

Received March 5, 2020, accepted March 25, 2020, date of publication April 6, 2020, date of current version April 27, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2986037

PhoneCursor: Improving 3D Selection Performance With Mobile Device in AR

MINGHUI SUN^{1,2}, MINGMING CAO³, LIMIN WANG^{1,2}, AND QIAN QIAN⁴

¹College of Computer Science and Technology, Jilin University, Changchun 130022, China

²Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun 130022, China

³College of Software, Jilin University, Changchun 130022, China

⁴Computer Technology Application Key Lab of Yunnan Province, Kunming University of Science and Technology, Shanghai 200000, China

Corresponding author: Limin Wang (42600225@qq.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61872164, and in part by the Program of Science and Technology Development Plan of Jilin Province of China under Grant 20190302032GX.

ABSTRACT This study proposes a novel technology as a creative method to interact with an augmented reality (AR) system called PhoneCursor, which is incorporated onto head-mounted displays. PhoneCursor is designed to offer more intuitive, portable and natural interaction. Its technical realization combines the gyroscope and acceleration sensor based on mobile phone. In view of the target selection ability as a base function of PhoneCursor, it also enhances the performance in solving some difficult selection problems in AR, including far and small object selection, 3D occlusion, dense object selection, and batch selection problem. An experiment based on the ISO 9241-9 standard is conducted to investigate this ability, wherein the PhoneCursor technique is compared with the widely used and well-performing head-movement-based selection technique. Results show that PhoneCursor performs better in terms of movement time and throughput than the conventional technique. Some applications are further proposed to showcase the potential of using the PhoneCursor technique in AR scenarios.

INDEX TERMS Human computer interaction, interactive systems.

I. INTRODUCTION

The use of AR equipment is considered to be a prevailing trend in the future, mobile phones have recently been used by billions of people and have already become a part of people's daily lives. Human's familiarity with mobile phones and various sensors on mobile phones make it a natural interactive device. Meanwhile, the 3D selection task is the basic function of the interaction between augmented reality (AR) and virtual reality (VR) and is widely applied to various scenes. For example, if a typical VR device (e.g., HTC Vive) is used, the rays emitted by the controller [1]–[4], [7], [11] or the controller itself is used to select menus and various objects. By contrast, if a typical AR device (e.g., HoloLens) is used in the AR scene, the menus and objects are selected by the ray based on head movement [5]. With the experience in AR and VR scenarios affected by 3D selection tasks, designing an efficient technology can improve the comfort, acceptability, and performance of such interactions.

The associate editor coordinating the review of this manuscript and approving it for publication was Arianna Dulizia.

The two main categories of 3D selection tasks are divided into virtual hand and ray casting technology [30]. AR and VR still encounter difficulties in far and small object selection, object occlusion, and dense object selection. Many researchers have proposed new methods to address these limitations, and various techniques have improved ray casting technology, including adding controllable cursors to rays [1], adding depth information [4], and combining multiple rays [7]. In addition, 3D selection performance has improved by changing input devices, using different kinds of devices [31], adding tracking sensors [5], [16], and carrying out progressive refinement [9], [10], [13].

Although these methods have greatly improved the limitation of 3D selection, direct selection in mid-air of a 3D space (e.g., virtual hands and ray casting technology) will increase user fatigue [29]. Utilizing a 2D plane selection technology to complete 3D selection tasks can reduce fatigue because putting up hands in mid-air is not necessary. Progressive refinement techniques demonstrate good performance on 3D selection task [13]. In the current study, 2D selection technique and progressive refinement are combined to develop a novel technique called PhoneCursor, which is designed to

offer a possible way for interacting with AR based on mobile phone. PhoneCursor is divided into two steps. When an object is selected in a 3D space, the first step is to control the image plane in the AR to scan the target by shaking the wrist. This step is done to map the angle data collected by the gyro sensor of the mobile phone to the depth information of the object in the space. When the image plane is positioned in the AR to the depth position where the target is located, all objects at the same depth, including the target object to be selected, are uniformly displayed on the touch screen of the mobile phone. The next step is to select the target object through the gesture of the finger sliding on the mobile phone's touch screen. The target is selected when the finger slides toward the target. This technology not only has advantages in selecting far and small objects in AR but also exhibits outstanding performance in solving 3D object occlusion and dense object selection. PhoneCursor can also facilitate batch selection and reduce user fatigue.

The contributions of this work are threefold. 1) A novel technique is presented for interacting with AR based on mobile phone. And PhoneCursor demonstrates outstanding performance in the aforementioned challenges. 2) A comparative analysis is conducted between the traditional AR 3D selection technique based on head movement and the proposed technique. The result of the experiment shows that the basic selection function of PhoneCursor technique perform better than the traditional technique. 3) Several applications are released and proposed with a possible way to interact with AR.

II. RELATED WORKS

Selection is a basic function for a technology interacting with AR system. Related studies on the selection technology in AR and VR are explored and divided into three categories: 2D selection technology, 3D selection technology, and the comparison between the two. In addition, the principle of Fitts' law is briefly discussed.

A. 3D SELECTION TECHNOLOGY IN AR AND VR

Many studies have been conducted to solve 3D pointing and selecting tasks in VR and AR scenarios. Baloup *et al.* [1] presented the RayCursor, which was a novel 3D pointing technology in VR based on ray casting technology. RayCursor technology was an enhancement of the traditional ray casting technology, which allows the controllable movement of the ray with six degrees of freedom by adding a controllable cursor. An investigation of crossing-based selection technology [2] under VR scenario reported that crossing was more effective than pointing in terms of time and accuracy and was highly fitted to Fitts' law. Baloup *et al.* [3] argued that the method for selecting far and small objects in the current VR scene was single and the efficiency and accuracy was low. Therefore, they proposed adding another cursor method based on the ray, called Bubble Cursor. Ro *et al.* [4] applied the depth information of the ray to the AR object to allow the user to register and manipulate the virtual object at any

position through the retained depth information in the real 3D space. Kytö *et al.* [5] proposed a new selection technique that combined accurate but highly fatigued head movement selection techniques with low accuracy and fast eye gaze movement selection techniques in the AR field. Moore *et al.* [6] presented a novel selection technology that voted for the indicated object to improve the 3D selection performance. Xu *et al.* [7] presented a novel 3D selection technology called guidance ray technology, which utilized a combination of three rays (one straight ray and two bendable rays) to select the object in a VR scenario. Park *et al.* [8] proposed SelectAhead to improve the efficiency and performance of 3D selection in an object density environment under a VR scenario. Progressive refinement can effectively solve the issues of small target selection, jitter, and density object selection in VR [9], [10]. Yu *et al.* [11] explored the performances of widely used selection techniques (ray casting, virtual hand, and hand extension) in solving far and small object selection, dense object selection, and occlusion problems in VR scenario through a detailed experimental design. Mendes *et al.* [12] proposed a novel mid-air method called PRECIOUS to solve the problem regarding out-of-reach object selection. An evaluation and tradeoff on accuracy and speed of 3D selection were investigated between progressive refinement technology and immediate technique [13]. Bhowmick *et al.* [14] explored an object selection method to solve a dense and occluded dense problem in VR based on body gesture. Wolfgang *et al.* [32] investigated the input method based on mobile phone to manipulate the AR object, however, it not focused on selection task in AR.

As previously mentioned, 3D selection tasks have inherent problems. Although many researchers have proposed various solutions, the balance between performance and comfort in 3D space selection were still a problem. We propose the PhoneCursor to improve the 3D selection performance based on mobile phone, in view of people's familiar with phone offers acceptability and comfort.

B. 2D SELECTION TECHNOLOGY IN AR AND VR

In recent years, numerous researchers have studied the selection task in a 2D interaction space. Wu *et al.* [15] proposed the HorizontalDragger technology to map the 2D selection to the 1D selection, selected the target from dense objects in a 2D selection, and improved the selection accuracy. Delamare *et al.* [16] used handheld devices to select 3D objects in AR and proposed the P2Roll and P2Slide selection methods to explore the balance between focus and performance among multiple AR objects. Unlike the proposed PhoneCursor, which used the wrist to shake (gyroscope) and select the plane in the located object in AR according to the depth information of the object, P2Roll directly selects the object through the roll of the wrist. Meanwhile, PhoneCursor utilizes the finger slide gesture to select the target in a specific depth plane, while P2Slide uses the finger slide gesture to select the object in 3D scenario. Debarba *et al.* [17] proposed a novel selection technology of utilizing two steps to select

objects in a VR scenario. The first step required the user to point to the region where the desired target was located using the handheld device. Then, the objects in the pointed region were arranged and mapped to the handheld device. Therefore, the user can select the target directly from the touch screen. However, this mechanism is different from the proposed selection technique. Vemavarapu and Borst [18] reported that using the handheld’s touch interface demonstrated a better target selection performance than the standard ray pointing in a 3D visualization environment. Prachyabrued *et al.* [19] proposed Handymap, which was a novel technology for the 3D selection of dense objects in VR. Kim and Bang [20] proposed a new technology called VRMouse, which used the VR controller to simulate the 2D selection of desktop mice. However, compared to traditional ray-based selection methods, VRMouse’s performance was less satisfactory. Teather and Stuerzlinger [21] investigated the method of using a 2D-projected 3D object to finish the selection task in 3D. Lubos *et al.* [22] analyzed the performance of direct selection in the 3D interface of a VR environment. Qian and Teather [23] explored the performance of three selection techniques, namely, eye-based, head-based, and the combination of the two. The results showed that head-based selection was the most suitable method among the three, whereas the eye-based selection was the least satisfactory. Therefore, the head-based selection was adopted for the comparative analysis with PhoneCursor. Ramcharitar and Teather [24] proposed and evaluated a head-coupled cursor to assist 2D selection in head-mounted displays (HMDs). EZCursorVR [31] explored the impact of different input devices on the performance of the selection tasks by mapping the 3D objects in VR to the virtual planes, thereby mimicking the working process of computer desktops. EZCursor compared the devices like Mouse, Joystick and controller in imitating the mouse to select objects on two-dimensional desktop scenario. However, people may interact with the world in UBICOMP. In our paper, we try to explore how people use mobile device to interact with HMD. This study will give the early exploration and fundamental result in this situation. Moreover, we provided table 1 to compare with these related works, which focus on using the cross-device approaches to make a selection task.

As previously mentioned above, 2D selection technology facilitated 3D selection task based on indirect selection. Moreover, through the method of progressive refinement, 2D selection technology has a good balance between performance and comfort for user. Therefore, we proposed a novel method called PhoneCursor to improve the performance on 3D selection.

C. FITTS’ LAW

The 3D selection task in this study is based on Fitts’ law. MacKenzie [26], [27] extended the principle of this law to 2D selection tasks and utilized the concept as an efficient tool for performance evaluation in the human–computer interaction (HCI) field. The predicted model of Fitts’ law was typically studied through a standard 3D selection task of ISO 9241-

TABLE 1. Utilizing Cross-device approaches to make a selecting task which included in related work.

| Article Method | 2D/3D VR/AR | Cross Device | Solving problems | Benefits |
|--|-------------|----------------------------|---|---|
| RayCursor[1], Goal-Crossing[2], Bubble-Cursor[3] | 3D/VR | Controller | Far/small-Object Batch - Selection | Low error Efficient High-acceptance Outperform med-Pointing |
| Guidance Ray[7] | 3D/VR | Wii mote | Far/small-Object Occluded-Object Dense-Object | Efficient Accurate-Selection |
| Depth variable Raycasting[4], Smart Phone-Based[32], Mobile Phone Pointing[16] | 3D/AR | Phone | Far object-Manipulate, Dense-Object | Promote-accuracy, Efficient High-acceptance |
| Disambiguation-Canvas[17], Handymap[19] | 2D/VR | Phone | Dense/Small-Object, Occluded-Object | Easy Efficient Accurate, Promote-Accuracy |
| VRMouse[20], EZCursor[31] | 2D/VR | Controller, Mouse Joystick | Far/small-Object | Less-satisfactory, Mouse-Outperform med |
| PhoneCursor | 2D/AR | Phone | Occluded-Object Dense-Object Far/small-Object Batch-Selection | Easy Efficient Anywhere Anytime Low-fatigue High-acceptance |

Utilizing Cross-device approaches to make a selecting task which included in related work.

9 [28]. In the previous study, Fitts’ law was treated as a predicted model that showed that the complexity of the task was linearly related to the task’s completion time. The model was defined by (1), where *ID* stands for index of difficulty

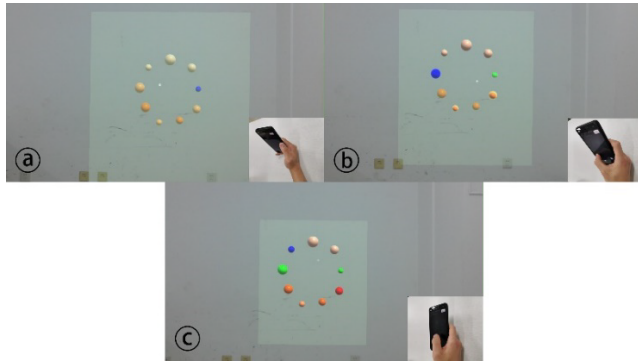


FIGURE 1. (a) When the user shakes his/her wrist to maintain the gesture of the phone to 90°, the image plane in space moves to the front of all objects. (b) When the user shakes his/her wrist to maintain the gesture of the phone at approximately 45°, the image plane in space just moves to the plane where the blue target is at the same depth. (c) When the user shakes his/her wrist to make a horizontal gesture of the phone, the image plane in the space moves behind all the objects.

and MT represented the movement time of the selection task.

$$MT = a + b \times ID \quad (1)$$

$$ID = \log\left(\frac{D}{W} + 1\right) \quad (2)$$

In addition, parameters a and b were the coefficients of the linear regression of Equation (1), D represented the distance between the starting position of the selection task and the target position, and W was the width of the target. Meanwhile, throughput (TP) was also used as a model for evaluating the performance of selection tasks and is calculated as

$$TP = \frac{ID}{MT} \quad (3)$$

III. PHONECURSOR

PhoneCursor is similar to many image plane selection methods and is also used to select objects in a 3D space by plane mapping. The cursor does not appear in the HMD field of view because the final selection is executed on the phone screen and not on the virtual plane of the AR. From the perspective of user experience, the selection process is consistent with the normal one-hand use of the phone, and the user only needs to be familiar with how to select the plane and determine the depth through the dynamic changes of the mobile gyroscope. The dynamic change in the mobile phone gyroscope is determined by the up and down shaking of the wrist holding the mobile phone, in which the dynamic change in the angle and speed between the mobile phone and the horizontal plane is used to move an image plane back and forth in the AR scenario. Unlike traditional image plane mapping, PhoneCursor does not map all objects in the field of view to the selection plane. The selection plane in the PhoneCursor technology will determine the depth at which the object will appear on the phone as well as the relative position of the objects on the phone screen. Figure 1 shows the specific workflow of the PhoneCursor technology.

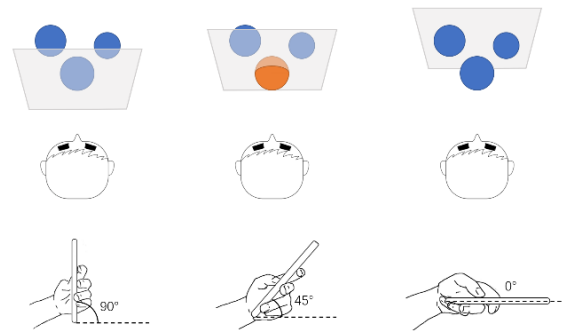


FIGURE 2. Depth mapping method based on the pitch angle of phone held by hand.

A. RENDERING OF THE SELECTED IMAGE PLANE

When completing the selection task, a translucent image plane appears in the user's field of view in the HMD, which is perpendicular to the user's line of sight. The size of this image plane can be changed according to the requirement of different scenarios and different users. At the same time, the constant change in the angle because of head movement will cause jitter and affect accuracy. Therefore, after the user starts the selection, the selection plane is set in the world's coordinate system and the angle is not changed until the selection is completed. In addition, although the translucent plane will overlap with the object (e.g., the plane will block the object when the plane is in front of the object, and the object will cross the plane when the plane intersects with the object), the depth of the plane will be clearly visible such that the user can clearly determine if the target object collides with the plane. To allow the user to clearly perceive the movement of the plane, the selection plane is also designed to be smaller than the actual field of view. The system will display and scale the objects touched by the selection plane on the screen of the mobile phone. The scale method is set according to the size of the selection plane and phone. The objects that are touched will be highlighted in light colors (e.g., red, blue, and green).

B. DEPTH MAPPING METHOD IN AR

One of the problems that must be solved when using a 2D plane to select objects in a 3D space is how to deal with the depth of the object and represent the depth information. When the user holds the phone in one hand, the front and rear swings of the wrist cause the posture of the mobile phone to change. At the same time, the mobile phone gyroscope and accelerometer can record the change in the posture of the mobile phone so the spatial depth can be mapped by the phone's attitude angle. Figure 2 illustrates the depth mapping method in 3D space through a mobile phone. Among the Euler angles of the three directions recorded by the gyroscope (i.e., pitch, yaw, and roll), the changing of the pitch angle is the easiest to operate and the most suitable for intuitive zooming in and out. Hence, this angle is mapped to the depth. In the proposed design, the horizontal state (0°) of the

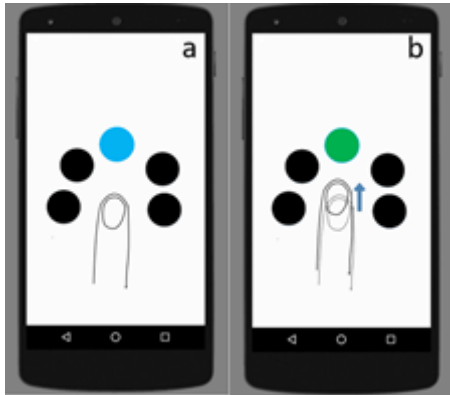


FIGURE 3. (a) Mapping of the image plane where the target is located to the screen of the phone. when the finger touches the screen. (b) When the finger slides toward the blue target, the target is selected and becomes green.

mobile phone corresponds to the farthest distance that can be selected, while the vertical state (90°) corresponds to the closest distance that can be selected. Matching the angular range (0° – 90°) where the mobile phone can shake to the depth range of the 3D space is equivalent to vertically dividing the 3D space into a number of “slices,” where the display content of the mobile phone is the content of the “slice” corresponding to a certain pitch angle. Therefore, combining the phone’s pitch angle (z-axis of the depth in 3D space) with the sliding operation on the phone screen (x- and y-axes of the spatial coordinates) will determine the position in the space.

C. SELECTION METHOD ON THE PHONE SCREEN

After the target object in the 3D space is mapped to the screen of the mobile phone, two selection methods will be applied to complete the selection task. Both methods start when the finger touches the screen and end when the finger leaves the screen. The differences between these methods are as follows. In the first option, the finger should be moved to the selected object after pressing and then lifted to complete the selection. According to the results of the pilot study, this selection method has high precision and success rate. However, the disadvantage of this technique is that the user needs certain visual feedback. Therefore, when the user frequently watches the mobile phone screen to complete the selection task, user fatigue is increased and selection speed is decreased. In the second option, when the finger touches the screen, the finger should slightly slide toward the direction of the target to complete the selection task. According to the results of the pilot study, the object on the screen of the mobile phone is selected by the sliding direction, which has the advantages of short selection time, low fatigue, and unnecessary visual feedback. Therefore, the latter selection method is adopted. Figure 3 displays the mechanism of the selection method on the phone screen.

D. COMPARISON WITH TRADITIONAL IMAGE PLANE SELECTION

Our selection method is different from the traditional image plane selection in two aspects. First, the latter directly projects the 3D space into a 2D plane from the user’s perspective. Despite the simplicity and intuitiveness, the depth information of the space is ignored during the mapping process, which makes the selection difficult when the distant object is blocked by the near object. Therefore, occlusion is an urgent problem that should be resolved. PhoneCursor preserves depth information by dividing the depth information space into a myriad of 2D planes, making the selection of occluded objects possible. The user only needs to select among the objects at a certain depth rather than among all objects, which reduces the difficulty of the user’s operation. Second, the PhoneCursor’s finger touch selection on the 2D screen of the mobile phone reduces the range of limb movement and does not require visual feedback, thereby reducing fatigue. After traditional image plane selection, the distant objects will be relatively smaller than the nearby objects. Thus, according to Fitts’ law, these objects will be more difficult to select. By contrast, PhoneCursor will maintain the original size information of the object in space and will not increase the difficulty of selection owing to the change in size of the projection itself.

IV. EXPERIMENT

A. PARTICIPANT

Twelve subjects (8 males, 4 females, and aged 20–30 years) from a local university were recruited. Nine of the subjects have experience in using AR/VR and the rest are not familiar with AR or VR. All subjects have used a smart phone no less than two years and are right-handed.

B. APPARATUS

The experiment was conducted on HoloLens, which is a product produced by Microsoft. The system development experiment included a laptop with an Intel Core i5 7200-U quad core processor, 8 GB of RAM, and Microsoft Windows 10 operating system. An ordinary smart phone with an Android operating system no older than the 3.4.7 version was used. The experiment was developed by the Unity and Microsoft Visual Studio 2017 community, which generated the code and designed an AR scenario before releasing the phone and HoloLens versions. The library of Mixed Reality Toolkit was used as development kits in Unity.

C. TASKS

The experiment task was based on an ISO-9241-9 selection task [28] and in accordance with the principle of Fitts’ law. Figure 5 shows that nine spherical targets are rounded into a circle and presented in different sizes, colors, depths, and diameters. The size of the spherical targets were set at three levels (0.06, 0.08, and 0.1 m). Moreover, the spherical targets in each circle have four colors: 1) white represents the initial



FIGURE 4. Experiment scenario where the subject wears the HoloLens while using a mobile phone to select the target.

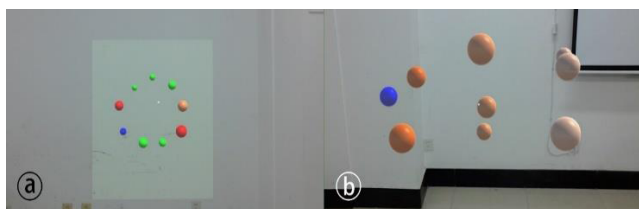


FIGURE 5. (a) Nine spherical targets with different colors rounded in a circle. (b) Nine spherical targets distributed in three depths.

state before the trial, 2) green signals the selection process, 3) blue is the target to be selected after starting the trial, and 4) red signifies that the target selection failed. The depths of the spherical targets were set at 3.3, 3.6, and 3.9 m, and the circle diameters were set at 0.3, 0.6, and 0.9 m, respectively. The input method consisted of two parts. The first part allowed the subject to move his/her head to focus on the target and use the pinch gesture as the command trigger. The second part allowed subjects to slide on the screen of the phone using one hand to select the target.

D. PROCEDURE

Before starting the first trial, the subjects were instructed to familiarize themselves with the selection task. The experiment was divided into two parts because of the different input methods; the first part used the head movement and pinch gestures, and the second part used the phone to indirectly select the target. The researchers showed the subjects how to use the HMD to finish the selection task and how to utilize the phone as a cursor to select the target. The subjects were required to do what the researchers did for about 15 minutes. The experiment data included username, success or failure, movement time, throughput, distance, width, and depth. The data generated during the training stage were used to instruct the subject to achieve better training than the previous one. However, these data were excluded from the normal analysis. After ensuring that the subjects were familiar with each trial, the experiment officially started. According to the balance of a Latin square, each subject decided the execution order of

the two parts of the experiment. In each trial, the subjects were required to press the start timing button in the first part of the experiment. This button is a white spherical target at the center of the other targets. Next, this gesture was replaced by the action of pressing on the phone screen in the second part of the experiment. As soon as the start timing button was triggered, the target would turn blue. The subjects were then required to select the blue spherical target. If the blue spherical target was correctly selected, the color would turn green. Otherwise, the target would turn red. To provide feedback on the state of the subject's selection, the currently selected spherical target was highlighted. The selection time of each trial was recorded only when the color of the target turned red or green.

The subjects were allowed to rest between the two parts of the experiment for 5 minutes. After the experiment was completed, the generated data and subjects' feedback (e.g., fatigue, satisfaction, advantages, and disadvantages) were recorded. The subjects were further asked to rank and evaluate their preferred input method. Finally, each subject was given 10 Chinese yuan as compensation for their participation. It took them 45 minutes to finish the experiment.

E. DESIGN

The experiment used a within-subject design to evaluate the effect of independent variables, including device, distance, width, and depth.

Device: Head, Smartphone

Distance: 0.3, 0.6, 0.9 (m)

Width: 0.06, 0.08, 0.1 (m)

Depth: 3.3, 3.6, 3.9 (m)

Three dependent variables were then selected, namely, movement time, throughput, and error rate. Movement time was calculated as the duration of the time from the pressing of the start button to the end of the selection, throughput was determined according to (3), and error rate is the percentage of missed targets. Each subject selected nine spherical targets per round, which was repeated three times. The targets were randomly generated according to the combination of width \times depth. In total, 17496 trials were generated (9 spherical targets \times 2 devices \times 3 distances \times 3 widths \times 3 depths \times 12 subjects).

V. RESULT AND DISCUSSION

A. THROUGHPUT

Repeated measures ANOVA showed that the independent variable for device has a significant effect ($F_{1,11} = 4.755$, $p = 0.05$) on throughput. Similarly, depth ($F_{2,22} = 9.128$, $p = 0.001$) and width ($F_{2,22} = 16.143$, $p = 0.000$) exert a significant effect on throughput. The respective interaction effects of device \times distance, device \times width, device \times distance \times width, device \times depth, distance \times depth, device \times distance \times depth, distance \times width \times depth, and device \times distance \times width \times depth are as follows: ($F_{2,22} = 17.5$, $p = 0.000$), ($F_{2,22} = 3.734$, $p = 0.04$), ($F_{4,44} = 2.893$,

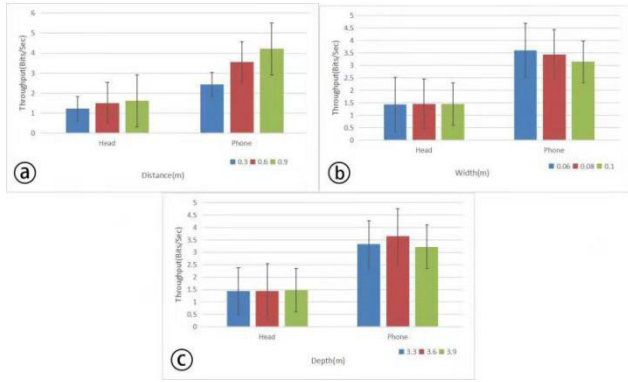


FIGURE 6. Result of the pairwise difference of throughput for distance, width, and depth.

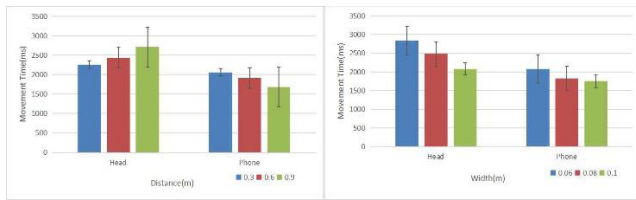


FIGURE 7. Result of movement time for distance and width.

$p = 0.033$), ($F_{4,44} = 3.532, p = 0.014$), ($F_{2,22} = 11.009, p = 0.000$), ($F_{4,44} = 2.915, p = 0.032$), ($F_{4,44} = 2.623, p = 0.047$), ($F_{8,88} = 3.067, p = 0.004$), and ($F_{8,88} = 2.861, p = 0.007$). The average throughput values on the head-based selection and PhoneCursor were 1.452 and 3.397 bps, respectively. The pairwise difference of throughput among the four independent variables are shown in Figure 6.

B. MOVEMENT TIME

Repeated measures ANOVA showed that the independent variable for device has a significant effect on movement time ($F_{1,11} = 4.755, p = 0.05$). Width has a significant effect on movement time ($F_{2,22} = 16.143, p = 0.000$) as well. Distance and depth, however, exert no significant effect. However, device \times distance and device \times width demonstrated significant interaction effects ($F_{2,22} = 16.134, p = 0.000$ and $F_{2,22} = 3.65, p = 0.043$, respectively). The average movement times of the head-based selection and PhoneCursor were 2469.19 and 1887.57ms, respectively. The pairwise difference of movement time among the four independent variables are displayed in Figure 7.

C. ERROR RATE

The head-based selection exhibited few failures and obtained an accuracy rate of approximately 100%, which can be attributed to the work process and principle of the method itself. This selection technique requires the subjects to move their heads to focus on the target and then make the pinch gesture as the command trigger [4]. The error rate result for the PhoneCursor is presented in Figure 8. The error rate decreased when the distance increased. The maximum error

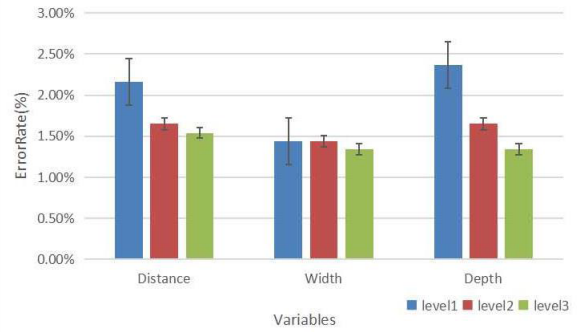


FIGURE 8. Error rates for distance, width, and depth.

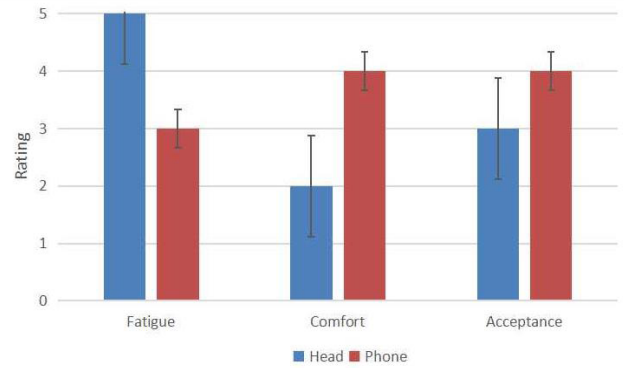


FIGURE 9. Fatigue, comfort, and acceptance ratings for head-based selection and PhoneCursor.

rate was 2.16% at level 1 (distance = 0.3 m), and the minimum is 1.54% at level 3 (distance = 0.9 m). The error rate was less fluctuation on width. Moreover, the error rate decreased when depth increased. The maximum error rate was 2.37% at level 1 (depth = 3.3 m), and the minimum was 1.34% at level 3 (depth = 3.9 m). The error rate at level 2 (depth = 3.6 m) was 1.65%.

D. SUBJECTIVE EVALUATION

All evaluations from the subjects during and after the experiment were collected and ranked on a five-point Likert scale. The results are shown in Figure 9. The average scores for fatigue level in head-based selection and PhoneCursor were 5 and 3, respectively. Moving the head and stretching the arm in mid-air for a long time increased the fatigue level of the subjects [29]. Moreover, the respective average scores for comfort were 2 and 4, which might be due to the fact that smart phones can be used to make selection in any way without the need to make large physical movements. The average scores for acceptance were 3 and 4, respectively. High fatigue and low comfort from the head-based selection is the reason the scores for PhoneCursor is higher than that for the former. Most of the highest scores were only 4 owing to the weaknesses and disadvantages of both methods.

E. DISCUSSION

Compared to the existing selection technology in head-mounted AR devices which utilizes head movement and gestures to make a selection, the proposed selection technology exhibited a more satisfactory performance in terms of movement time and throughput. The average movement time of head control was 2469.19ms, which was higher than that of phone control at 1887.57ms. The subjects also rapidly finished the selection task during the initial stage of the first part of the experiment, but the time increased eventually. The subject also specified that the long-time arm stretch and the frequent head movement caused distraction and high fatigue. In the process of using phone control selection technology, in which the objects are selected by using the gyroscope and by sliding the finger on the screen, the distraction and fatigue levels decreased. In addition, the greater familiarity of the users with mobile phones compared with AR devices is a crucial factor for the analysis. The average throughput of the head control experiment was 1.452 bps, while that of phone control was 3.397 bps. According to (3), the lower the movement time, the higher the throughput.

The result of the above experiment indicates the overall effect of independent variables on the two parts of the experiment. Therefore, the specific impact of the independent variables (distance, width, and depth) on the proposed technique were further analyzed. Repeated measures ANOVA showed that distance has a significant effect on movement time ($F_{2,22} = 4.421, p = 0.024$) and throughput ($F_{2,22} = 24.094, p = 0.000$). The longer the distance, the larger the radius of the ring formed by the nine spherical targets and, consequently, the more discrete the targets are. This outcome is the reason the sliding direction of the finger on phone screen is utilized to make selections faster and easier. Moreover, the result shows that width has a significant effect on movement time ($F_{2,22} = 9.26, p = 0.001$), and depth has a significant effect on throughput ($F_{2,22} = 3.905, p = 0.035$). However, width has no significant effect on throughput, and depth has no significant effect on movement time. According to the principle of the proposed technology, object selection in 3D space is simulated using the 2D direction selection on the screen of the mobile phone. Therefore, depth does not necessarily affect movement time.

The proposed selection technology has many advantages compared to other methods (e.g., effective selection of small and remote targets in AR scenario). The experiment results suggest that width and distance influence the selection task. The smaller/farther the target, the more difficult the selection process will be. However, the PhoneCursor technology uses 2D selection to simulate the 3D selection in AR, which overcomes this disadvantage to some extent. Overcoming 3D object occlusion is another advantage of the proposed method. PhoneCursor technology divides the objects occluded in 3D into different 2D planes according to the vertical distance from the subject (i.e., depth of the target) and then selects the plane where the target is located, thereby solving the occlusion problem. PhoneCursor also facilitates batch

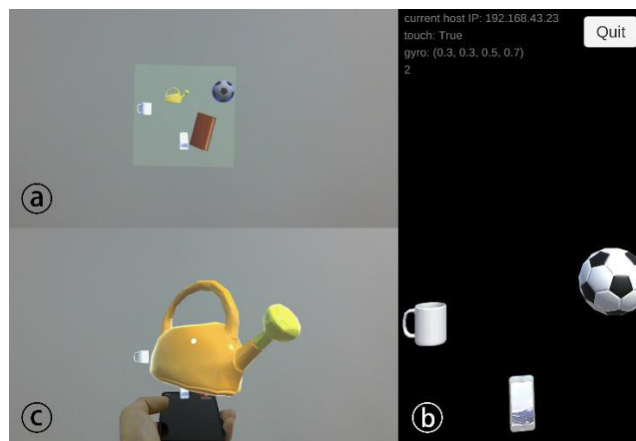


FIGURE 10. (a) Some items are disorderly distributed in the AR scene (football, cup, phone, book, etc.). (b) Using the PhoneCursor technique to scan the item and map them in the touch screen. (c) When the finger slides to select the desired item, the item appears at the top of phone and can be easily manipulate using the device.

selection. The traditional way to select objects in AR is from one object after the other; no method can be used to directly select multiple objects at once. PhoneCursor uses the finger to slide on the screen and cross multiple targets to implement batch selection (e.g., online shopping). The last advantage of PhoneCursor is the fatigue reduction effect. PhoneCursor technology does not require frequent head movements or arm stretching to select the targets. Slight wrist and finger (sliding) movements can accomplish the selection task.

VI. FUTURE APPLICATION

The advantages mentioned in the previous section have broad prospects in real life. In view of these prospects, future applications are designed and described by utilizing PhoneCursor technology.

A. ONLINE SHOPPING

Online shopping is popular around the world. When shopping online, only the preferred products are selected from the pictures, along with some text descriptions (e.g., size, style, and color). However, selecting a suitable product from the picture and text descriptions always yields mistakes. Therefore, online shopping in AR or VR is a good solution to satisfy the requirements of users. A typical online shopping scenario is designed to evaluate the selection performance of PhoneCursor (Figure 10). The product categories include toy, vase, kettle, and computer. These products are randomly distributed in front of the users at different heights and depths. They are created according to the actual scale in real life. In the first application, if the user wants to buy all the products in front of him/her, he/she does not need to select one by one and then pay. Instead, the user can use a smart phone to scan all objects. The specific scanning method involves the user controlling the virtual plane in the AR to move back and forth through the dynamic changes of the mobile gyroscope. If the virtual plane collides with the object, then

the object is scanned and will appear on the user's mobile phone screen. The function of a virtual plane is like a cross section consisting of a myriad of rays, which are then used to scan an object. After the scanned products appear on the phone screen, the user can slide a finger to select and pay for the items all at once. If numerous objects are selected, the advantage of batch selection will be highlighted.

B. MANIPULATE OBJECTS

We can also use PhoneCursor technology to promote interaction performance for some operations including zoom, scale and pinch. The second application is used to describe the detail of the products. Although the items in the AR can be manipulated through gestures, such as zoom, scale, rotate, and move, these operations are undoubtedly inaccurate and will increase the user's fatigue. In PhoneCursor, the selected object will move directly above the phone screen (Figure 10c). The mobile phone gyroscope can then be manipulated to control the precise rotation of the product on the top of the mobile phone as well as the enlargement and reduction of the product size by sliding the finger on the screen of the mobile phone. Moreover, the position of the mobile phone can be changed to achieve the effect of moving the product. According to user feedback, as expected, using mobile phones to control AR objects can reduce fatigue and increase the interest for using phone to interact with AR.

C. WORD-GESTURE TYPING IN AR

Gupta *et al.* [33] proposed a new method called RotoSwype that utilized the orientation of a ring device on a finger to input text. Considering the gyroscope and acceleration sensor, we can also use PhoneCursor for word-gesture typing and make it a new input method. We arrange the letters according to 3×3 grid, and place three letters in each grid, at different depths or at the same depth. Then, we control the orientation and angle of mobile phone based on wrist shaking. Corresponding to AR, a virtual plane move forth and back to find which grid is our target. If we find the location of the target, when the finger touches the touch screen, the target will appear on the mobile phone according to the original layout, and then select the letter input by sliding the finger. The advantage of PhoneCursor for word-gesture typing is that it does not need visual feedback, single hand input, and can input at any time, such as walking, standing, etc.

D. OTHER APPLICATIONS

PhoneCursor technology has many other applications in AR that have not been designed yet. For instance, it can be employed as an assistant tool for sketching in AR, just like the combination of 2D and 3D sketching proposed by Arora [34]. When a mobile phone is used to make a call, the user information is displayed and an AR postcard is created. During videotaping, the video information can be displayed in the AR through the proposed technology. Moreover, when walking in the street or in the crowd, the accuracy of voice recognition in a noisy environment is low and the gestures will attract

people's attention. Therefore, using PhoneCursor technology to interact with AR devices can protect privacy. PhoneCursor technology can use the mobile phone as a controller of an AR device, which can map the menu of the AR device to the screen of the mobile phone. The icon on the screen can simply be clicked to trigger the corresponding AR application. In conclusion, using PhoneCursor to interact with AR is a novel and complementary method during several occasions.

VII. CONCLUSION

This study presents a novel method for a possible way to interact with AR with HMD called PhoneCursor. Microsoft HoloLens and a smart phone are used to evaluate the selection performance of the proposed technology. Object selection is the basic function of PhoneCursor. The core idea of this selection technique is to perform selection in a 2D plane to replace the selection in a 3D space. A fixed plane was controlled by the orientation and angle of mobile phone. When the object for selection touches the plane, the selected target is mapped in the screen of the mobile phone. The target can be selected by finger gesture.

A control experiment is designed to explore the performance of PhoneCursor technology and collect feedback from the subjects regarding their acceptance of the technology. PhoneCursor is compared with the head based selection technology used in the HMD, which uses head movements to focus on the target object and the pinch gesture to confirm a selection. Results show that PhoneCursor performs better than the selection method using head movement and gesture in terms of movement time and throughput. Although the error rate of the PhoneCursor technology is similar to that of the traditional one, the average error rate is acceptable to the users. The effect of three independent variables, namely, distance, width, and depth, on the performance of the proposed technology is analyzed and discussed as well.

PhoneCursor has several advantages, including batch selection capability, highly efficient selection of small or remote targets, low fatigue, high user acceptance, and the novel interaction method of combining HMD and smart phone through the gyro sensors in the phone. The main contributions are threefold. 1) A new selection method that utilizes the gyro sensors of smart phones is presented to control the selection in HMD. 2) A comparative experiment is performed by comparing PhoneCursor with a widely used selection technology in AR (i.e., method involving head and pinch gestures). 3) Several real-life application scenarios that demonstrate the usability of PhoneCursor technology are designed.

VIII. FUTURE WORK

In future work, the impact of visual and haptic feedback on the PhoneCursor selection technology should be explored. On the basis of user feedback, certain visual feedback will reduce the error rate. The shape and color of the target will likewise have a certain impact on visual feedback, and haptic feedback will provide valuable insights regarding the

selection performance using the finger. The performance of PhoneCursor in solving 3D occlusion problems is also an interesting research area. The difference between the application scenarios and selection tasks of the PhoneCursor in VR and AR will be further investigated. Other selection tasks in VR and AR will also be evaluated in future work.

REFERENCES

- [1] M. Baloup, T. Pietrzak, and G. Casiez, "RayCursor: A 3D pointing facilitation technique based on raycasting," in *Proc. CHI Conf. Hum. Factors Comput. Syst. (CHI)*, New York, NY, USA, vol. 1, 2019, pp. 1–12, doi: [10.1145/3290605.3300331](https://doi.org/10.1145/3290605.3300331).
- [2] H. Tu, S. Huang, J. Yuan, X. Ren, and F. Tian, "Crossing-based selection with virtual reality head-mounted displays," in *Proc. CHI Conf. Hum. Factors Comput. Syst. (CHI)*, vol. 618, 2019, pp. 1–14, doi: [10.1145/3290605.3300848](https://doi.org/10.1145/3290605.3300848).
- [3] M. Baloup, V. Oudjail, T. Pietrzak, and G. Casiez, "Pointing techniques for distant targets in virtual reality," in *Proc. 30th Conf. Interact. Homme-Mach. (IHM)*, 2018, pp. 100–107.
- [4] H. Ro, S. Chae, I. Kim, J. Byun, Y. Yang, Y. Park, and T. Han, "A dynamic depth-variable ray-casting interface for object manipulation in ar environments," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2017, pp. 2873–2878.
- [5] M. Kytö, B. Ens, T. Piumsomboon, G. A. Lee, and M. Billinghurst, "Pinpointing: Precise head- and eye-based target selection for augmented reality," in *Proc. CHI Conf. Hum. Factors Comput. Syst. (CHI)*, 2018, p. 81.
- [6] A. G. Moore, J. G. Hatch, S. Kuehl, and R. P. McMahan, "VOTE: A ray-casting study of vote-oriented technique enhancements," *Int. J. Hum.-Comput. Stud.*, vol. 120, pp. 36–48, Dec. 2018.
- [7] C. Xu, M. Zhou, D. Zhang, W. Shui, and Z. Wu, "Guidance rays: 3D object selection based on multi-ray in dense scenario," in *Proc. 12th ACM SIGGRAPH Int. Conf. Virtual-Reality Continuum Appl. Ind. (VRCAI)*, 2013, pp. 91–100.
- [8] S. Park, S. Kim, and J. Park, "Select ahead: Efficient object selection technique using the tendency of recent cursor movements," in *Proc. 10th Asia Pacific Conf. Comput. Hum. Interact. (APCHI)*, 2012, pp. 51–58.
- [9] F. Bacim, R. Kopper, and D. A. Bowman, "Design and evaluation of 3D selection techniques based on progressive refinement," *Int. J. Hum.-Comput. Stud.*, vol. 71, nos. 7–8, pp. 785–802, Jul. 2013.
- [10] R. Kopper, F. Bacim, and D. A. Bowman, "Rapid and accurate 3D selection by progressive refinement," in *Proc. IEEE Symp. 3D User Interfaces (3DUI)*, Mar. 2011, pp. 67–74.
- [11] D. Yu, H. N. Liang, F. Lu, V. Nanjappan, K. Papangelis, and W. Wang, "Target selection in head-mounted display virtual reality environments," *J. Universal Comput. Sci.*, vol. 24, no. 9, pp. 1217–1243, 2018.
- [12] D. Mendes, D. Medeiros, M. Sousa, E. Cordeiro, A. Ferreira, and J. A. Jorge, "Design and evaluation of a novel out-of-reach selection technique for VR using iterative refinement," *Comput. Graph.*, vol. 67, pp. 95–102, Oct. 2017.
- [13] F. B. de Araujo e Silva, "Increasing selection accuracy and speed through progressive refinement," Ph.D. dissertation, Virginia Tech, Blacksburg, VA, USA, 2015.
- [14] D. A. W. Rosa and H. H. Nagel, "Selection techniques for dense and occluded virtual 3D environments, supported by depth feedback: Double, bound and depth bubble cursors," in *Proc. 29th Int. Conf. Chilean Comput. Sci. Soc.*, 2010.
- [15] S. Wu, A. Chellali, S. Otmane, and G. Moreau, "HorizontalDragger: A freehand remote selector for object acquisition," in *Proc. IEEE Virtual Reality (VR)*, Mar. 2015, pp. 307–308.
- [16] W. Delamare, C. Coutrix, and L. Nigay, "Mobile pointing task in the physical world: Balancing focus and performance while disambiguating," in *Proc. 15th Int. Conf. Hum.-Comput. Interact. Mobile Devices Services (MobileHCI)*, 2013, pp. 89–98.
- [17] H. G. Debarba, J. G. Grandi, A. Maciel, L. Nedel, and R. Boulic, "Disambiguation canvas: A precise selection technique for virtual environments," in *Proc. IFIP Conf. Hum.-Comput. Interact.* Berlin, Germany: Springer, Sep. 2013, pp. 388–405.
- [18] P. V. Vemavarapu and C. W. Borst, "Evaluation of a handheld touch device as an alternative to standard ray-based selection in a geosciences visualization environment," in *Proc. Workshop Off-Shelf Virtual Reality*, 2013, pp. 1–8.
- [19] M. Prachyabrued, D. L. Ducrest, and C. W. Borst, "Handymap: A selection interface for cluttered VR environments using a tracked hand-held touch device," in *Proc. Int. Symp. Vis. Comput.* Berlin, Germany: Springer, Sep. 2011, pp. 45–54.
- [20] W. Kim and S. Bang, "VRMouse: Mouse emulation with the VR controller for 2D selection in VR," *SSRN Electron. J.*, 2018.
- [21] R. J. Teather and W. Stuerzlinger, "Pointing at 3D target projections with one-eyed and stereo cursors," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst. (CHI)*, 2013, pp. 159–168.
- [22] P. Lubos, G. Bruder, and F. Steinicke, "Analysis of direct selection in head-mounted display environments," in *Proc. IEEE Symp. 3D User Interfaces (3DUI)*, Mar. 2014, pp. 11–18.
- [23] Y. Y. Qian and R. J. Teather, "The eyes don't have it: An empirical comparison of head-based and eye-based selection in virtual reality," in *Proc. 5th Symp. Spatial User Interact. (SUI)*, 2017, pp. 91–98.
- [24] A. Ramcharitar and R. J. Teather, "A head coupled cursor for 2D selection in virtual reality," in *Proc. 5th Symp. Spatial User Interact. (SUI)*, 2017, p. 162.
- [25] T. S. Young, R. J. Teather, and I. S. MacKenzie, "An arm-mounted inertial controller for 6DOF input: Design and evaluation," in *Proc. IEEE Symp. 3D User Interfaces (3DUI)*, Mar. 2017, pp. 26–35.
- [26] I. S. MacKenzie and W. Buxton, "Extending Fitts' law to two-dimensional tasks," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst. (CHI)*, 1992, pp. 219–226.
- [27] I. S. MacKenzie, "Fitts' law as a research and design tool in human-computer interaction," *Hum. Comput. Interact.*, vol. 7, no. 1, pp. 91–139, Mar. 1992.
- [28] N. Bevan, "International standards for HCI and usability," *Int. J. Hum.-Comput. Stud.*, vol. 55, no. 4, pp. 533–552, Oct. 2001.
- [29] S. Jang, W. Stuerzlinger, S. Ambike, and K. Ramani, "Modeling cumulative arm fatigue in mid-air interaction based on perceived exertion and kinetics of arm motion," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2017, pp. 3328–3339.
- [30] F. Argelaguet and C. Andujar, "A survey of 3D object selection techniques for virtual environments," *Comput. Graph.*, vol. 37, no. 3, pp. 121–136, May 2013.
- [31] A. Ramcharitar and R. J. Teather, "EZCursorVR: 2D selection with virtual reality head-mounted displays," in *Proc. 44th Graph. Interface Conf.*, 2018, pp. 123–130.
- [32] W. Büschel, A. Mitschick, T. Meyer, and R. Dachsel, "Investigating smartphone-based pan and zoom in 3D data spaces in augmented reality," in *Proc. 21st Int. Conf. Hum.-Comput. Interact. Mobile Devices Services (MobileHCI)*, vol. 2, 2019, pp. 1–13, doi: [10.1145/3338286.3340113](https://doi.org/10.1145/3338286.3340113).
- [33] A. Gupta, C. Ji, H.-S. Yeo, A. Quigley, and D. Vogel, "RotoSwype: Word-gesture typing using a ring," in *Proc. CHI Conf. Hum. Factors Comput. Syst. (CHI)*, vol. 14, 2019, pp. 1–12, doi: [10.1145/3290605.3300244](https://doi.org/10.1145/3290605.3300244).
- [34] R. Arora, K. R. Habib, T. Grossman, G. Fitzmaurice, and K. Singh, "SymbiosisSketch: Combining 2D & 3D sketching for designing detailed 3D objects in situ," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2018, pp. 1–15.



MINGHUI SUN received the Ph.D. degree in computer science from the Kochi University of Technology, Japan, in 2011. He is currently an Associate Professor with the College of Computer Science and Technology, Jilin University, China. He is interested in using HCI methods to solve challenging real-world computing problems in many areas, including tactile interface, pen-based interface, and tangible interface.



MINGMING CAO received the bachelor's degree from Baicheng Normal University, China, in 2016. He is currently pursuing the master's degree with the College of Computer Science and Technology, Jilin University. His research interests include human-computer interaction (HCI) and AR and cross device interaction.



LIMIN WANG received the Ph.D. degree in computer science from Jilin University, China, in 2005. He is currently a Professor with the College of Computer Science and Technology, Jilin University. He has published innovative articles in journals, such as *Knowledge-Based Systems*, *Expert System With Applications*, and *Progress in Natural Science*. His research interests include probabilistic logic inference and Bayesian networks.



QIAN QIAN received the Ph.D. degree in computer science from the Kochi University of Technology, Japan, in 2011. He is currently an Associate Professor with the Kunming University of Science and Technology, China. His research interests include visual cognitive psychology and cognitive computational modeling.

• • •