Evaluation of Hand-Based Interaction for Near-Field Mixed Reality with Optical See-Through Head-Mounted Displays

Zhenliang Zhang[1], Benyang Cao[1], Dongdong Weng[1,2,*], Yue Liu[1,2], Yongtian Wang[1,2], Hua Huang[3]

[1] Beijing Engineering Research Center of Mixed Reality and Advanced Display, School of Optics and Photonics, Beijing Institute of Technology, Beijing, China, 100081

[2] Advanced Innovation Center for Future Visual Entertainment, 4 Xitucheng Rd, Haidian, Beijing, China, 100088

[3] School of Computer Science, Beijing Institute of Technology, Beijing, China, 100081

Abstract

Hand-based interaction is one of the most widely-used interaction modes in the applications based on optical see-through head-mounted displays (OST-HMDs). In this paper, such interaction modes as gesture-based interaction (GBI) and physics-based interaction (PBI) are developed to construct a mixed reality system to evaluate the advantages and disadvantages of different interaction modes for near-field mixed reality. The experimental results show that PBI leads to a better performance of users regarding their work efficiency in the proposed tasks. The statistical analysis of T-test has been adopted to prove that the difference of efficiency between different interaction modes is significant.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Graphical user interfaces

1 Introduction

Human-computer interaction is an important issue in mixed reality. With the appearance of commercial depth sensors, many researchers begin to combine depth sensors with other devices, thus constructing new interaction modes for mixed reality. The commonly-used depth sensors, such as Microsoft Kinect, Intel RealSense and Leap Motion controller, can not only provide the depth images of scenes, but also make estimation about the positions of the skeleton of human hands or bodies. There are different kinds of interaction modes when the depth sensors are integrated, in which two most typical modes are gesture based interaction (GBI) and physics-based interaction (PBI). GBI is based on the recognition of different gestures, and widely used in various systems such as WUW [4] and Mime [1]. PBI is based on the detection of hand shape and construction of virtual hand skeletons. Users can use the virtual hands to directly manipulate virtual objects, such as SpaceTop [3] and HoloDesk [2]. However, because the hands of users can be seen by themselves directly, the application of GBI and PBI in mixed reality is complicated.

In this paper, we evaluate the hand-based interaction for near-field mixed reality with optical see-through head-mounted displays (OST-HMDs). Our contributions are as follows. (1) Based on an OST-HMD, we have designed the GBI method and the PBI method for mixed reality applications. (2) Comparative experiments containing two tasks are implemented to analyze the performance of two interaction modes (i.e., GBI and PBI); (3) Experimental data are analyzed with statistical methods.

2 Construction of GBI and PBI

To perform GBI, two gestures are defined in this paper. One is the open palm, and the other is the fist, as shown in the left side of Fig 1. During the implementation, a gesture with different directions of movement denotes different meanings. For example, an open palm gesture with a movement to the right denotes that the selected virtual object should be moved to the right, and its velocity of movement is decided by the velocity of hands. In addition, another parameter is the start threshold of accumulated frames for a certain gesture command, which can prevent the error resulting from jitters. The two parameters (i.e. moving velocity and start threshold) above should be determined by a pilot test before the formal experiment.

*crgj@bit.edu.cn

2018 IEEE Conference on Virtual Reality and 3D User Interfaces
18-22 March, Reutlingen, Germany
978-1-5386-3365-6/18/$31.00 ©2018 IEEE

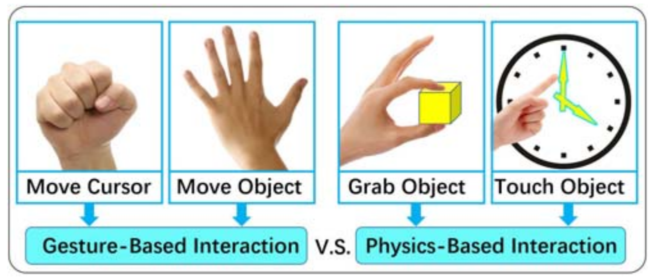

Figure 1: Four specific operations by GBI and PBI. The two on the left are two gestures used in GBI, and the two on the right are two interaction situations in direct interaction with PBI.

To perform PBI, the real-time positions of hands are required. There are two means to realize the goal. First, the spatial positions could be used to determine whether one action should be triggered. For a certain virtual object, it can be defined that the object is grabbed only if the virtual object is collided by the collision bodies of index and thumb. Second, certain tasks are implemented by physical collision between real hands and virtual objects. The collision effects are affected by the settings of physical properties. The example scene of PBI is shown in the right side of Fig. 1.

Especially, when the hand is integrated into a HMD-based mixed reality system, the situation should be considered for two interaction modes respectively. First, the mode using GBI does not need the calibration of OST-HMD, because the spatial position of hands is not required to be registered to the OST-HMD. Second, the mode using PBI relies on an accurate calibration of OST-HMD. To ensure that the real hands coincide with the virtual hands, the calibration of the depth sensor and OST-HMD should be performed using the single point active alignment method [5] during the initialization of the system.

3 Experiment

3.1 System Setup

The system is a mixed reality system based on an OST display as shown in Fig. 2 (a). The OST display is a core component to construct an OST-HMD. The resolution of the internal micro projector is 800*600 pixels. A Leap Motion controller is attached to the OST display firmly. The OST display and the depth sensor are combined using a metal bracket.

3.2 Experimental Design

20 volunteers (9 males and 11 females) are invited to join in the experiment, whose ages range from 22 to 28 and all have never joined in similar experiments. The volunteers are divided into two groups (Group A and Group B). The volunteers in Group A are

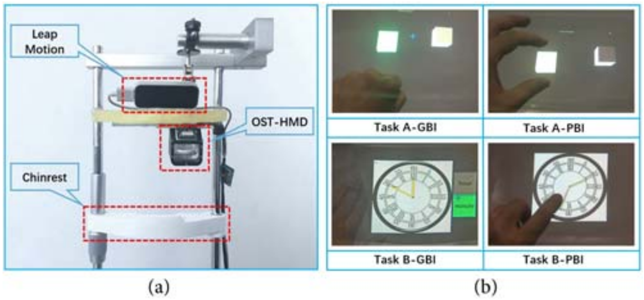

Figure 2: System overview. (a) is the device used in experiments. (b) is the first-person views for different tasks.

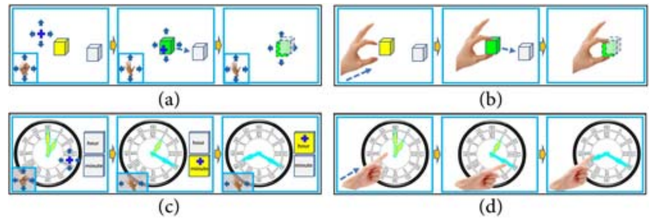

Figure 3: Detailed processes of tasks. (a)The description of Task A-GBI. (b)The description of Task A-PBI. (c)The description of Task B-GBI. (d)The description of Task B-PBI.

composed of 4 males and 6 females, while the volunteers in Group B are composed of 5 males and 5 females. Two tasks are designed to test the efficiency of users. Intuitively, the higher efficiency the users can obtain, the better the corresponding interaction mode is.

3.2.1 Task A: Pick-and-Place

Task A is a pick-and-place task, in which users are asked to pick the yellow cube to the position of white cube for 10 times. When the cube is selected, its color will change to green, which helps the user to know whether the cube is selected. The yellow cube and the white cube are in similar depths, and both have three random positions. After the user moves the yellow cube to the white one, the yellow cube and the white cube will first disappear and then appear in new positions.

Task A-GBI denotes task A with GBI mode, which is shown in Fig. 3 (a). In this task, we have defined two gestures: fist and open palm. When the gesture is the fist, the blue cursor on the screen will move as the hand moves. When the cursor is moved onto the yellow cube, the yellow cube is thought to be selected. When the gesture is the open palm, the selected cube will move as the hand moves.

Task A-PBI denotes task A with PBI mode, which is shown in Fig. 3 (b). In this task, the spatial position of every fingertip is detected. When the cube is collided by the thumb fingertip and the index fingertip, it is thought to be selected. And after the yellow cube is selected, the user can hold it and grab it to the position of the white cube.

3.2.2 Task B: Select-and-Alignment

Task B is a select-and-alignment task. Users are asked to align the hour and minute hands of a virtual clock according to a given time. The user is asked to complete the task for 4 times, then the consumed time of each round is recorded.

Task B-GBI denotes the task B with GBI mode, which is shown in Fig. 3 (c). In this task, there are two virtual buttons to assist the user to confirm which hand (hour hand or minute hand) is selected. When the cursor is moved onto one of the virtual button with the fist gesture, the user can change to an open palm gesture to make the hour hand or the minute hand rotate around the central point.

Task B-PBI denotes the task B with PBI mode, which is shown in Fig. 3 (d). In this task, the user can directly touch the clock hands with the real fingers. The real finger can directly move the virtual clock hands with the detection of collision body and rigid body.

3.2.3 Experimental Arrangement

Participants of Group A should first perform two tasks with GBI first, and then perform the same two tasks with PBI. Group B and Group A are in the opposite order regarding interaction modes. Before every task, the users are trained until they know how to successfully perform the task. The first-person views of experiments are shown in Fig. 2 (b).

4 Results

All the subjects are divided into two groups with different orders of tasks (GBI–>PBI versus PBI–>GBI), whose purpose is to verify whether the different orders of tasks would have different effects on the subjects in the experiment. Independent T-tests are performed to identify whether there are significant differences between different orders of tasks (GBI–>PBI versus PBI–>GBI) for subjects regarding their consumed time in different tasks. The statistical result shows that there is no significant difference between different orders of tasks regarding their consumed time in all tasks (Task A-GBI, $p = 0.235 > 0.05$; Task A-PBI, $p = 0.234 > 0.05$; Task B-GBI, $p = 0.809 > 0.05$; Task B-PBI, $p = 0.170 > 0.05$). For the same group of subjects, two different interaction modes (GBI and PBI) are used for both Task A and Task B. Considering the interaction mode, Task A is divided into Task A-GBI and Task A-PBI. Similarly, Task B is divided into Task B-GBI and Task B-PBI, whose purpose is to verify whether different interaction modes would have significant differences on the consumed time of tasks. A paired T-test is performed to identify whether there is a significant difference regarding the average consumed time between different interaction modes for the same task in the same group of subjects. The result of paired T-test indicates that there is a significant difference between their consumed time on different interaction modes in both Task A and Task B (In Task A, $p \approx 0 < 0.05$; In Task B, $p = 0.021 < 0.05$). It can be seen that the subjects would obtain a significantly better performance in the proposed tasks using PBI instead of GBI.

5 Conclusion

Two interaction modes for near-field mixed reality have been compared in this paper. Before the experiments, the GBI is designed and the mechanism of PBI is also determined. Two mixed reality applications are developed to test whether there is different effect on the efficiency of users when different interaction modes are used. Experiments prove that, under the current condition, the PBI would lead to a better performance of users.

Acknowledgments

This work has been supported by the National Key R&D Program of China (No. 2016YFB1001502) and the National Natural Science Foundation of China under Grant 61727808.

References

[1] A. Colao, A. Kirmani, H. S. Yang, N. W. Gong, C. Schmandt, and V. K. Goyal. *Mime: Compact, low-power 3D gesture sensing for interaction with head-mounted displays*. ACM, 2013.

[2] O. Hilliges, D. Kim, S. Izadi, M. Weiss, and A. Wilson. Holodesk: direct 3d interactions with a situated see-through display. In *Sigchi Conference on Human Factors in Computing Systems*, pp. 2421–2430, 2012.

[3] J. Lee, A. Olwal, H. Ishii, and C. Boulanger. Spacetop:integrating 2d and spatial 3d interactions in a see-through desktop environment. In *Sigchi Conference on Human Factors in Computing Systems*, pp. 189–192, 2013.

[4] P. Mistry, P. Maes, and L. Chang. Wuw - wear ur world:a wearable gestural interface. In *International Conference on Human Factors in Computing Systems, CHI 2009, Extended Abstracts Volume, Boston, Ma, Usa, April*, pp. 4111–4116, 2009.

[5] M. Tuceryan and N. Navab. Single point active alignment method (spaam) for optical see-through hmd calibration for ar. In *IEEE and ACM International Symposium on Augmented Reality*, pp. 149–158, 2000.