2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
Macau, China, November 4-8, 2019

Toward an Efficient Hybrid Interaction Paradigm for Object Manipulation in Optical See-Through Mixed Reality

Zhenliang Zhang[1] Dongdong Weng[1,2,*] Jie Guo[1] Yue Liu[1,2] Yongtian Wang[1,2]

***Abstract*— Human-computer interaction (HCI) plays an important role in the near-field mixed reality, in which the hand-based interaction is one of the most widely-used interaction modes, especially 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. The ultimate goal is to find an efficient hybrid paradigm for mixed reality applications based on OST-HMDs to deal with the situations that a single interaction mode cannot handle. The results of the experiment, which compares GBI and PBI, show that PBI leads to a better performance of users regarding their work efficiency in the proposed two tasks. Some statistical tests, including T-test and one-way ANOVA, have also been adopted to prove that the difference regarding the efficiency between different interaction modes is significant. Experiments for combining both interaction modes are put forward in order to seek a good experience for manipulation, which proves that the partially-overlapping style would help to improve work efficiency for manipulation tasks. The experimental results of the proposed two hand-based interaction modes and their hybrid forms can provide some practical suggestions for the development of mixed reality systems based on OST-HMDs.**

I. Introduction

Mixed reality based on optical see-through head-mounted displays (OST-HMDs) is a new interaction paradigm in recent years. However, a mature interaction mode with high efficiency and low learning cost is still not fully discovered from various interaction modes. Among these interaction modes, hand-based interaction is a commonly-used one, so finding a suitable hand-based interaction method is of much importance. With the emergence of commercial depth sensors, many researchers began to combine depth sensors with other display devices, thus constructing new interaction modes in mixed reality [1]. The depth sensors, such as Microsoft Kinect, Intel RealSense, and Leap Motion controller, can provide the depth images of scenes and make an estimation about the positions of the skeletons of human hands or bodies. There are different kinds of hand-based interaction modes when the depth sensors are integrated, and the most typical modes are the gesture-based interaction (GBI) which exploits gesture recognition, and the physics-based interaction (PBI) which exploits physical detection.

With the rapid development of mixed reality systems [2], a natural human-computer interaction mode is required, especially in the near-field mixed reality applications based on head-mounted displays. Both GBI and PBI have been widely used to construct interaction systems. For example, GBI is used in interactive computer games [3] and human-robot interaction [4], and PBI is used in such operations that are aimed to create physical effects as screwing and grabbing. However, the two interaction modes can also be used to finish similar tasks, such as controlling the movement of objects and some other detailed manipulation tasks. So there may be some cases that need developers to determine which mode is more suitable to be adopted in these cases and whether a combination of different interaction modes can provide a better performance.It is much helpful for developers to know the different features of different interaction modes (GBI and PBI), when developing a mixed reality interaction system based on optical see-through head-mounted displays. Since the hands of users can be seen by themselves directly, the application of GBI and PBI in mixed reality is more complicated than that in virtual reality. Therefore, the comparison of these hand-based interaction modes appear to be significant.

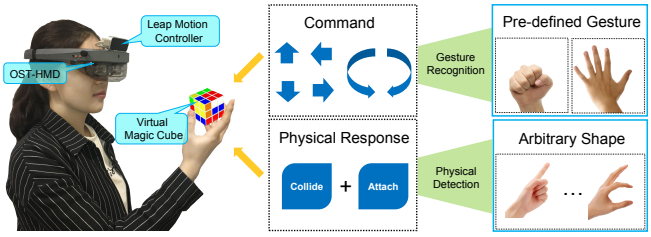


Fig. 1. A typical scene in near-field mixed reality with two hand-based interaction modes.

The study about human-robot interaction is deeply related to hand-based information exchange between humans and robots. A shared workspace for humans and robots should be useful for sharing perceived information and knowledge with each other in real time. With the support of mixed reality, the human and the robot can enter a mixed reality environment and directly show the information to each other. For example, in a manipulation task of objects, humans can teach the robot how to manipulate a certain real object (like a magic cube) by demonstrating the task with a virtual object, and directly supervise the robot's implementation by seeing through the HMD. Therefore, developing an efficient manipulation paradigm with hands is of much importance for a shared workspace involving both humans and robots.

In this paper, we evaluate the two hand-based interaction methods for near-field mixed reality with OST-HMDs (a small part was preliminarily presented as a short poster [5]), and build a typical near-field mixed reality system with a hybrid interaction mode, which is shown in Fig. 1. Our contributions are listed as follows: (i) We have designed the discrete-state GBI method and the continuous-state PBI method for HMD-based mixed reality applications to analyze the different features of them in two given manipulation tasks; (ii) We integrate GBI and PBI to construct a hybrid interaction paradigm in order to find an efficient combination way of GBI and PBI, and the performance of different feasible combination ways has been studied; (iii) Subjective and objective experimental data are analyzed with various statistical methods, which indicate that the hybrid interaction mode constructed by GBI and PBI has potential use for manipulating virtual objects.

* Corresponding Author
[1]Beijing Engineering Research Center of Mixed Reality and Advanced Display, School of Optics and Photonics, Beijing Institute of Technology, Beijing, China.
[2]AICFVE of Beijing Film Academy, 4 Xitucheng Rd, Beijing, China.
Emails: zzlyw10@gmail.com, {crgj, guojie, liuyue, wyt}@bit.edu.cn.

II. Related Work

The related works mainly contain two parts. Some of these are focused on GBI, and the others are related to PBI. We explore the works respectively.

A. Gesture-Based Interaction

Many researchers performed experiments on GBI, and they defined different gestures as different commands to perform spatial manipulation tasks [6]. A much general style of hand-based interaction method is using hands as a kind of graphical interfaces, so that the users can perform spatial interaction with bare hands [7] or some assistant devices [8] such as depth cameras and projectors. In recent years, depth sensors for hand detection have achieved a rapid development. For example, the Leap Motion controller has been proved to be an accurate device [9], which can be used to capture hands. At the same time, the Intel RealSense is also a suitable choice for detecting hands, which has been successfully applied to the Project Alloy [10]. With the help of various depth sensors, some hand-based interaction systems are proposed, especially for some HMD-based systems [11]. Mistry et al. [12] introduced Wear Ur World (WUW), a wearable gestural interface, which brought information out into the physical world. WUW could project information onto surfaces, walls and physical objects in the world, and let the user interact with the projected information through natural hand gestures and arm movements. Colaco et al. [13] designed a device called Mime, which allowed its users to interact with a virtual interface with a hand. After the hand's 3D position was obtained, a user could use different gestures to input different information to the HMD-based augmented reality (AR) system. In addition to the above two AR systems, Khattak et al. [14] introduced a virtual reality system consisting of an Oculus Rift head-mounted display, an RGB-D camera, and a Leap Motion controller, where the finger information could be captured in order to support the hand-related interaction.

Even though gesture-based interaction is a traditional topic, it becomes particular when it is linked to the HMD-based interaction system, and different interaction ways may result in different effects. That is why more attention should be paid to the HMD-based mixed reality.

B. Physics-Based Interaction

PBI is based on the accurate registration between virtual and real objects, especially the registration of hands. The hand and eye can be calibrated using the single point active alignment method (SPAAM) [15]. Some works extended SPAAM by introducing automatic calibration frameworks [16]. Based on an accurate calibration, PBI can promote the development of PBI-based systems.

The direct hand-object interaction [17] has been a hot topic for a long time, especially for the interaction method with feedback [18]. This has facilitated the research about the direct interaction between hands and virtual elements via hand-based interaction methods. Some typical systems are developed to implement this kind of direct interaction with different forms. Hilliges et al. [19] proposed the HoloDesk, which was an interactive system combining an optical see-through display and Kinect camera to create the illusion that users were directly interacting with 3D graphics. They introduced a new technique for interpreting raw Kinect data to approximate and track rigid (e.g., books, cups) as well as non-rigid physical objects and support a variety of PBI between virtual and real objects. This was a classical system which showed that physical hands can directly contact the virtual objects and interact with them. Özacar et al. [20] explored a mid-air direct multi-finger interaction technique to efficiently perform fundamental object manipulations on an interactive stereoscopic tabletop display, which allowed multiple users to directly observe and interact in a shared workspace. Lee et al. [21] proposed an interaction system that allowed users to type, click, draw in 2D, and directly manipulate interface elements that floated in the 3D space above the keyboard. They put forward a concept called SpaceTop, which gave users a feeling that they could manipulate the virtual objects in space. In addition, Piumsomboon et al. [22] presented an AR framework that allowed natural hand and tangible augmented reality interaction for physics-based interaction environment. The above four systems only gave the virtual-real fusion at the virtual level, so Kervegant et al. [23] used HoloLens and a haptic device to construct a human-computer system which could give feedback and enhance the presence of virtual objects. Therefore, the direct interaction, which is a specific kind of physics-based interaction, is also of much importance to construct a hand-based interaction system.

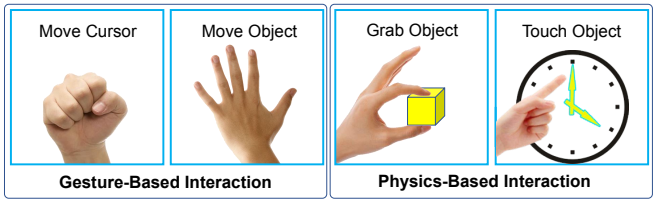


Fig. 2. Four specific operations by GBI and PBI. The left figure shows two gestures that are used in GBI, and the right figure shows two typical interaction situations in PBI.

Many works about GBI and PBI have been presented, but they usually studied about a single method (GBI or PBI) for specific tasks. It is still demanded to compare GBI and PBI together so that the advantages and disadvantages can be clearly presented for researchers and designers.

III. Design of Interaction Mode

There are two design rules in designing the mentioned GBI and PBI. The first rule is simplicity, which leads to a low learning cost and a low use error. The second rule is efficiency, which indicates the value of the given interaction mode. To compare GBI and PBI, as shown in Fig. 2, a 3D interaction system is needed to support two interaction modes.

A. Construction of Gestures

According to the design rule of simplicity, only two gestures are defined in this paper. One is the "open palm", and the other is the "fist". Hands data collected by the depth sensor should be recorded by frames. First, the training data are collected by recording some hands with different sizes. In our research, we ask users with different hand sizes to make "open palm" and "fist" gestures in the detection field of the depth sensor and keep the gestures in different positions and different rotations until enough data are recorded. For every frame of hands data, we compute the sum of Euclidean distances of every finger to other four. When the training data are completed, we use Fine Gaussian Support Vector Machine to classify the data into two classes. Meanwhile, the predictor is created. Since the data are a set of one-dimension data, the classification threshold could be obtained.

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 determined by the velocity of hands. In addition, another parameter is the start threshold for a certain gesture command. For example, if a continuous movement to a direction lasts for more than 5 frames, the gesture command controlling movement in that direction will be triggered, in which the threshold is 5. The two parameters (movement velocity and start threshold) should be determined

by a pilot test before the formal experiment. In our research, the threshold is set to be 5 frames, because the rendered image on the screen is updated at a rate of 60 frames per second, so 5 frames will only take 1/12 second. Such small delay of the rendered image is acceptable for most users. Obviously, the start threshold is a trade-off between the delay of rendered image and the robustness of hands.

For the specific meaning of every gesture, it can be defined according to different tasks. Since we just adopt two simple gestures, the redefinition of meanings for different tasks should be relatively easy. The reason why we only use two gestures for the experiment is that two gestures are enough for our designed tasks, and at the same time it minimizes the effort of users to learn how to use the system. In addition, the reason for selecting the "fist" and the "open palm" is just that the two gestures are easy to be recognized, which can ensure the stability of the system.

B. Physical Mechanism for Direct Interaction

To perform a direct PBI, the real-time poses of hand skeletons 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 has collided with both the index finger and the thumb. Then the virtual object will be attached to the center of the index finger and the thumb. The virtual object can be released by separating the index finger and the thumb. Second, certain tasks are implemented by the physical collision between real hands and virtual objects. The collision effects are affected by the settings of physics. For the two means above, the parameters should be set up according to the physical world.

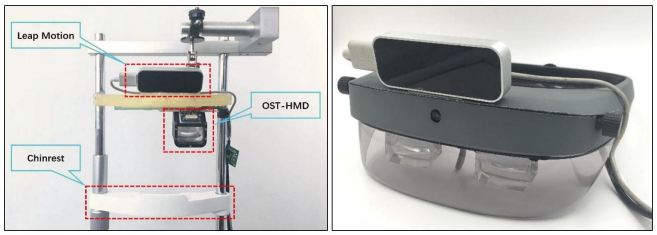
More specifically, the PBI in this paper can be realized through 3 steps. First, the hand of a user is captured by a sensor, and the hand shape, including all poses of skeletons of this hand, can be reconstructed in real time. Second, the data of the reconstructed hand is used to create an identical virtual hand, whole spatial parameters are the same as the user's real hand. Third, the virtual hand can attach a mesh collision body, so that it can interact with other objects in virtual space based on the physical collision rules. In this way, PBI can give the user a feeling that the real hand can interact with the virtual objects at the visual level. This kind of interaction mode is the same as what people use in daily life, so it is expected to be more easily adopted by users.

C. Near-Field Interaction Based on OST-HMD

A general goal to use the OST-HMD is to create a mixed reality environment. Especially, when the hand is integrated into the system, we should consider different interaction modes respectively.

First, the mode using GBI does not need the calibration of OST-HMDs, because the spatial position of hands is not required to be registered to the OST-HMD. During the process of performing gestures, the absolute position of hands is not related to the corresponding gesture commands. GBI is only depended on the gesture and the relative movement between two adjacent frames. Though the relative movement is determined by the change of the physical position, GBI is mainly based on the recognition of gestures.

Second, the mode using PBI relies on an accurate calibration of OST-HMDs. To ensure that the real hands coincide with the virtual hands, the calibration of the depth sensor and the OST-HMD should be implemented during the initialization of the system, in which SPAAM is used to calibrate the transformation between the depth sensor and the OST-HMD. In a typical SPAAM, at least 6 pairs of 3D-2D corresponding points are needed.



(a) (b)

Fig. 3. System overview. (a) is used in the experiments for comparing two interaction modes. (b) is a binocular OST-HMD, which is used in an simple application combining two different interaction modes.

IV. Experimental Design

A. System Setup

The system is a mixed reality system based on an OST display as shown in Fig. 3 (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. Fig. 3 (b) is a binocular OST-HMD (NED+ X1), which can be used to develop applications.

In order to make the system support both PBI and GBI, some preparation work should be done before the formal experiments. For PBI, the calibration of the hand and the HMD should be performed. Theoretically, every subject should calibrate for his own experiment, but for general tasks that do not need an extremely high precision, we can just use a set of pre-calibrated parameters for all subjects to simplify the procedures. For GBI, we should train the adopted gestures in advance so as to make new subjects can use the system directly. In sum, both PBI and GBI should rely on some preparation works, so we only focus on the online implementation when comparing the interaction modes.

B. Experiment I: Comparison between GBI and PBI

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 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 subjects. Intuitively, the efficiency of subjects would indicate the acceptance of different interaction modes.

1) Task A: Pick-and-Place: Task A is performed in two different ways: one uses GBI and the other uses PBI. For convenience, the one using GBI is denoted as Task A-GBI, while the other using PBI is denoted as Task A-PBI.

Task A is a pick-and-place task, in which subjects are asked to pick the yellow cube to the position of the white cube for 10 times. When the cube is selected, its color will change to green. Here the visual feedback of changing color is adopted in determining whether the object is selected.The yellow cube and the white cube are in similar depths, and both have three random positions. After the subject 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. 4 (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. 4 (b). In this task, the spatial position of every fingertip is detected. When the cube is collided by the thumb fingertip

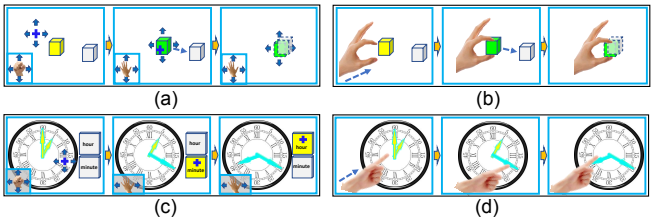


Fig. 4. Detailed processes of tasks. (a) Task A-GBI: grab a cube from one position to another. The yellow color means its initial position. The green color means the cube is being moved by hands. The white cube denotes the target position. The cube is controlled by "fist" gesture and "open palm" gesture. (b) Task A-PBI: use hands directly move the yellow cube. The color's meanings are the same as that in (a). (c) Task B-GBI: align clock hands to specific positions using gestures. (d) Task B-PBI: align clock hands to specific positions using direct hand interaction based on physics.

and the index fingertip, it is thought to be selected. And after the yellow cube is selected, the subject can hold it and grab it to the position of the white cube.

2) Task B: Select-and-Alignment: Select-and-alignment is another perspective to assess an interaction method. This task requires subjects to perform a more precise operation, especially when PBI is used.

Task B is a select-and-alignment task. Subjects are asked to align the hour and minute hands of a virtual clock according to a given time. To avoid particularity, the subject is asked to complete the task for 4 times, then the time of each round is recorded. Task B asks the users to do 4 times which are not as many as that in Task A, because every attempt for task B requires much more effort and time, so it is not necessary to collect a lot of data to compute the average. For every attempt, the initial state of the clock is "1:00", and the four final states are "8:20", "10:10","3:00" and "6:30". When the clock hands are kept still at the right final positions for more than 3 seconds, it means that the subject has successfully finished the current task. Then the clock hands are reset to the initial positions waiting for a new attempt.

Task B-GBI denotes the task B with GBI mode, which is shown in Fig. 4 (c). In this task, there are two virtual buttons to assist the subject to confirm which hand (hour hand or minute hand) is selected. When the cursor is moved onto one of the virtual buttons with the "fist" gesture, the subject can change to an "open palm" gesture to make the hour hand or the minute hand rotate around a central point.

Task B-PBI denotes the task B with PBI mode, which is shown in Fig. 4 (d). In this task, the subject can directly touch the clock hands with the real fingers. It appears that the real finger can move the virtual clock hands with the physical detection.

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 for interaction modes. Before every task, the subjects are trained for 5 minutes in order that they know how to successfully perform the task. The detailed experimental arrangement is listed in Table I. And the real process of these experiments is shown in Fig. 5.

TABLE I

DETAILED EXPERIMENTAL ARRANGEMENT.

Steps	Group A	Group B
(1-a): Learning	Task A-GBI	Task A-PBI
(1-b): Performing	Task A-GBI	Task A-PBI
(2-a): Learning	Task B-GBI	Task B-PBI
(2-b): Performing	Task B-GBI	Task B-PBI
(3-a): Learning	Task A-PBI	Task A-GBI
(3-b): Performing	Task A-PBI	Task A-GBI
(4-a): Learning	Task B-PBI	Task B-GBI
(4-b): Performing	Task B-PBI	Task B-GBI

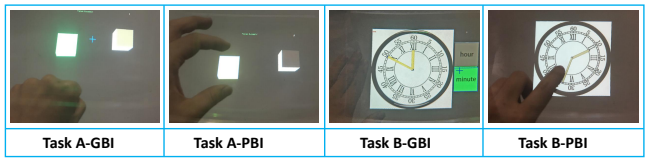


Fig. 5. Real scenes captured in the first-person perspective.

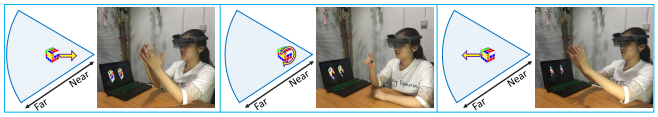


Fig. 6. The task of manipulating magic cube using the proposed hybrid method. (a) Dragging the cube toward the user. (b) Detailed manipulation in near-filed. (c) Pushing the cube.

C. Experiment II: Hybrid Interaction

Since Sect. IV-B has compared GBI and PBI regarding their advantages and disadvantages, there should be some tasks that may need both two interaction modes. Based on this motivation, we propose an experiment combining GBI and PBI to find a feasible combination way of the proposed two modes. Since PBI is more natural than GBI, PBI should be used as the main interaction mode, especially in the hand-accessible area in front of the eyes. Meanwhile, GBI should be used beyond the hand-accessible area as a supplement of PBI. The combination method is designed in this way because PBI cannot manipulate the virtual objects that are not located in the hand-accessible area.

The task is manipulating a magic cube because this task is complicated enough to use different kinds of interaction modes. The whole process is shown in Fig. 6. This task of manipulating the magic cube is decomposed into 3 procedures. First, the magic cube is initialized at a far position from the subject's viewpoint (the viewpoint is located at the center of two eyes), the subject should move the magic cube to a near position to perform some more detailed actions like rotating. Second, the subject should solve the magic cube by rotating its different surfaces. Here the magic cube can be solved by exact 3 rotations. Third, move the solved magic cube back to its original position. In the whole process, the subject can use different interactions according to corresponding settings that are designed for different combination ways of GBI and PBI. The working space can be divided into two subspaces: the hand-accessible area and the hand-inaccessible area. There are three combination ways, which are shown in Fig. 7. For the zero-overlapping, the two subspaces are divided by a surface, and the distance between the subject and the surface is the length of subject's arm, so that the subject can cover all the hand-accessible area. For the partially-overlapping, the hand-accessible area and the hand-inaccessible area share a part of space, and two kinds of modes can be used in this space. For the completely-overlapping, the hand accessible area is totally covered by the hand-inaccessible area. In the hand-accessible area, the PBI can be performed easily, since all objects in this area can be reached by stretching out the arm. But if the objects are in a far position beyond the arm's reach, GBI is needed to do some remote control.

We have asked 10 subjects (graduate students) to perform this task. The first-person view of subjects is shown in Fig. 8. First, every subject should learn how to do the task, and both of the two interaction modes used in tasks are taught to subject. Then the subject performs the same task for 3 times corresponding to 3 combination ways of GBI and PBI. The consumed time for every attempt is recorded to represent the efficiency. Since a lot of experiments have been performed in Sect. IV-B, here we use the consumed time to intuitively show the performance of different interaction modes. Under the condition that all other factors are fully controlled, a shorter consumed time of tasks could indicate a better performance.

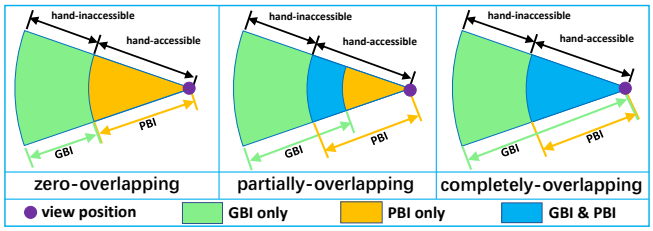


Fig. 7. The combination ways considering the hand-accessible depth. (left) GBI and PBI are combined with a parting surface. (middle) GBI and PBI are combined with an overlapped space. (right) GBI and PBI can be used in the whole space.

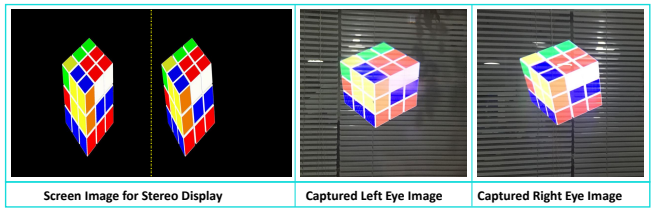


Fig. 8. Stereo images displayed on the OST-HMD. An image for stereo can be divided into two parts at the vertical central line, and shown for the left eye and right eye respectively.

V. Experimental Result

We analyze the experimental results for Experiment I and Experiment II respectively.

A. Result of Experiment I

1) Work Efficiency: 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. Table II indicates that there is not a 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, $p = 0.170 > 0.05$).

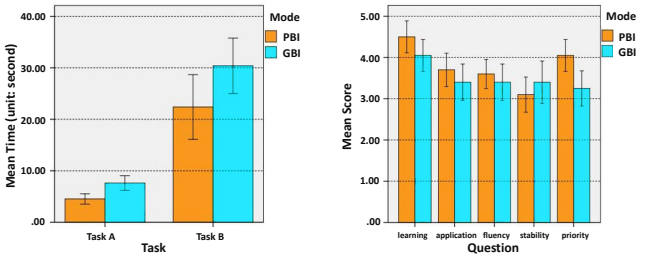
For the same group of subjects, two different interaction modes (GBI and PBI) are used for both Task A and Task B. Considering the interaction mode, Task A is divided into Task A-GBI and Task A-PBI. Similarly, Task B is divided into Task B-GBI and Task B-PBI, whose purpose is to verify whether different interaction modes have a significant difference regarding consumed time of tasks.

A paired T-test is performed to identify if there is a significant difference between the average consumed time of different interaction modes for the same task in the same group of subjects. The paired T-test result in Table III indicates that there are significant differences between their consumed time on different interaction modes in both Task A and Task B (In Task A, $p < 0.001 < 0.05$; In Task B, $p = 0.021 < 0.05$).

TABLE II

Comparison of the effect of different order of tasks on performance of different subjects.

Variable	M	SD	t	df	p
Time in Task A-GBI			1.23	18	0.235
GBI–>PBI	8.44	4.06			
PBI–>GBI	6.80	1.15			
Time in Task A-PBI			-1.232	18	0.234
GBI–>PBI	3.94	2.24			
PBI–>GBI	5.11	2.01			
Time in Task B-GBI			0.246	18	0.809
GBI–>PBI	31.03	13.35			
PBI–>GBI	29.73	10.08			
Time in Task B-PBI			1.461	11.70	0.170
GBI–>PBI	26.67	17.18			
PBI–>GBI	18.14	6.73			



(a) (b)

Fig. 9. Comparison of two interaction modes. (a) Comparison result of mean time that consumed in regarding different tasks. (b) Comparison result regarding the scores of five questions in the questionnaire. The meanings of the labels at horizontal axis: "learning" denotes learning difficulty, "application" denotes application difficulty, "fluency" denotes degree of fluency, "stability" denotes stability of control, "priority" denotes degree of recommendation.

It can be seen from the result that the subjects would obtain a significantly better performance in the proposed tasks using PBI instead of GBI. The difference of the consumed time between PBI and GBI is shown in Fig. 9 (a).

2) User Experience: To know the subjective assessment of subjects regarding the two interaction modes, we have designed a questionnaire according to the principle of the Likert scale, which is shown in Table IV. There are five aspects to be scored, i.e., the learning difficulty, the application difficulty, the degree of fluency, the stability of control and the degree of recommendation. In the questionnaire, the subjects are asked to give each aspect a score from 1 to 5. Generally, a higher score denotes a higher performance or lower difficulty.

All the subjects are asked to fill the questionnaire after finishing the whole experiment, whose purpose is to verify whether the different orders of tasks would have different effects on the scores of subjects in the experiment.

Mann-Whitney U tests are performed to identify whether there are significant differences between different orders of tasks for the subjects regarding their scores for the given questions. The results of Mann-Whitney U tests indicate that there is not a significant difference between different orders of tasks at a significance level of 5% regarding the scores for every question.

For the same group of subjects, the scores of two interaction modes for every question are given, whose purpose is to verify whether there is a significant difference between the two interaction modes for every question regarding the scores.

A Wilcoxon test is performed to identify whether there is a significant difference between the scores on two different interaction modes for the same group of subjects. The results of the Wilcoxon test indicate that there is a significant difference between the scores of the two interaction modes about the degree of recommendation ($p = 0.024 < 0.05$). However, there is no significant difference between the scores

TABLE III

Comparison of the effect of different interaction modes on the same task.

Mode	M	SD	t	df	p
Task A			4.263	19	<0.001
Task A-GBI	7.62	3.03			
Task A-PBI	4.52	2.15			
Task B			2.524	19	0.021
Task B-GBI	30.38	11.53			
Task B-PBI	22.40	13.43			

TABLE IV

Questionnaire about user experience.

Questions for Evaluation	Score
Q1: learning difficulty	(not easy)1 – 2 – 3 – 4 – 5(very easy)
Q2: application difficulty	(not easy)1 – 2 – 3 – 4 – 5(very easy)
Q3: degree of fluency	(not fluent)1 – 2 – 3 – 4 – 5(very fluent)
Q4: stability of control	(not stable)1 – 2 – 3 – 4 – 5(very stable)
Q5: recommendation	(very low)1 – 2 – 3 – 4 – 5(very high)

of the two interaction modes about the other four questions (for learning difficulty, $p = 0.058 > 0.05$; for application difficulty, $p = 0.395 > 0.05$; for the degree of fluency, $p = 0.521 > 0.05$; for the stability of control, $p = 0.383 > 0.05$). A more intuitive result is in Fig. 9 (b), which shows more detailed results about the comparison.

B. Result of Experiment II

We have designed one task for subjects using this system. For an actual application, we modify the experimental system to a typical binocular OST-HMD, which is in Fig. 3 (b).

This adopted task aims to test the efficiency in detailed manipulation that may need both GBI and PBI. The subject is asked to restore a disrupted magic cube, which only needs three steps to recover. In this experiment, three different combining ways of GBI and PBI are employed. To compare different combining ways, one-way ANOVA is implemented. A statistically significant difference ($p < 0.001 < 0.05$) is found among the three combining ways of the given interaction modes regarding the consumed time in the magic cube task, which is shown in Table V. Meanwhile, the mean of the consumed time is 41.60 seconds for zero-overlapping combination, 34.30 seconds for partially-overlapping combination, and 56.40 seconds for completely-overlapping combination. Tukey's HSD Post Hoc Tests indicate that there is a significant difference between the completely-overlapping way and the zero-overlapping way ($p < 0.001 < 0.05$). Similarly, there is also a significant difference between the completely-overlapping way and the partially-overlapping way ($p < 0.001 < 0.05$). Notice that the mean time using partially-overlapping way is obviously lower than that using zero-overlapping way, though the difference is not statistically significant ($p = 0.099 > 0.05$).

VI. 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. During Experiment I, two mixed reality tasks are adopted to test whether there are different influences on the efficiency and experience of users when different interaction modes are exploited. Experiments prove that, under the current condition, the PBI would lead to a better performance of users. And the PBI has been recommended with a higher priority, though its stability of control is not as good as GBI. As for Experiment II, a good combination way of PBI and GBI may be necessary to obtain an efficient hybrid mode for near-field mixed reality applications. The partially-overlapping hybrid mode provides a feasible reference plan for seeking excellent hand-based hybrid interaction mode.

This paper has focused on the evaluation of two typical hand-based interaction modes in mixed reality and the exploration of the hand-based hybrid interaction paradigm. With the obtained results, we can design more suitable virtual interfaces for robot systems and make more efficient and innovative interactive environments for human-robot interaction.

Acknowledgment

This work was supported by the National Key Research and Development Program of China (No.2017YFB1002805), the National Natural Science Foundation of China (No.61731003) and the 111 Project (B18005).

TABLE V

One-way analysis of variance summary table comparing different combining ways regarding consumed time.

Source	df	SS	MS	F	p
Consumed Time					
Between groups	2	2535.80	1267.90	21.96	<0.001
Within groups	27	1558.90	57.74		
Total	29	4094.70			

References

[1] C. Weichel, M. Lau, D. Kim, N. Villar, and H. W. Gellersen, "Mixfab: a mixed-reality environment for personal fabrication," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2014, pp. 3855–3864.

[2] H. Benko, R. Jota, and A. Wilson, "Miragetable: freehand interaction on a projected augmented reality tabletop," in *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 2012, pp. 199–208.

[3] H.-S. Yeo, B.-G. Lee, and H. Lim, "Hand tracking and gesture recognition system for human-computer interaction using low-cost hardware," *Multimedia Tools and Applications*, vol. 74, no. 8, pp. 2687–2715, 2015.

[4] H. Liu, Y. Zhang, W. Si, X. Xie, Y. Zhu, and S.-C. Zhu, "Interactive robot knowledge patching using augmented reality," in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 1947–1954.

[5] Z. Zhang, B. Cao, D. Weng, Y. Liu, Y. Wang, and H. Huang, "Evaluation of hand-based interaction for near-field mixed reality with optical see-through head-mounted displays," in *Proceedings of IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2018, pp. 739–740.

[6] S. Wu, A. Ricca, A. Chellali, and S. Otmane, "Classic3d and single3d: Two unimanual techniques for constrained 3d manipulations on tablet pcs," in *Proceedings of IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 2017, pp. 168–171.

[7] S. Gustafson, D. Bierwirth, and P. Baudisch, "Imaginary interfaces: spatial interaction with empty hands and without visual feedback," in *Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology*. ACM, 2010, pp. 3–12.

[8] C. Harrison, H. Benko, and A. D. Wilson, "Omnitouch: wearable multitouch interaction everywhere," in *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*. ACM, 2011, pp. 441–450.

[9] F. Weichert, D. Bachmann, B. Rudak, and D. Fisseler, "Analysis of the accuracy and robustness of the leap motion controller," *Sensors*, vol. 13, no. 5, pp. 6380–6393, 2013.

[10] D. Diakopoulos and A. K. Bhowmik, "Project alloy: An all-in-one virtual and merged reality platform," *Journal of the Society for Information Display*, vol. 48, no. 1, pp. 19–22, 2017.

[11] J.-Y. Lee, H.-M. Park, S.-H. Lee, S.-H. Shin, T.-E. Kim, and J.-S. Choi, "Design and implementation of an augmented reality system using gaze interaction," *Multimedia Tools and Applications*, vol. 68, no. 2, pp. 265–280, 2014.

[12] P. Mistry, P. Maes, and L. Chang, "Wuw-wear ur world: a wearable gestural interface," in *Proceedings of CHI'09 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2009, pp. 4111–4116.

[13] A. Colaço, 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," in *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*. ACM, 2013, pp. 227–236.

[14] S. Khattak, B. Cowan, I. Chepurna, and A. Hogue, "A real-time reconstructed 3d environment augmented with virtual objects rendered with correct occlusion," in *Proceedings of IEEE Games Media Entertainment (GEM)*. IEEE, 2014, pp. 1–8.

[15] M. Tuceryan, Y. Genc, and N. Navab, "Single-point active alignment method (spaam) for optical see-through hmd calibration for augmented reality," *Presence: Teleoperators & Virtual Environments*, vol. 11, no. 3, pp. 259–276, 2002.

[16] A. Plopski, Y. Itoh, C. Nitschke, K. Kiyokawa, G. Klinker, and H. Takemura, "Corneal-imaging calibration for optical see-through head-mounted displays," *IEEE Transactions on Visualization and Computer Graphics*, vol. 21, no. 4, pp. 481–490, 2015.

[17] X. Xie, H. Liu, M. Edmonds, F. Gaol, S. Qi, Y. Zhu, B. Rothrock, and S.-C. Zhu, "Unsupervised learning of hierarchical models for hand-object interactions," in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 1–9.

[18] M. Edmonds, F. Gao, X. Xie, H. Liu, S. Qi, Y. Zhu, B. Rothrock, and S.-C. Zhu, "Feeling the force: Integrating force and pose for fluent discovery through imitation learning to open medicine bottles," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 3530–3537.

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

[20] K. Özacar, K. Takashima, and Y. Kitamura, "Direct 3d object manipulation on a collaborative stereoscopic display," in *Proceedings of the 1st Symposium on Spatial User Interaction*. ACM, 2013, pp. 69–72.

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

[22] T. Piumsomboon, A. Clark, A. Umakatsu, and M. Billinghurst, "Poster: Physically-based natural hand and tangible ar interaction for face-to-face collaboration on a tabletop," in *Proceedings of IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 2012, pp. 155–156.

[23] C. Kervegant, F. Raymond, D. Graeff, and J. Castet, "Touch hologram in mid-air," in *Proceedings of ACM SIGGRAPH 2017 Emerging Technologies*. ACM, 2017, p. 23.