Accurate and Low-Latency Sensing of Touch Contact on Any Surface with Finger-Worn IMU Sensor

Yizheng Gu[12], **Chun Yu**[12†], **Zhipeng Li**[2], **Weiqi Li**[2], **Shuchang Xu**[12], **Xiaoying Wei**[12], **Yuanchun Shi**[12]
[1]Key Laboratory of Pervasive Computing, Ministry of Education, China
[2]Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China
{guyz17,zp-li16,wq-li16,xusc18,wei-xy17}@mails.tsinghua.edu.cn, {chunyu,shiyc}@tsinghua.edu.cn

ABSTRACT

Head-mounted Mixed Reality (MR) systems enable touch interaction on any physical surface. However, optical methods (i.e., with cameras on the headset) have difficulty in determining the touch contact accurately. We show that a finger ring with Inertial Measurement Unit (IMU) can substantially improve the accuracy of contact sensing from 84.74% to 98.61% (f1 score), with a low latency of 10 ms. We tested different ring wearing positions and tapping postures (e.g., with different fingers and parts). Results show that an IMU-based ring worn on the proximal phalanx of the index finger can accurately sense touch contact of most usable tapping postures. Participants preferred wearing a ring for better user experience. Our approach can be used in combination with the optical touch sensing to provide robust and low-latency contact detection.

Author Keywords

Mixed reality, head-mounted display, smart ring, touch interaction.

CCS Concepts

•**Human-centered computing** → Gestural input;

INTRODUCTION

MR (Mixed Reality) technologies, such as Hololens and MagicLeap, bring rich possibilities for novel human-computer interaction paradigms. With the depth camera sensing the physical environment (including users' hands) and the 3D glass rendering virtual elements, mixed reality in principle enables interaction anywhere. One promising and valuable setting is to project virtual user interface on an arbitrary physical surface, and allow users to interact with direct finger touch. This extends "touch" – the most usable input method of human beings – which is now restricted to digital touchscreen devices to any physical surface. Compared with mid-air interaction, MR-enabled surface interaction can provide "real" haptic feedback that is an essential component of natural touch experience. It can also capture rich information of the tapping finger and hand (e.g., finger identification and posture) with the headset camera. These advantages all together provide a great potential to augment touch input in the future.

To sense touch, it is straightforward to leverage the cameras on the MR headset. However, optical methods have inherent drawbacks for detecting touch contact: First, with the camera looking behind the tapping finger, it is difficult to accurately detect when the finger contacts the surface. Second, optical methods usually require considerable processing and introduce a latency of variant length. For instance, the state-of-the-art work exploring touch sensing with depth cameras of Hololens [50] reported a high rate of both missed touches (3.5%) and spurious extra touches (19.0%), and a system latency of about 180 ms. In literature, numerous works have been carried out to study and improve contact sensing [5, 31], emphasizing the importance of delay [21, 10, 32] and spatial accuracy [18, 43, 2, 6] on touch experience. Therefore, camera-based contact sensing does not provide a satisfying solution.

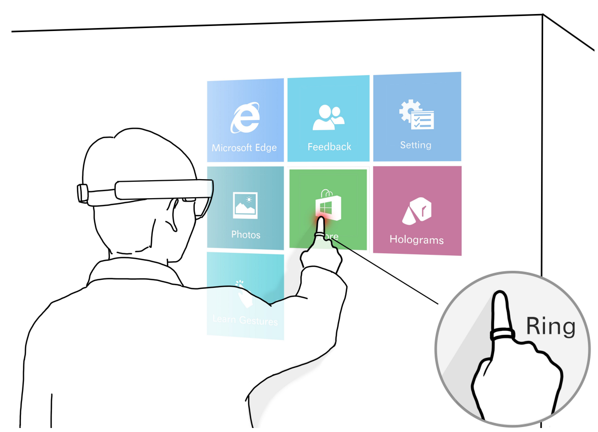

Figure 1: Our envisioned use scenario of mixed-reality interaction on any surface.

To address this problem, we envision combined use of an MR headset and a smart ring (Figure 1) in the future. The camera on the headset is responsible for detecting finger location and posture, while the smart ring, embedded with an IMU (Inertial

† denotes the corresponding author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

UIST 2019, Dec 20–23, 2019, New Orleans, LA, USA

© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ISBN 978-1-4503-6708-0/20/04. . . $15.00

DOI: https://doi.org/10.1145/3313831.XXXXXXX

Measurement Unit) sensor, is responsible for detecting touch contact. The first advantage of this setting is that an IMU sensor worn on the finger can directly detect the sudden finger contact on the surface. Second, processing IMU data is usually efficient, which ensures low latency. To our knowledge, prior works have explored the possibilities of using finger-worn IMU sensors to augment touch input (e.g., finger tracking [24, 34] and finger identification [29]). We acknowledge using a finger-worn IMU sensor to detect touch is not new [24]. But we are surprised to see that no research in literature has been conducted to optimize the accuracy and latency of touch contact detection, which is essential for the natural touch experience.

In this work, we investigate sensing touch contact with a finger-worn IMU sensor in the context of MR-enabled surface interaction. We are interested in identifying comfortable ring wearing positions preferred by users, and the associated accuracy and latency for sensing various tapping postures (e.g., using an IMU sensor worn on the index finger to sense tapping of the middle finger). Our results suggest that an IMU sensor worn on the proximal phalanx of the index or middle finger provides the best user preference and sensing capability: The F1 score can be as high as 98.61% (Precision = 98.61%, Recall = 98.62%), while the detection latency can be as low as 10ms. The empirical results obtained in our research provide practical guidelines on deploying an IMU-based smart ring to optimize the touch experience on any surface.

Specifically, the contributions of this work are four-fold.

- We investigated user preference on the tapping postures and the ring placements.
- We identify a set of usable hand postures during tapping and validated the feasibility of recognizing them with optical methods.
- We empirically demonstrate that the SVM-based method substantially outperforms traditional threshold-based method for sensing touch contact in terms of accuracy.
- We find the best ring wearing positions to be on the proximal phalanx of the index or middle finger in terms of both user preference and sensing capability.

RELATED WORK

Touch input on surfaces

Touch is the most common input method for modern handheld devices [9], e.g., smartphone and touchpad. However, most current devices provide touch sensing by instrumenting the surface itself, e.g., with capacitive [25, 44], optical [13, 30, 45] and acoustic [36, 48] sensors. It is not practicable to support anywhere touch by changing the whole environment.

Cameras allow touch sensing without instrumenting the surfaces. Several optical schemes have been proposed for touch sensing in the literature, including LIDAR [35], RGB cameras [27, 8, 1, 42], infrared cameras [23] and thermal cameras [39]. The recent emergence of inexpensive depth cameras has led to a wide research interest in touch sensing techniques based on depth cameras. Researchers started to focus on interaction design [1, 47, 50] and the spatial accuracy of touch sensing [46, 4, 14]. These approaches required fixing the cameras in the lab environment or using wearable cameras [50, 14].

However, optical touch sensing has difficulty in determining whether a finger has contacted the surface or not [47, 50]. Most optical techniques use threshold method to sense contact [1, 46, 14, 20, 47, 50]. For example, a contact is declared if the distance between fingertip and surface descends below 10 mm, and ended if the distance ascends past 15 mm. This method is not robust enough. First, the contact sensing can be affected by the noise, delay and occlusion of cameras. Second, the thresholds force users to control the hand carefully to avoid accidental touch.

From an overhead view, solving the contact problem is hard for optical touch sensing. Therefore, a robust contact detection is required.

Contact sensing based on vibration

Touch generates vibration and sound, which can be used to sense touch interactions. Some works use the sensors on devices for detection [17, 19, 22, 36, 48], while the others place sensors on the fingers.

To our knowledge, prior work on finger-worn sensor did not focus on contact sensing or achieve a satisfying recognition accuracy. They focused on finger tracking (relative motion) [24, 34], touch finger identification [29] or touch surface identification [41], but neglected the quality of contact sensing. They used simple threshold methods to sense contact [24, 34, 33], yielding an accuracy of up to 89.8%.

In this paper, we used an optical method to track the fingers and focused on contact sensing based on IMU ring. An accurate and low-latency contact sensing technique is crucial for optical touch sensing, and can naturally complement the ring interactions above as well.

Tapping postures

There have been some approaches to enrich the input vocabulary of touch. For a conventional touch screen, the spatial and temporal relationship of touches is used to extend the touch interaction, e.g., tap-and-hold gestures [11] and multi-fingers interaction [26]. Researchers built Ad hoc devices to enrich touch input, for examples, by adding pressure [37], velocity [17, 19], tangential force [16] and finger orientation [49, 38].

In the scenarios of head-mounted AR systems, vision information is available. A straightforward way to enrich touch interaction is to identify tapping postures (e.g., which fingers and which part of the fingers touch the surface). Prior work has shown the value of recognizing tapping postures in touch interaction [15, 7]. In this paper, our contact sensing algorithm also supports different tapping postures.

EXP. 1A: USER PREFERENCE OF TAPPING POSTURE

We conducted this experiment to collect tapping postures that most users are willing to use in daily routines.

We first defined a comprehensive set of tapping postures. For each posture in the set, we asked participants to perform the posture and then rate it in a questionnaire. We chose the most popular postures according to their ratings.

Tapping postures set

We focused on the tapping postures at the moment of contact. As figure 2 shows, we explored tapping postures in a three-dimensional design space:

- **Which fingers touch the surface?** The *thumb*, the *index*, *middle*, *ring*, *pinkie* fingers, *two fingers* and *three fingers*.
- **Which part of the finger touches the surface?** We refer to TapSense [15] to explore this dimension. Users may touch with the *pad*, *tip*, *knuckle*, *side* or *nail* of a finger.
- **Posture of the non-touching fingers.** While some fingers touch the surface, the others could be in a *closed fist* position or in an *open palm* position.

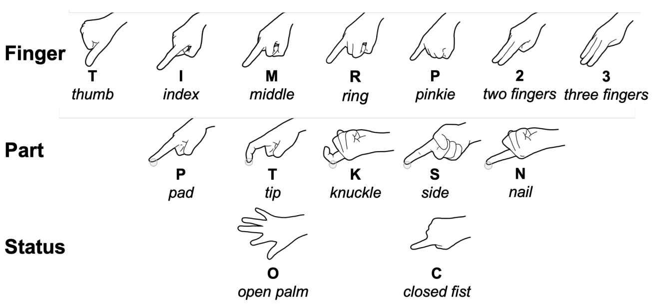

Figure 2: Three-dimensional design space for tapping postures. The bold labels indicate the abbreviation of each condition.

Finally, we had $7 \times 5 \times 2 = 70$ types of tapping postures in our set. We defined abbreviation for them, e.g., *IPO* for touching with the index finger pad (open palm).

Design

We recruited 20 participants from the campus (7 females; aged from 18 to 27, M = 22.0). The experiment has two sessions of surface orientations: a horizontal desk and a vertical wall. They are common surface orientations in our daily life. We counter-balanced the order of orientations across participants.

Participants had to touch in $70 \times 2 = 140$ conditions. They rated for each tapping posture through a questionnaire. The questionnaire evaluated three aspects of each tapping posture on a 7-point Likert Scale:

- **Comfort:** the physical and mental ease of performing the posture (1 - not easy, 7 - very easy).
- **Memory:** the ease of remember the posture (1 - not easy, 7 - very easy).
- **Preference:** the willingness to use the tapping posture (1 - not at all, 7 - very much).

In the end of the experiment, we conducted an brief interview for the concerns below:

- Is there any available tapping posture outside our set?
- How many different tapping postures are you willing to identify in daily use.

Procedure

During the experiment, the participant sat on an adjustable chair. We asked the participant to adjust the chair so that he can touch in the most comfortable position.

For each tapping posture, the experimenter first demonstrated it. The participant then performed the posture in person for two or three times and then rated it in the questionnaire. After each touch, the participant was allowed to modify the previous ratings through comparison.

The participant rested for five minutes every ten tapping postures. The whole experiment lasted for one hour.

Result

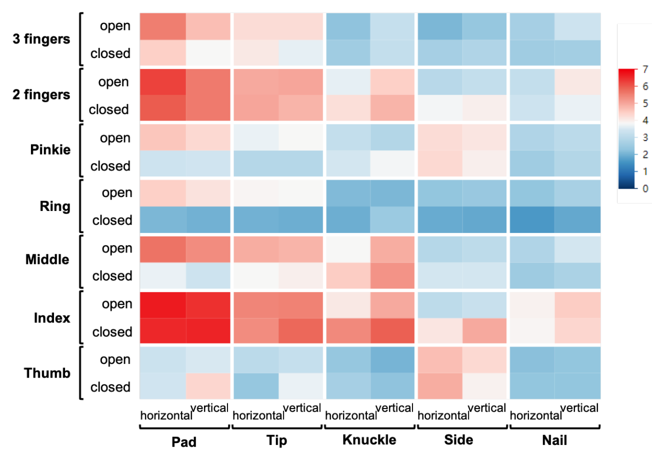

Figure 3: User preference of different tapping postures (1 - worst; 7 - best).

Figure 3 shows participants' preference for all 70 postures in both horizontal and vertical conditions. The top ten postures were *IPO*, *IPC*, *2PO*, *2PC*, *ITO*, *ITC*, *MPO*, *IKC*, *2TO* and *3PO* (Figure 4). Friedman test found no significant effects of orientation on user preference to the ten postures. In the interview, no participant reported available postures outside our set. Thus, we deemed these ten postures as the most popular postures in touch interaction.

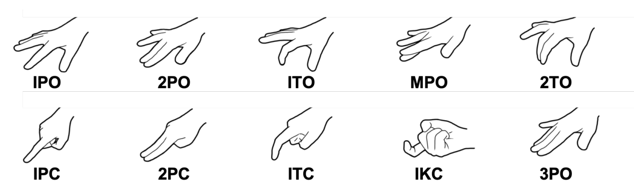

Figure 4: The top ten tapping postures and their abbreviation. Please see their ratings in appendix I.

Friedman test showed significant effects of touch finger ($\chi^2 = 767.70, p < .0001$) and finger part ($\chi^2 = 423.86, p < .0001$) on subjective preference. Participants preferred to touch with the index, middle fingers, two fingers and three fingers. Participants accepted only touching with the pad, tip and knuckle of a finger. Other conditions could be excluded from the touch interaction design.

Participants reported that they are willing to identify 7.45(SD=2.61) postures in average. Thus, we deemed that the ten popular tapping postures were enough for the follow-up research.

EXP. 1B: USER PREFERENCE OF RING PLACEMENT

We conducted this short experiment to investigate user preference of ring placement (Figure 5(a)) on a 7-point Likert Scale (1 - worst; 7 - best):

- **Comfort**: the physical and mental ease of performing touch interaction with a ring on this position.
- **Acceptance**: the willingness to wear the ring on this position in daily life.
- **Preference**: the willingness to perform touch interaction with a ring on this position.

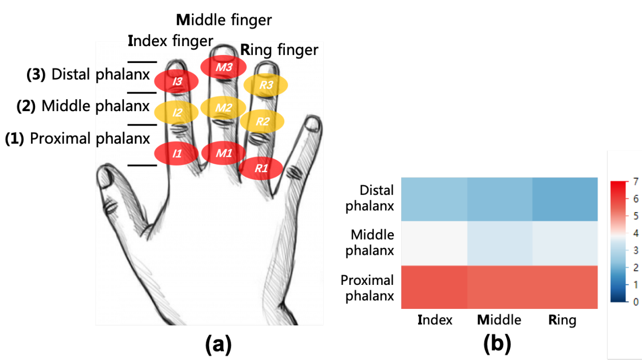

Figure 5: (a) shows different positions to wear a ring. We defined abbreviation for each position, e.g., *I1* for the first phalanx of the index finger. The red color indicates the tested ring placement in experiment two. (b) is the user preference of ring positions (1 - worst; 7 - best).

The twelve participants in experiment 1A attended this experiment. They should touch for several times with a normal ring before they could rate the preference.

Figure 5(b) shows that participants prefer to wear the rings on *I1* (5.65), *M1* (5.45) and *R1* (5.45). Touch with the ring worn on these positions is comfortable (5.40 ↑) and acceptable (5.32 ↑).

EXP. 2: TOUCH DATA COLLECTION

In this experiment, we sampled motion and camera data that the participant touches with an IMU ring. The motivation was to provide data for two follow-up works. The first was to evaluate the identification of tapping postures based on camera. The second was to design the contact sensing algorithm based on IMU ring.

Design and procedure

We recruited twelve participants from the campus (4 females; aged from 20 to 29, M = 23.1). The experiment had two sessions of surface orientations (horizontal and vertical). We counter-balanced the surface orientation across participants.

Each session consisted of five blocks. The participant wore the IMU ring on five different positions (Figure 5): *I1*, *M1*, *R1*, *I3* and *M3*. Experiment 1B shows that users prefer to wear the rings on *I1*, *M1* and *R1*. We added also *I3* and *M3* because an IMU sensor on fingertip may detect a stronger vibration.

Each block consisted of ten trials. The participant touched the surface for 20 times with the ten popular tapping postures (Figure 4). Participants were asked to touch in a natural way. Each participant performed $2 \times 5 \times 10 \times 20 = 2000$ touches in total.

Then, we collected mid-air gestures as negative samples. The participant wore the IMU ring on different positions and performed gestures such as drawing circle, swiping and Hololens gestures. The participant was not allowed to collide his fingers (e.g., pinch). The sampling of each ring position lasted for one minute.

Similar to the first experiment, we asked the participant to touch in a natural way in the most comfortable position. The participant rested for two minutes after every 200 touches. The whole experiment lasted for one and a half hour.

Apparatus

Figure 6 illustrates the experimental apparatus. The participant wore an IMU ring and a head-mounted leap motion. He touched on a low-latency touch screen. During the experiment, we sampled acceleration and angular velocity data from the IMU ring, the skeleton of hand from leap motion, and contact conditions from the touch screen.

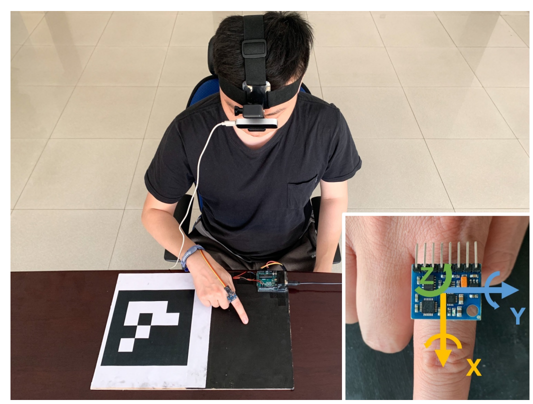

Figure 6: The experimental setting in experiment two. A participant touched on the touch screen with an IMU ring. The subfigure shows the coordinate of the IMU ring.

The IMU ring was a 9-axis accelerator GY-85 attached to a regular finger ring. We made several IMU rings to fit different finger sizes. The ring connected to an Arduino Uno R3 with Dupont lines. We attached the Dupont lines on the user's wrist with a velcro strap.

The touch screen was a wooden board covered with conductive ink. The capacitance of the board increases when a finger touches on it. We leveraged this phenomenon to judge the contact condition [3]. Analysis of high-speed camera data showed that the latency of the touch screen was below 5 ms.

The touch screen was also connected with the Arduino so that it shared the same timestamps with the IMU sensor. The frequency was 200 Hz.

Leap motion sampled the positions and orientations of the palm and all finger joints. We placed a marker on the plane of the touch surface to calculate the distance between each joint and the surface. The frame rate of leap motion was 60 Hz. The latency between camera data and the Arduino was about 20 ms. We controlled the light condition (bright; avoiding sunlight) to ensure the sensing quality. The touch screen was black in IR images, which was a perfect background for Leap motion.

Result

The experiment collected $12 \times 2000 = 24000$ raw positive samples. We used an interactive program to remove wrong data, for example, when the fingernail contacts earlier than the finger pulp does, the capacitive screen can not detect the contact in time. Finally, we held more than 23900 positive samples.

We randomly sampled negative samples from the mid-air gestures. The numbers of positive and negative samples are the same.

TAPPING POSTURE CLASSIFICATION

In this session, we evaluate the identification of tapping postures based on optical method. The motivation was to verify the feasibility. The classification method was for evaluation but not our contribution.

Method

We referred to [52] to extract hand skeleton features, including fingertip distances, adjacent fingertip distances, fingertip elevations, and fingertip angles. These values were concatenated to be a hand shape feature of 19 dimensions. We trained a SVM model for the classification.

Result

We used leave one out cross-validation to evaluate the posture classification (Table 1). The four classification of *IPO*, *IPC*, *2PC* and *IKC* achieved an accuracy of 99.0%. The accuracy of identifying seven postures (*2PO*, *MPO* and *3PO* added) was acceptable (88.5%). The classification of ten postures was not satisfying yet.

	4 classes	7 classes	10 classes
Horizontal	99.1%(1.3%)	89.5%(3.9%)	76.4%(6.8%)
Vertical	99.0%(1.4%)	87.6%(4.8%)	77.6%(6.7%)

Table 1: Average classification accuracy for four, seven and ten tapping postures in leave-one-out cross validation. Standard deviations were showed in parenthesis.

The result shows that head-mounted leap motion can robustly identify four to seven tapping postures. With the development of hand tracking techniques [12, 40, 51], we argue that enhancing touch interaction with various tapping postures will be feasible soon.

TOUCH CONTACT SENSING

In this session, we designed a contact sensing algorithm based on IMU ring. The aim was to sense the contact of various tapping postures with a low-latency.

We have three conclusions in this session. First, observations suggest that available information from IMU ring is rich. A machine learning method can largely improve the accuracy of contact sensing compared with prior threshold methods. Second, it is the best choice to wear the ring on the proximal phalanx of the index or middle finger. The two ring positions optimize the performance of recognition and are most preferred by users. Third, a significant vibration can transmit to any ring position within 20 ms. Thus, the latency of IMU-based contact sensing can be low.

Observation

The raw data of the accelerator was fused with gravity. We used a filter [28] to split the raw acceleration into true acceleration and gravity. In total, we had nine dimensions of motion data (3-axis acceleration, 3-axis angular velocity and 3-axis gravity).

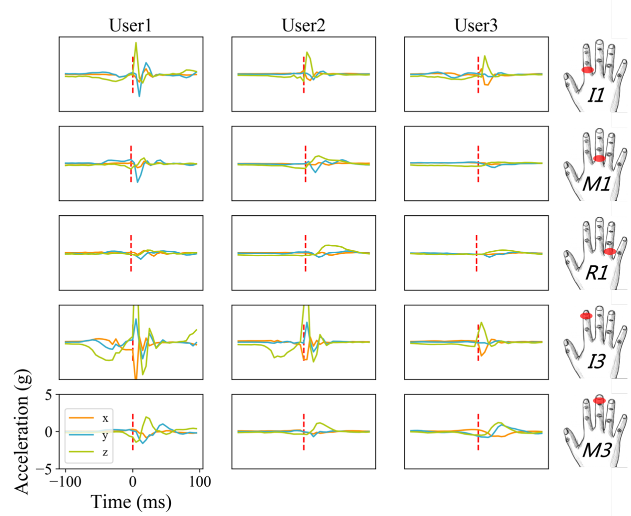

Figure 7: Illustration of acceleration data over users and ring positions. Multiple features such as *mean*, *minimum*, *maximum*, *skewness* and *kurtosis* could be valuable to describe the patterns.

We use the tapping posture *IPO* as an example to illustrate the data. Figure 7 shows the acceleration data of different users and ring positions. The acceleration reached a peak within 30 milliseconds after a contact. We speculated that the peak was caused by the collision at the touch moment.

For each ring wearing position, the detected patterns of acceleration among users were similar. We inferred that multiple features such as *maximum*, *minimum*, *mean*, *skewness* and *kurtosis* could be helpful to the contact sensing. For examples, the ring on *I1* (Figure 7, Row 1) detected strong vibration on z-axis, so *maximum* could be a good feature here; the ring on

M1 (Figure 7, Row 2) detected peaks in the same direction and duration on y-axis, so *skewness* and *kurtosis* were also valuable to describe the patterns.

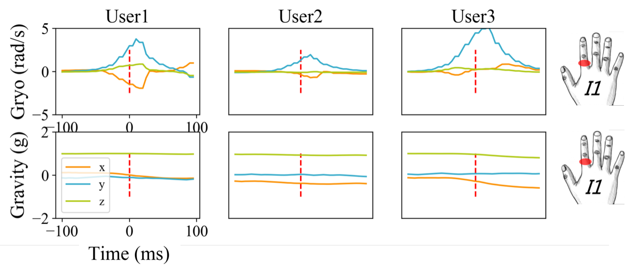

Figure 8: Illustration of the gyroscope and gravity data over users. We inferred that features extracted from gyroscope and gravity data can contribute to the contact sensing.

Figure 8 illustrates the angular velocity and the gravity over users. These patterns were regular. For example, the gravity data was similar for all the users. It indicates that different users touch with a near orientation to the surface. We inferred that both the angular velocity and the gravity can contribute to the contact sensing.

The result shows that available information from IMU ring is rich. Prior work used threshold on a single feature (e.g., acceleration [34, 29] or sound [33]) to sense touch contact. In this paper, we decided to extract multiples features from the IMU ring and use SVM for the classification.

Classifier

We extracted features from a time window of ten frames (50 ms). For each dimension of the 9-axis IMU data, we calculated its *maximum*, *minimum*, *mean*, *skewness* and *kurtosis*. Then, we concatenated these values to obtain a feature of 45 dimensions.

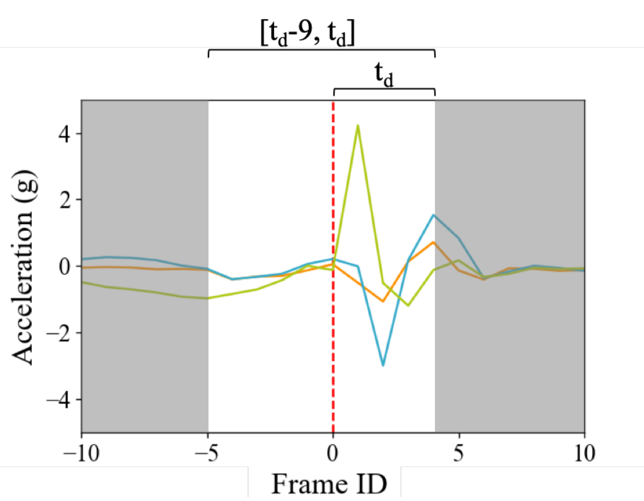

Figure 9: The acceleration data of a positive sample. Model latency t_d indicated how we choose the time windows of samples for the training.

It was a problem how to choose the time window for the training because it takes an unknown time for the vibration of touch to transmit to the IMU ring. We **defined** $t_d (0 < t_d < 10)$ as *Model Latency* (Figure 9) of a classifier, when the classifier was trained by samples in the time windows $[t_d - 9, t_d]$. There was a trade-off: the larger t_d is, the more accurate the classifier will be, but the recognition delay may also increase. Thus, we had to test different model latency to find an optimal one.

Given a model latency t_d, we trained the classifier as follow. We extracted features from time window $[t_d - 9, t_d]$ as positive samples. To avoid reporting the contact in advanced, we extracted features from window $[-14, -5]$ as negative samples. Also, we extracted negative sample features from the mid-air gestures. Finally, we ran SVM to train the classifier.

Optimization of the Classifier

Model Latency

Figure 10 illustrates the enhancement of accuracy over model latency. Mixed ANOVA showed significant effects of model latency ($F_{3,33} = 133.4, p < .0001$) in the first 20 ms. After the first 20 ms, the curves start to converge ($F_{2,22} = 0.011, p = 0.99$). The result shows that the contact sensing performs the best with a model latency of 20 ms (at most 99.3%).

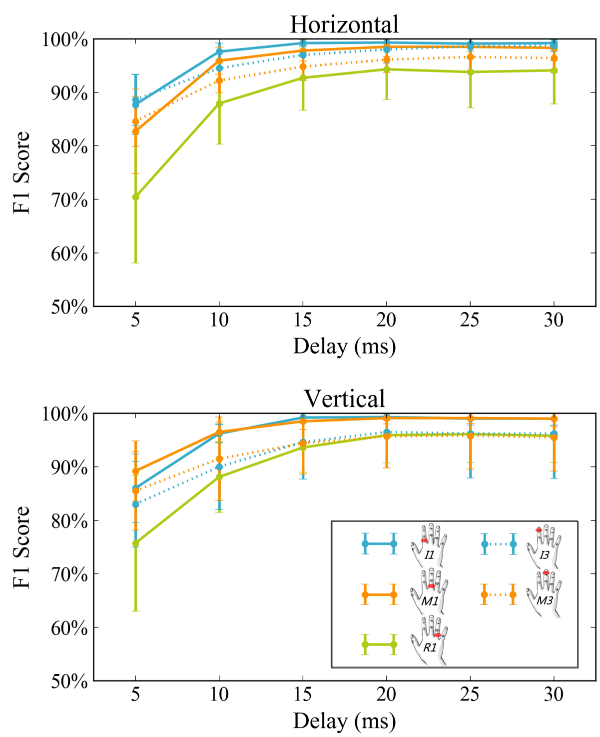

Figure 10: Average f1 score of contact detection over model latency. Error bars represent standard deviation.

The high recognition accuracy also indicates that the vibration of touch can transmit to any ring position in 20 ms. Thus, the IMU ring can be a low-latency approach to sense touch contact.

	Ring Position	I1	M1	R1	I3	M3
Horizontal	Precision	99.7%(SD=0.5%)	99.2%(1.0%)	97.6%(1.9%)	99.1%(1.3%)	98.3%(1.4%)
	Recall	98.9%(1.6%)	97.9%(3.0%)	91.7%(9.2%)	97.1%(4.4%)	94.1%(4.1%)
	F1 Score	99.3%(1.0%)	98.5%(1.8%)	94.3%(5.6%)	98.0%(2.5%)	96.1%(2.5%)
Vertical	Precision	99.7%(0.6%)	99.3%(1.1%)	98.1%(1.5%)	98.3%(2.1%)	98.4%(1.9%)
	Recall	99.0%(0.9%)	98.8%(1.8%)	94.0%(8.4%)	95.4%(10.8%)	93.7%(9.8%)
	F1 Score	99.3%(0.6%)	99.1%(1.1%)	95.9%(5.3%)	96.5%(6.7%)	95.7%(5.9%)

Table 2: Average accuracy of contact sensing over ring positions (model latency = 20 ms).

Ring Position

Table 2 shows the accuracy over different ring positions (model latency = 20 ms). We considered both horizontal and vertical conditions in the following comparison. The classifier performed the best with the ring worn on *I1* (99.3%). The next was *M1* (98.8%). RM ANOVA showed a significant effect of ring position ($F_{4,44} = 6.45, p < .001$) but no significant effect of surface orientation ($F_{1,11} = 0.09, p = .76$) on f1 score. Results showed *I1* significantly better than *R1*($p < .001$), *I3*($p < .005$) and *M3*($p < .005$); *M1* significantly better than *R1*($p < .005$), *I3*($p = .046$) and *M3*($p < .01$).

The results of ring positions *I3* and *M3* were not as good as expected. We found two reasons. First, the IMU ring on *I3* or *M3* could indeed detect a stronger vibration, but the noise was also enlarged. Second, the vibration generated by a finger could not well transmit to the tips of other fingers.

As the first experiment implicates, users prefer to wear the ring on *I1*, *M1* and *R1*. Thus, we recommend to wear the ring on *I1* or *M1* (the proximal phalanxes of the index and middle fingers). These two positions performs the best in both recognition accuracy and user preference.

Evaluation

The results above show that the classifier performed the best with a model latency of 20 ms and with the IMU ring on *I1*. We present the evaluation on this setting.

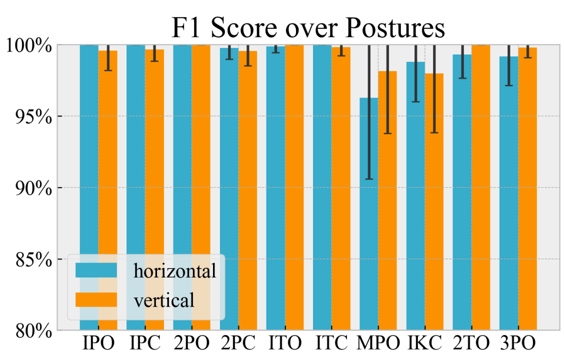

Figure 11: Average f1 score of contact sensing over tapping postures. Error bars represent standard deviation.

Figure 11 shows the f1 scores of contact sensing over different tapping postures. The accuracy exceeded 95% when the user performs tapping postures *MPO* and *IKC*. The accuracy was nearly 100% for other tapping postures.

To evaluate our contact sensing over other methods, we implemented two baselines for comparison:

- The first was the threshold method based on accelerator data [24, 34]. We ran a simulation to find the optimal threshold for each setting. Taking the setting of *I1* and *Horizontal* for example, we found $\|A_z\|$ as the best identifier, where A_z is the z-axis acceleration. A threshold of $1.08G$ optimizes the accuracy.

- The second baseline was based on vision, which declares a contact when the distance between fingertip and surface declines below 10 mm [46, 20, 47, 50].

Figure 12 shows the comparison between our method and the two baselines. ANOVA shows that our method significantly improved the precision ($F_{1,11} = 10.4, p < .001$) and the recall rate ($F_{1,11} = 59.8, p < .0001$; $F_{1,11} = 124.7, p < .0001$) of contact sensing.

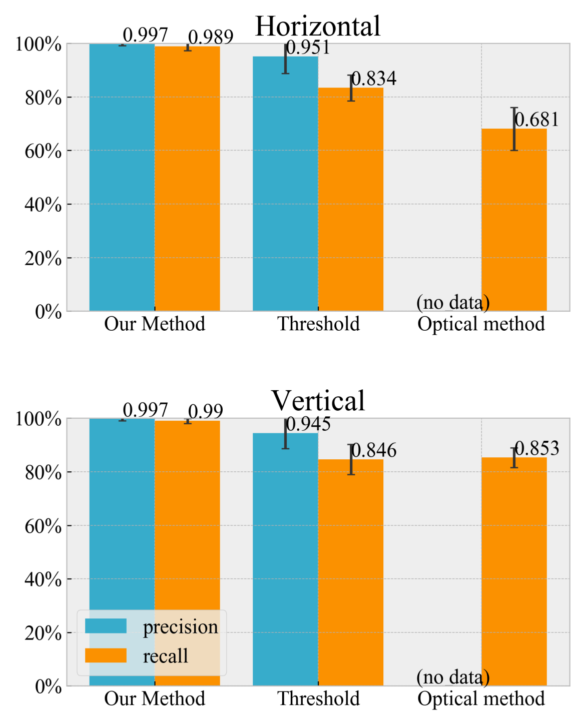

Figure 12: Average precision and recall rates of contact sensing (our method vs. baselines). Error bars represent standard deviation. We has no negative samples to evaluate the precision of the vision method.

Contact Sensing Algorithm

The contact classifier was not enough for sensing contact in runtime. First, it would trigger repeated touch events when touching. Second, though the prediction accuracy was as high as 99.3%, it would still cause spurious extra touches in the continuous runtime. To address these problems, we designed a contact sensing algorithm based on the contact classifier.

- The algorithm do not reports touch event if there has been a contact in the past ten frames (50 ms).
- The algorithm reports touch event only if the classifier detects two consecutive frames of contact.

The two statements above lead to another frame of delay. However, they can greatly reduce spurious extra touches and report only one event for each contact. In the next experiment, we evaluated the contact sensing algorithm by a real application.

Discussion: Why Machine Learning?

The reason why machine learning beat threshold methods is that multiple features are valuable. A single feature can not robustly detect touch contact. Here are some examples:

- Threshold methods failed in the case of soft tapping. In this case, the kurtosis of acceleration was a good feature for the machine learning method. Our method works no matter weather the tapping is soft or hard.
- When performing *IPC* with the ring worn on *M1*, the ring did not vibrate in a usual direction (z-axis), which sometimes made threshold methods fail. The combination of gravity and acceleration was helpful for the machine learning method.
- Mid-air tapping mostly led to false positives with threshold methods. For machine learning, multiple features are helpful to reject these false positives.

EXP. 3: EVALUATION

In this experiment, we evaluated our contact sensing algorithm by real application and compared it with the optical method.

Design and procedure

We recruited twelve participants (3 females; aged from 20 to 28, M = 23.2). The participant touched on the low-latency touch screen as in the last experiment. The touch screen provided ground truth for the evaluation.

The task was the "Piano Tiles" game (Figure 13). We presented the game on a regular display. The participants could see his virtual hand in the game scene. The control display ratio was 1. The participant's objective was to tap on the black tiles as they appeared from the top of the screen while avoiding the white. The screen moved manually, at the rate which the tiles were touched. If the participant tapped on a white tile, the screen would flicker to inform the error.

We had two sessions in the experiment. In session one, we compared our contact sensing with the optical method. The participants touched on a horizontal touch screen with the two techniques. They touched with the most common postures

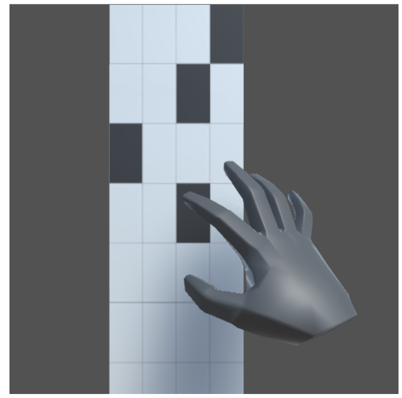

Figure 13: The experiment task: "Piano Tiles".

(*IPO* or *IPC*). The participant touched 100 black tiles to finish the game. They were asked to touch these tiles as fast as possible. This session lasted for 10 minutes. We asked participants to rate the two techniques on preference, subjective recognition accuracy, and subjective delay.

In session two, the participant touched with the ten popular postures wearing the IMU ring. The motivation was to evaluate the performance of our method over different tapping postures. For each posture, the participant touched 30 black tiles to finish. This session lasted for 20 minutes.

In the last experiment, the head-mounted leap motion had problems of occlusion and spatial accuracy, which led to bad results of the optical method. In this experiment, we placed the leap motion 20 cm right above the interaction area to improve its performance.

Result

Session one

Table 3 shows that the IMU ring improve the contact sensing on both precision and recall rate. The accuracy was measured by the difference between the tested methods and the touch screen (ground truth), so it did not matter if a participant touched a white tile.

	Our method	Optical method
Precision	98.62%(2.50%)	85.42%(10.42%)
Recall Rate	98.61%(1.33%)	84.08%(9.24%)
Completion Time (s)	35.74(13.69)	44.30(19.19)
Delays (ms)	6.61(3.41)	2.98(15.07)

Table 3: The comparison between our method and the baseline. Standard deviations were showed in parenthesis. Notice that the delay here is the gap between the tested methods and the touch screen (ground truth), which has an additional delay of 5 ms.

The task in this experiment required participants to touch quickly. The optical method could not handle with this situation well. For example, the user's finger sometimes did not left the surface more than 15 mm, which affected the recognition of the next touch.

The delay of our method was low and stable. Though we trained the contact classifier using samples of 20 ms delay, most touches could be recognized in less than 20 ms. Considering that the touch screen (ground truth) also had a delay of 5 ms, the average recognition delay of our method was about 10 ms.

The delay of the optical method varied a lot among touches. The optical method sometimes even sensed touch in advanced, which was reported by some participants. This is because the optical method declares a contact if the finger declines below the 10 mm threshold. The participant can feel the early touch when he touches slowly.

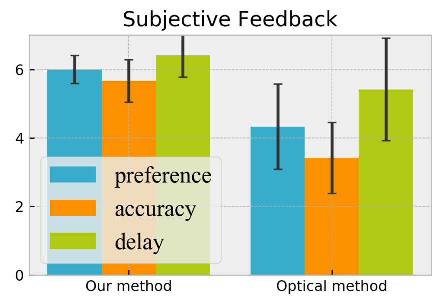

Figure 14: User ratings of the two tested methods (1 - worst; 7 - best).

Figure 14 shows the subjective feedback. Fridman test showed that participants prefer our sensing technique ($\chi^2 = 7.24, p < .01$). They could significantly feel the improvement of accuracy ($\chi^2 = 8.52, p < .01$) in our prototype. Participants felt that the delay of our method was better ($\chi^2 = 5.07, p < .05$), mainly because they found that the optical method sometimes reported touch in advanced.

Session two
Figure 15 shows that our algorithm can sense the contact of various tapping postures accurately. The precision and recall rates exceeded 98% except *IKC*, *MPO* and *2TO*.

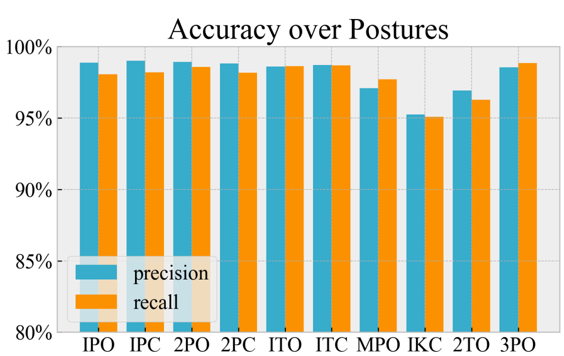

Figure 15: Precision and recall rate of contact detection over tapping postures.

When touching with *IKC* quickly, the participant sometimes made mistakes (e.g, multi-touch), which affected the accuracy (95.1%). The f1 scores of recognizing *MPO* and *2TO* were 97.3% and 96.6%. We acknowledge that a very light touch with these postures may cause recognition error, because the ring can hardly detect such a light vibration.

LIMITATION AND FUTURE WORK
This research has a numbers of limitations, which suggests new directions for future work.

Touch up
The proposed system can only detect touch down and relies on optical methods for touch up detection. It will affect operations based on touch up such as swipe and long press, but will not affect operation based on touch down like single/double tap and typing.

Currently, the touch up event needs to be detected by cameras as prior work did. Our method will not affect the touch up detection based on cameras.

Also, we propose to overcome the limitation in future work. First, a similar machine learning approach is possible to detect touch up. As the lifting direction of touch up is predictable, the acceleration in that direction can be a good feature. Second, the combination of cameras and the finger-worn IMU sensor may improve the detection of touch up.

Implementation
First, we used simple devices to develop the optical part of touch sensing. Better cameras may improve the performance of hand tracking. However, the IMU channel can always be used to improve the optical method.

Second, we used simple machine learning method in this research. We tested SVM and RF (Random Forest) and found that the performance of SVM was slightly better. More sophisticated algorithms such as HMM and LSTM may further improve the performance. We acknowledge that the obtained performance does not reflect the ceiling rate, but it is appropriate to figure out the motion pattern of touch.

Third, the IMU rings in our experiments were wired and not small enough, which may affect the user preference of our proposal. We should make a small and wireless IMU ring in the further to improve the user experience.

CONCLUSION
Touch on any surface is perhaps an input modality in the future. Head-mounted MR systems can affix virtual interface on physical surfaces, which makes it possible to support anywhere touch. Prior work has proposed fingers tracking by the cameras of MR headset, but it has difficulty in sensing contact. To our knowledge, our research is the first to focus on touch contact sensing by IMU ring. Result show that our method can recognize touch contact in 10 ms with the recall rate of 98.61% and spurious extra touch rate of 1.40%. Users prefer to wear an IMU ring for a better touch experience.

In particular, we summarized usable tapping postures with an user preference investigation. We also found that an IMU ring on the proximal phalanx of the index or middle finger can better recognize the contact of various tapping postures.

ACKNOWLEDGMENTS

This work is supported by the National Key Research and Development Plan under Grant No. 2016YFB1001200, the Natural Science Foundation of China under Grant No. 61572276 and No. 61672314, Tsinghua University Research Funding No. 20151080408, and also by Beijing Key Lab of Networked Multimedia.

REFERENCES

[1] Ankur Agarwal, Shahram Izadi, Manmohan Chandraker, and Andrew Blake. 2007. High precision multi-touch sensing on surfaces using overhead cameras. In *Horizontal Interactive Human-Computer Systems, 2007. TABLETOP'07. Second Annual IEEE International Workshop on*. IEEE, 197–200.

[2] Shiri Azenkot and Shumin Zhai. 2012. Touch behavior with different postures on soft smartphone keyboards. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*. ACM, 251–260.

[3] Paul Badger. 2018. Capacitive Sensing Library. (2018). `https://playground.arduino.cc/Main/CapacitiveSensor/`

[4] Hrvoje Benko, Ricardo Jota, and Andrew Wilson. 2012. MirageTable: freehand interaction on a projected augmented reality tabletop. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 199–208.

[5] Stephen J Bisset and Bernard Kasser. 1998. Multiple fingers contact sensing method for emulating mouse buttons and mouse operations on a touch sensor pad. (Oct. 20 1998). US Patent 5,825,352.

[6] Daniel Buschek, Alexander De Luca, and Florian Alt. 2015. Improving accuracy, applicability and usability of keystroke biometrics on mobile touchscreen devices. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 1393–1402.

[7] Xiang Cao, Andrew D Wilson, Ravin Balakrishnan, Ken Hinckley, and Scott E Hudson. 2008. ShapeTouch: Leveraging contact shape on interactive surfaces. In *2008 3rd IEEE International Workshop on Horizontal Interactive Human Computer Systems*. IEEE, 129–136.

[8] Jae Sik Chang, Eun Yi Kim, KeeChul Jung, and Hang Joon Kim. 2005. Real time hand tracking based on active contour model. In *International Conference on Computational Science and Its Applications*. Springer, 999–1006.

[9] Xiang'Anthony' Chen, Tovi Grossman, Daniel J Wigdor, and George Fitzmaurice. 2014. Duet: exploring joint interactions on a smart phone and a smart watch. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 159–168.

[10] Jonathan Deber, Ricardo Jota, Clifton Forlines, and Daniel Wigdor. 2015. How much faster is fast enough?: User perception of latency & latency improvements in direct and indirect touch. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 1827–1836.

[11] John Greer Elias, Wayne Carl Westerman, and Myra Mary Haggerty. 2010. Multi-touch gesture dictionary. (Nov. 23 2010). US Patent 7,840,912.

[12] Liuhao Ge, Yujun Cai, Junwu Weng, and Junsong Yuan. 2018. Hand PointNet: 3D hand pose estimation using point sets. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 8417–8426.

[13] Jefferson Y Han. 2005. Low-cost multi-touch sensing through frustrated total internal reflection. In *Proceedings of the 18th annual ACM symposium on User interface software and technology*. ACM, 115–118.

[14] Chris Harrison, Hrvoje Benko, and Andrew D Wilson. 2011a. OmniTouch: wearable multitouch interaction everywhere. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*. ACM, 441–450.

[15] Chris Harrison, Julia Schwarz, and Scott E Hudson. 2011b. TapSense: enhancing finger interaction on touch surfaces. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*. ACM, 627–636.

[16] Seongkook Heo and Geehyuk Lee. 2011a. Force gestures: augmented touch screen gestures using normal and tangential force. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*. ACM, 1909–1914.

[17] Seongkook Heo and Geehyuk Lee. 2011b. Forcetap: extending the input vocabulary of mobile touch screens by adding tap gestures. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*. ACM, 113–122.

[18] Christian Holz and Patrick Baudisch. 2011. Understanding touch. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2501–2510.

[19] Ken Iwasaki, Takashi Miyaki, and Jun Rekimoto. 2009. Expressive typing: a new way to sense typing pressure and its applications. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems*. ACM, 4369–4374.

[20] Shahram Izadi, David Kim, Otmar Hilliges, David Molyneaux, Richard Newcombe, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Dustin Freeman, Andrew Davison, and others. 2011. KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*. ACM, 559–568.

[21] Ricardo Jota, Albert Ng, Paul Dietz, and Daniel Wigdor. 2013. How fast is fast enough?: a study of the effects of latency in direct-touch pointing tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2291–2300.

[22] Shaun K Kane, Daniel Avrahami, Jacob O Wobbrock, Beverly Harrison, Adam D Rea, Matthai Philipose, and Anthony LaMarca. 2009. Bonfire: a nomadic system for hybrid laptop-tabletop interaction. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology*. ACM, 129–138.

[23] Hideki Koike, Yoichi Sato, and Yoshinori Kobayashi. 2001. Integrating paper and digital information on EnhancedDesk: a method for realtime finger tracking on an augmented desk system. *ACM Transactions on Computer-Human Interaction* 8, 4 (2001), 307–322.

[24] Alan HF Lam, Wen J Li, Yunhui Liu, and Ning Xi. 2002. MIDS: micro input devices system using MEMS sensors. In *Intelligent Robots and Systems, 2002. IEEE/RSJ International Conference on*, Vol. 2. IEEE, 1184–1189.

[25] SK Lee, William Buxton, and KC Smith. 1985. A multi-touch three dimensional touch-sensitive tablet. In *Acm Sigchi Bulletin*, Vol. 16. ACM, 21–25.

[26] G Julian Lepinski, Tovi Grossman, and George Fitzmaurice. 2010. The design and evaluation of multitouch marking menus. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2233–2242.

[27] Julien Letessier and François Bérard. 2004. Visual tracking of bare fingers for interactive surfaces. In *Proceedings of the 17th annual ACM symposium on User interface software and technology*. ACM, 119–122.

[28] Sebastian Madgwick. 2010. An efficient orientation filter for inertial and inertial/magnetic sensor arrays. *Report x-io and University of Bristol (UK)* 25 (2010), 113–118.

[29] Damien Masson, Alix Goguey, Sylvain Malacria, and Géry Casiez. 2017. Whichfingers: identifying fingers on touch surfaces and keyboards using vibration sensors. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*. ACM, 41–48.

[30] Nobuyuki Matsushita and Jun Rekimoto. 1997. HoloWall: designing a finger, hand, body, and object sensitive wall. In *Proceedings of the 10th annual ACM symposium on User interface software and technology*. ACM, 209–210.

[31] Rishi Mohindra. 2015. Identifying hover and/or palm input and rejecting spurious input for a touch panel. (July 14 2015). US Patent 9,081,450.

[32] Albert Ng, Julian Lepinski, Daniel Wigdor, Steven Sanders, and Paul Dietz. 2012. Designing for low-latency direct-touch input. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*. ACM, 453–464.

[33] Takehiro Niikura, Yoshihiro Watanabe, and Masatoshi Ishikawa. 2014. Anywhere surface touch: utilizing any surface as an input area. In *Proceedings of the 5th Augmented Human International Conference*. ACM, 39.

[34] Ju Young Oh, Jun Lee, Joong Ho Lee, and Ji Hyung Park. 2017. Anywheretouch: Finger tracking method on arbitrary surface using nailed-mounted imu for mobile hmd. In *International Conference on Human-Computer Interaction*. Springer, 185–191.

[35] Joseph A Paradiso, Kai-yuh Hsiao, Joshua Strickon, Joshua Lifton, and Ari Adler. 2000. Sensor systems for interactive surfaces. *IBM Systems Journal* 39, 3 (2000), 892–914.

[36] Joseph A Paradiso, Che King Leo, Nisha Checka, and Kaijen Hsiao. 2002. Passive acoustic sensing for tracking knocks atop large interactive displays. In *Sensors, 2002. Proceedings of IEEE*, Vol. 1. IEEE, 521–527.

[37] Gonzalo Ramos, Matthew Boulos, and Ravin Balakrishnan. 2004. Pressure widgets. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 487–494.

[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*. ACM, 927–936.

[39] Elliot N Saba, Eric C Larson, and Shwetak N Patel. 2012. Dante vision: In-air and touch gesture sensing for natural surface interaction with combined depth and thermal cameras. In *2012 IEEE International Conference on Emerging Signal Processing Applications*. IEEE, 167–170.

[40] Adrian Spurr, Jie Song, Seonwook Park, and Otmar Hilliges. 2018. Cross-modal deep variational hand pose estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 89–98.

[41] Lee Stearns, Uran Oh, Leah Findlater, and Jon E Froehlich. 2018. TouchCam: Realtime Recognition of Location-Specific On-Body Gestures to Support Users with Visual Impairments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 164.

[42] Naoki Sugita, Daisuke Iwai, and Kosuke Sato. 2008. Touch sensing by image analysis of fingernail. In *SICE Annual Conference, 2008*. IEEE, 1520–1525.

[43] Feng Wang and Xiangshi Ren. 2009. Empirical evaluation for finger input properties in multi-touch interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1063–1072.

[44] Dong Wei, Steven Zhiying Zhou, and Du Xie. 2010. MTMR: A conceptual interior design framework integrating Mixed Reality with the Multi-Touch tabletop interface. In *Mixed and Augmented Reality (ISMAR), 2010 9th IEEE International Symposium on*. IEEE, 279–280.

[45] Andrew D. Wilson. 2004. TouchLight: An Imaging Touch Screen and Display for Gesture-based Interaction. In *Proceedings of the 6th International Conference on Multimodal Interfaces (ICMI '04)*. ACM, New York, NY, USA, 69–76.

[46] Andrew D Wilson and Hrvoje Benko. 2010. Combining multiple depth cameras and projectors for interactions on, above and between surfaces. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*. ACM, 273–282.

[47] Robert Xiao, Scott Hudson, and Chris Harrison. 2016. DIRECT: Making Touch Tracking on Ordinary Surfaces Practical with Hybrid Depth-Infrared Sensing. In *Proceedings of the 2016 ACM on Interactive Surfaces and Spaces*. ACM, 85–94.

[48] Robert Xiao, Greg Lew, James Marsanico, Divya Hariharan, Scott Hudson, and Chris Harrison. 2014. Toffee: enabling ad hoc, around-device interaction with acoustic time-of-arrival correlation. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*. ACM, 67–76.

[49] 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*. ACM, 47–50.

[50] Robert Xiao, Julia Schwarz, Nick Throm, Andrew D Wilson, and Hrvoje Benko. 2018. MRTouch: Adding Touch Input to Head-Mounted Mixed Reality. *IEEE transactions on visualization and computer graphics* 24, 4 (2018), 1653–1660.

[51] Shanxin Yuan, Guillermo Garcia-Hernando, Björn Stenger, Gyeongsik Moon, Ju Yong Chang, Kyoung Mu Lee, Pavlo Molchanov, Jan Kautz, Sina Honari, Liuhao Ge, and others. 2018. Depth-based 3d hand pose estimation: From current achievements to future goals. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2636–2645.

[52] Dan Zhao, Yue Liu, and Guangchuan Li. 2018. Skeleton-based Dynamic Hand Gesture Recognition using 3D Depth Data. *Electronic Imaging* 2018, 18 (2018), 1–8.

APPENDIX I

	Preference	Comfort	Memory
IPO	6.60	6.67	6.40
IPC	6.60	6.67	6.64
2PO	5.95	6.10	5.81
2PC	5.81	5.88	5.86
IKC	5.69	5.90	6.05
ITC	5.62	6.10	5.76
MPO	5.55	6.17	5.79
ITO	5.50	6.07	5.74
3PO	5.14	5.31	5.38
2TO	5.02	5.50	5.12
2TC	4.95	5.29	5.17
MKC	4.90	5.55	5.60
MTO	4.90	5.83	5.33
ISC	4.57	5.07	4.90
2KC	4.55	4.81	5.14
IKO	4.55	4.93	5.12
TSO	4.52	5.26	4.57
PPO	4.48	5.14	5.12
TSC	4.45	4.88	4.55
MKO	4.40	5.17	5.00
RPO	4.33	5.29	4.64
3TO	4.29	4.57	4.69
INO	4.24	5.00	4.57
PSO	4.21	5.05	4.45
3PC	4.17	4.55	4.90
PSC	4.17	5.05	4.50
INC	4.14	4.76	4.83
2KO	4.07	4.50	4.74
MTC	3.93	5.19	4.71
TPC	3.90	3.93	4.98

Table 4: User preference to the top 25 tapping postures.