490099.3511123

[9] Sovann En, Alexis Lechervy, and Frédéric Jurie. 2018. RPNet: An end-to-end network for relative camera pose estimation. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*. 738–745. https://doi.org/10.1007/978-3-030-11009-3_46

[10] Scott Frees and G Drew Kessler. 2005. Precise and rapid interaction through scaled manipulation in immersive virtual environments. In *IEEE Proceedings of Virtual Reality*. IEEE, 99–106. https://doi.org/10.1109/VR.2005.1492759

[11] Scott Frees, G Drew Kessler, and Edwin Kay. 2007. PRISM interaction for enhancing control in immersive virtual environments. *ACM Transactions on Computer-Human Interaction (TOCHI)* 14, 1 (2007), 2–es. https://doi.org/10.1145/1229855.1229857

[12] Alix Goguey, Géry Casiez, Daniel Vogel, and Carl Gutwin. 2018. Characterizing finger pitch and roll orientation during atomic touch actions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–12. https://doi.org/10.1145/3173574.3174163

[13] Lawrence Alan Gust. 2006. Compact optical pointing apparatus and method. U.S. Patent No. 7 102 617.

[14] Chris Harrison and Scott Hudson. 2012. Using shear as a supplemental two-dimensional input channel for rich touchscreen interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 3149–3152. https://doi.org/10.1145/2207676.2208730

[15] Chris Harrison, Julia Schwarz, and Scott E Hudson. 2011. TapSense: enhancing finger interaction on touch surfaces. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*. 627–636. https://doi.org/10.1145/2047196.2047279

[16] Ke He, Yongjie Duan, Jianjiang Feng, and Jie Zhou. 2022. Estimating 3D finger angle via fingerprint image. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 1 (2022), 1–22. https://doi.org/doi.org/10.1145/3517243

[17] Seongkook Heo, Dongwook Lee, and Minsoo Hahn. 2008. FloatingPad: a touchpad based 3D input device. In *International Conference on Artificial Reality and Telexistence*. 335–338. https://doi.org/10203/160076

[18] Ken Hinckley. 2002. Input technologies and techniques. In *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*. L. Erlbaum Associates Inc., 151–168. https://doi.org/10.5555/772072.772085

[19] Ken Hinckley, Joe Tullio, Randy Pausch, Dennis Proffitt, and Neal Kassell. 1997. Usability analysis of 3D rotation techniques. In *Proceedings of the 10th Annual ACM Symposium on User Interface Software and Technology*. 1–10. https://doi.org/10.1145/263407.263408

[20] Christian Holz and Patrick Baudisch. 2010. The generalized perceived input point model and how to double touch accuracy by extracting fingerprints. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 581–590. https://doi.org/10.1145/1753326.1753413

[21] Fang Hu, Peng He, Songlin Xu, Yin Li, and Cheng Zhang. 2020. FingerTrak: Continuous 3D hand pose tracking by deep learning hand silhouettes captured by miniature thermal cameras on wrist. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 2 (2020), 1–24. https://doi.org/10.1145/3397306

[22] Kaori Ikematsu and Shota Yamanaka. 2020. ScraTouch: Extending interaction technique using fingernail on unmodified capacitive touch surfaces. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 3 (2020), 1–19. https://doi.org/10.1145/3411831

[23] Sungyeon Kim, Minkyo Seo, Ivan Laptev, Minsu Cho, and Suha Kwak. 2019. Deep metric learning beyond binary supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2288–2297. https://doi.org/10.1109/CVPR.2019.00239

[24] Sven Kratz, Patrick Chiu, and Maribeth Back. 2013. Pointpose: finger pose estimation for touch input on mobile devices using a depth sensor. In *Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces*. 223–230. https://doi.org/10.1145/2512349.2512824

[25] Huy Viet Le, Sven Mayer, and Niels Henze. 2019. Investigating the feasibility of finger identification on capacitive touchscreens using deep learning. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. 637—-649. https://doi.org/10.1145/3301275.3302295

[26] Davide Maltoni, Dario Maio, Anil K. Jain, and Salil Prabhakar. 2009. *Handbook of Fingerprint Recognition*. Springer. https://doi.org/10.1007/978-1-84882-254-2_1

[27] Sven Mayer, Perihan Gad, Katrin Wolf, Paweł W Woźniak, and Niels Henze. 2017. Understanding the ergonomic constraints in designing for touch surfaces. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–9. https://doi.org/10.1145/3098279.3098537

[28] Sven Mayer, Huy Viet Le, and Niels Henze. 2017. Estimating the finger orientation on capacitive touchscreens using convolutional neural networks. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces*. 220–229. https://doi.org/10.1145/3132272.3134130

[29] Sven Mayer, Michael Mayer, and Niels Henze. 2017. Feasibility analysis of detecting the finger orientation with depth cameras. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–8. https://doi.org/10.1145/3098279.3122125

[30] Daniel Mendes, Filipe Relvas, Alfredo Ferreira, and Joaquim Jorge. 2016. The benefits of dof separation in mid-air 3d object manipulation. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*. 261–268. https://doi.org/10.1145/2993369.2993396

[31] Sundar Murugappan, Niklas Elmqvist, and Karthik Ramani. 2012. Extended multitouch: recovering touch posture and differentiating users using a depth camera. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*. 487–496. https://doi.org/10.1145/2380116.2380177

[32] Hyoungsik Nam, Ki-Hyuk Seol, Junhee Lee, Hyeonseong Cho, and Sang Won Jung. 2021. Review of capacitive touchscreen technologies: Overview, research trends, and machine learning approaches. *Sensors* 21, 14 (2021), 4776. https://doi.org/10.3390/s21144776

[33] Ian Oakley, Carina Lindahl, Khanh Le, DoYoung Lee, and MD Rasel Islam. 2016. The flat finger: exploring area touches on smartwatches. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 4238–4249. https://doi.org/10.1145/2858036.2858179

[34] Chang Peng, Mengyue Chen, and Xiaoning Jiang. 2021. Under-display ultrasonic fingerprint recognition with finger vessel imaging. *IEEE Sensors Journal* 21, 6 (2021), 7412–7419. https://doi.org/10.1109/JSEN.2021.3051975

[35] Narjes Pourjafarian, Anusha Withana, Joseph A Paradiso, and Jürgen Steimle. 2019. Multi-Touch Kit: A do-it-yourself technique for capacitive multi-touch sensing using a commodity microcontroller. (2019).

[36] Qualcomm. 2019. The world's largest Ultrasonic In-Display Fingerprint Sensor. https://www.qualcomm.com/products/3d-sonic-max.

[37] Simon Rogers, John Williamson, Craig Stewart, and Roderick Murray-Smith. 2011. AnglePose: robust, precise capacitive touch tracking via 3D orientation estimation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing*

Systems. 2575–2584. https://doi.org/10.1145/1978942.1979318

[38] Anne Roudaut, Eric Lecolinet, and Yves Guiard. 2009. MicroRolls: expanding touch-screen input vocabulary by distinguishing rolls vs. slides of the thumb. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 927–936. https://doi.org/10.1145/1518701.1518843

[39] Jeff Sauro and James R Lewis. 2010. Average task times in usability tests: what to report?. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2347–2350. https://doi.org/10.1145/1753326.1753679

[40] Florian Schroff, Dmitry Kalenichenko, and James Philbin. 2015. Facenet: a unified embedding for face recognition and clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 815–823. https://doi.org/10.1109/CVPR.2015.7298682

[41] Paul Streli and Christian Holz. 2021. CapContact: super-resolution contact areas from capacitive touchscreens. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14. https://doi.org/10.1145/3411764.3445621

[42] Yao-Hung Hubert Tsai, Paul Pu Liang, Amir Zadeh, Louis-Philippe Morency, and Ruslan Salakhutdinov. 2018. Learning factorized multimodal representations. *arXiv preprint arXiv:1806.06176* (2018). https://doi.org/10.48550/arXiv.1806.06176

[43] Manuel Veit, Antonio Capobianco, and Dominique Bechmann. 2009. Influence of degrees of freedom's manipulation on performances during orientation tasks in virtual reality environments. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*. 51–58. https://doi.org/10.1145/1643928.1643942

[44] Jonas Vogelsang, Francisco Kiss, and Sven Mayer. 2021. A design space for user interface elements using finger orientation input. In *Mensch und Computer 2021*. 1–10. https://doi.org/10.1145/3473856.3473862

[45] Feng Wang, Xiang Cao, Xiangshi Ren, and Pourang Irani. 2009. Detecting and leveraging finger orientation for interaction with direct-touch surfaces. In *Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology*. 23–32. https://doi.org/10.1145/1622176.1622182

[46] Yoichi Watanabe, Yasutoshi Makino, Katsunari Sato, and Takashi Maeno. 2012. Contact force and finger angles estimation for touch panel by detecting transmitted light on fingernail. In *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications*. Springer, 601–612. https://doi.org/10.1007/978-3-642-31401-8_53

[47] Robert Xiao, Julia Schwarz, and Chris Harrison. 2015. Estimating 3D finger angle on commodity touchscreens. In *Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces*. 47–50. https://doi.org/10.1145/2817721.2817737

[48] Ping-Hung Yin, Chih-Wen Lu, Jia-Shyang Wang, Keng-Li Chang, Fu-Kuo Lin, and Poki Chen. 2021. A 368× 184 optical under-display fingerprint sensor comprising hybrid arrays of global and rolling shutter pixels with shared pixel-level ADCs. *IEEE Journal of Solid-State Circuits* 56, 3 (2021), 763–777. https://doi.org/10.1109/JSSC.2020.3042894

[49] Qihao Yin, Jianjiang Feng, Jiwen Lu, and Jie Zhou. 2021. Joint estimation of pose and singular points of fingerprints. *IEEE Transactions on Information Forensics and Security* 16 (2021), 1467–1479. https://doi.org/10.1109/TIFS.2020.3036803

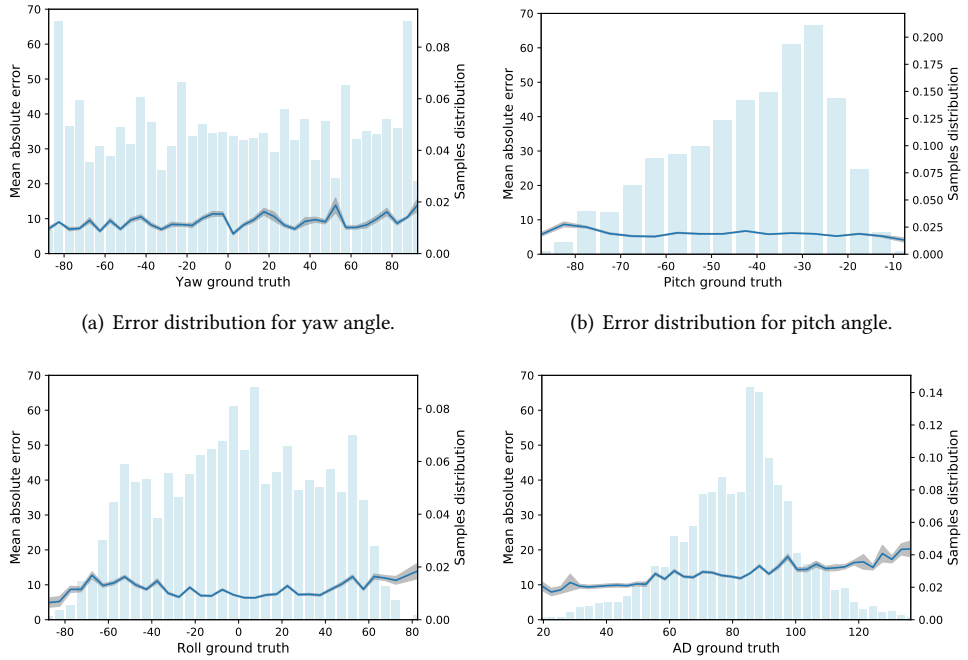
[50] Vadim Zaliva. 2012. 3D finger posture detection and gesture recognition on touch surfaces. In *2012 12th International Conference on Control Automation Robotics & Vision (ICARCV)*. IEEE, 359–364. https://doi.org/10.1109/ICARCV.2012.6485185

[51] Wenzhao Zheng, Jiwen Lu, and Jie Zhou. 2020. Structural deep metric learning for room layout estimation. In *European Conference on Computer Vision*. Springer, 735–751. https://doi.org/10.1007/978-3-030-58523-5_43

A SUPPLEMENTARY MATERIALS

A.1 Experimental Results

The error distribution of our method under different 3D absolute finger angle ground truth of the first frame is shown in Figure A1, which aims to explore the effects of initial finger angles on estimation performance. As shown in the figure, the proposed model is robust to 3D absolute finger angle of the first frame, initial finger orientation in other words.



(a) Error distribution for yaw angle.

(b) Error distribution for pitch angle.

(c) Error distribution for roll angle.

(d) Error distribution for AD.

Fig. A1. Error distribution of the proposed method for 3D absolute finger orientation (of the first one within a pair of images) represented by all three angles and angular distance (AD) in the test subset. 95% CI is shown as gray area.

A.1.1 Analysis of absolute finger orientation. The performance of approaches that calculate 3D relative finger angles based on 3D absolute finger angles, is affected by the estimation accuracy of single absolute finger angle. Therefore, performance of 3D absolute models, i.e. CNN in Mayer [28] and multi-task CNN [16], is also shown in Table A1. Similarly, the estimation performance increases with higher image resolution. The deeper network multi-task CNN [16] performs much better than the simple and shallow CNN model in Mayer et al. [28]. Meanwhile, based on the results of these models in Tables 3 and A1, we find that the prediction accuracy of relative finger angles decreases compared to absolute finger angles.

A.1.2 Analysis of rotation range. Compared with 3D absolute finger orientation based interaction, input range is enlarged using 3D relative finger angles, e.g. range of yaw angle extended from $[-90°, 90°]$ to $[-180°, 180°]$ and pitch angle from $[0°, 90°]$ to $[-90°, 90°]$. Intuitively, users tend to perform large rotation for rough manipulation quickly and achieve fine rotation by scaling.

Table A1. Quantitative results for 3D absolute finger angle models. Errors are reported in degrees.

Algorithm	PPI	Yaw			Pitch			Roll			AD		
		MAE	RMSE	SD	MAE	RMSE	SD	MAE	RMSE	SD	MAE	RMSE	SD
CNN in Mayer et al. [28]	10	23.61	38.29	30.15	10.17	12.85	7.86	20.43	28.44	19.79	33.09	47.36	33.89
	30	20.11	32.77	25.87	9.04	11.53	7.15	18.08	24.11	15.96	28.36	39.78	27.90
	180	15.51	22.19	15.87	8.10	10.09	6.01	15.67	20.24	12.81	22.59	27.52	15.71
Multi-task CNN [16]	30	17.84	30.92	25.26	8.81	11.22	6.95	17.14	23.97	16.75	25.29	37.07	27.10
	180	10.28	15.98	12.24	7.04	8.87	5.39	13.81	18.18	11.83	17.51	21.05	11.69

Therefore, robust prediction accuracy for different actual finger rotation angles is important in real applications. In Table A2, we show the 3D relative finger orientation estimation accuracy when the relative rotation angle ground truth is less than 90° and greater than 90° respectively. We observe that the proposed 3D relative model perform robustly on different rotation ranges, making it possible to utilize large relative finger pose transformation for input interaction in real applications.

Table A2. Quantitative results for 3D relative finger orientation estimation models when actual rotation angle is smaller than 90° and greater than 90° respectively. Default image resolution is 180 ppi. Errors are reported in degrees.

	Algorithm	Yaw			Pitch			Roll			AD		
		MAE	RMSE	SD	MAE	RMSE	SD	MAE	RMSE	SD	MAE	RMSE	SD
AD≤ 90°	CNN in Mayer et al. [28][1]	27.90	45.87	36.41	14.83	20.02	13.46	15.85	21.68	14.79	37.61	51.06	34.53
	CNN in Mayer et al. [28]	16.90	24.14	17.23	10.78	14.02	8.97	12.25	16.51	11.06	24.44	29.10	15.79
	Multi-task CNN [16]	11.46	17.95	13.81	8.53	11.43	7.61	10.30	14.04	9.54	18.32	22.63	13.28
	Naïve Siamese	8.94	**12.65**	**8.94**	7.45	9.24	5.46	8.66	11.47	7.52	14.25	**16.26**	**7.82**
	Pose-aware Siamese	**8.55**	14.87	12.17	**6.20**	**8.02**	**5.09**	**7.47**	**10.49**	**7.36**	**13.42**	17.72	11.58
AD> 90°	CNN in Mayer et al. [28][1]	44.06	65.85	48.94	17.01	23.31	15.94	21.85	30.89	21.83	50.40	68.00	45.65
	CNN in Mayer et al. [28]	27.37	41.36	31.00	11.24	14.80	9.63	17.50	26.26	19.59	31.26	40.25	25.35
	Multi-task CNN [16]	16.07	25.77	20.15	8.93	11.84	7.77	12.96	19.05	13.96	20.90	26.84	16.85
	Naïve Siamese	10.94	16.42	12.24	7.29	9.22	5.66	11.03	15.52	10.91	15.27	17.60	**8.76**
	Pose-aware Siamese	**9.66**	**14.79**	**11.19**	**5.85**	**7.69**	**4.99**	**9.23**	**13.41**	**9.74**	**13.49**	**16.25**	9.06

[1] down-sample the input fingerprint image to 10 ppi for capacitive image simulation.

Received 2023-02-15; accepted 2023-05-30