



An empirical practice of design and evaluation of freehand interaction gestures in virtual reality

Ying Wu^{1,2} · Yigang Wang¹ · Xiaolong Lou^{1,3} · Mingwei Zhang¹

Received: 18 May 2023 / Revised: 25 September 2023 / Accepted: 30 October 2023 /
Published online: 8 November 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Nowadays, virtual reality (VR) technologies are increasingly used in various domains. Thus, there is an urgent need to sufficiently explore a natural approach to human–computer interaction that can be customized to various situations across different user groups. In this study, we explored hand gesture types that are more effective and natural for manipulating three-dimensional (3D) objects for application in freehand interaction within a VR environment. More specifically, gestures for 3D manipulation including directional navigation, rotation, scaling, and teleportation were tracked and recognized through a Leap Motion sensor equipped in front of a head-mounted VR device and on top of 3D glasses, respectively. To systematically validate the efficiency, usability, and reliability of the hand gestures, we designed a series of representative tasks in a specifically designed VR application and then recruited 40 participants to complete all required interaction tasks using the newly designed gestures and baseline hand controller. The results showed that the gestural interaction method was sufficiently effective for accomplishing most of the interactive tasks, indicating that this method could be a viable alternative to the conventional hand controller interaction method. Moreover, users preferred the new gestural interaction method more than the conventional one in terms of user experience and immersion. Based on these results, guidelines and strategies were discussed for developing freehand interaction techniques in general VR applications.

Keywords Freehand interaction · Hand gestures · Virtual reality · Motion-sensing interaction · Human–computer interaction evaluation

✉ Yigang Wang
yigang.wang@hdu.edu.cn

¹ School of Media and Design, Hangzhou Dianzi University, Hangzhou 310018, Zhejiang, China

² Jiaying Nanhu University, Jiaying 314001, Zhejiang, China

³ State Key Laboratory of Virtual Reality System and Technology, Beihang University, Beijing, China

1 Introduction

Nowadays, virtual reality (VR) technologies are widely used in various fields, such as medicine, education, entertainment, and art exhibition [1, 2]. The technical progress in VR, in terms of both hardware and software, allows highly immersive virtual experiences to be created for users. The most common interaction is through a handle. However, this has many limitations; for instance, it is cumbersome and unnatural to implement interaction using external input devices [3]. Currently, research on gesture-based interaction is expanding. Studies have shown that the sense of body ownership in VR can provide a more immersive experience [4], as basic body movements and freehand gestures are meaningful and expressive, and using hands to operate the model directly is a good means of human–computer interaction (HCI) [5].

To develop truly natural gesture interaction, there are multiple hardware-based approaches. Some modern commercial VR devices, e.g., Oculus Quest, provide limited gesture-based interaction by establishing a special physics-based model on the shape of the tracking hand. Media Pipe Hand, a gesture tracking system from Google, uses a camera and 2.5D technology to approximate 3D simulation [6]. However, there are still limitations given that gesture-recognition technology is immature, such as the finite and complex nature of gestures.

Moreover, the majority of the earlier studies investigating on the hand gestural interaction were based on a VR device of head-mounted display (HMD), but had little investigation on a large display-based VR application. Quite a few researchers had proven that information searching and interaction on large displays had qualitative advantages owing to the larger space and deeper immersion [7]. From this perspective, a more comprehensive investigation on both an HMD- and a large display- based VR application is quite necessary. The goal of our research is to explore the most natural and efficient freehand gestures that can be widely used in universal VR applications and tasks, across different VR devices; at the same time, to explore respective advantages of freehand gestural interaction in HMD- and large display- based VR environments.

As the gesture recognition method based on Leap Motion has high accuracy, real-time performance, and low cost [8], the interaction method proposed in this paper adopts a Leap Motion device. We designed and developed natural hand gestures to manipulate the model, including panning and rotating, scene navigation, teleportation, menu manipulation, grasping, and scaling. In addition, we used a collision detection algorithm to enhance the target manipulation task. We evaluated the actual effects of different hand gestures on manipulating virtual objects, and a user study (Section 5) was conducted to compare the task performance between the freehand interaction method and traditional hand controller-based method.

The three main contributions of this study are as follows:

- (1) We propose a convenient large-screen VR interactive system that can be adapted to a wide range of scenes.
- (2) A user-verified freehand gesture set is proposed, which can be generally applied to natural hand gestural interaction in various VR scenes.
- (3) By comparing to the controller-based interaction method, we systematically validate the practical advantages of the freehand gestural interaction method in terms of task completion efficiency, usability and reliability. It implies guidelines for developing natural user interaction applications in VR environment.

2 Related work

In terms of HCI, new interaction metaphors, designs, and tools have been implemented; various existing virtual motion technologies have been updated; and new technologies have been developed and studied [9].

Meanwhile, motion solutions based on classical input devices have lost their appeal because they break the illusion of interacting directly with the virtual world. Researchers have begun to investigate gesture-based VR content interaction by using contactless somatosensory devices, such as Microsoft's Kinect and Leap Motion [10]. VR HMDs have also achieved hardware advances by integrating more tracking solutions into consumer hardware, such as eye tracking integration or direct manual tracking [11, 12]. Natural user interfaces, such as freehand gestures in VR, are likely to become built-in standards. Therefore, future VR should require only an HMD and no additional hardware.

Over the past few years, many studies have been conducted on gesture interaction in VR. For instance, Tian et al. [13] developed some simple applications to use gestures as a “natural mouse” for tasks such as instruction and drawing so that users perform these tasks in a familiar manner. Wu et al. [14] elicited user-defined gestures for shopping tasks in an immersive VR environment. Masurovsky et al. [15] compared hand tracking and controller performance in grab-and-place tasks. Their studies indicate that gestural interaction can substantially improve users' operating experiences, as it allows users to directly control the information space in physical space with two hands. However, using free-hand gestures is still immature compared with the user's expectations [16]. In one recent research by Venkatakrishnan et al. [17], a host of design issues such as the input modality, hand triggering mechanisms and the interface geometry had been investigated, and they summarized guidelines for designing future mid-air freehand drawing and writing applications in VR.

Studies have extensively explored different tasks, such as pointing, reaching, moving, and rotating [18, 19]. Argelaguet et al. [20] conducted a survey of different 3D object selection techniques. Schafer et al. [21] studied the hand gestures of distant pinch and template grab for picking up virtual objects and found that there were no significant differences in performance and accuracy. Meanwhile, if the users want to select or grasp an object that is far away, they may need to move to the front of the object through walking or teleporting. Bowman et al. [22] evaluated the grabbing and manipulating techniques for remote objects in immersive virtual environments and discussed their characteristics and limitations. In addition, rotation is a fundamental input task. Su et al. [23] developed an improved device-based technique within handheld mobile augmented reality (AR) interfaces to solve the large-range 3D object rotation problem, as well as issues related to position and orientation deviation in manipulating 3D objects. To imitate interactions with physical objects as closely as possible, Mendes et al. [24] used 6 degrees-of-freedom (DOF) information from users' hands with direct manipulation, where the grabbed object follows the movement of the hand, and they implemented the technique as presented. Hands-free menu selection can greatly reduce the occupation of space and enrich operation content; previous works have explored different kinds of menus. Gerber and Bechmann [25] developed a technique called “rotating menu,” which places the menu on the user's wrist. Xia et al. [26] designed a gesture-based menu to assist users in learning the spatial layout of a virtual environment more efficiently. Although the approaches are effective in specific scenarios, they need to be optimized.

Travel is considered the most basic and important component of the VR experience, through which the user's viewpoint position can be changed and rotated. Various virtual

locomotion techniques have been developed and studied, aiming to offer a natural, user-friendly, and efficient way of travel in a virtual environment (VE) [27, 28]. Zhang et al. [29] proposed a motion control method using two-handed gestures: the left palm is used to control the movement forward or backward, while the right thumb controls the movement left or right. Caggianese et al. [30] compared gesture-, gaze-, and controller-based movement techniques in which participants had to move along predefined paths in a VE. However, there is still potential to explore possible ways to integrate various techniques into seamless and intuitive interaction. The applications of most basic interactions in VEs need to be optimized [31].

Designing and optimizing these technologies can be challenging, and the main problem is how to choose a gesture that can easily adapt to all users. In addition, fatigue can affect users, especially when repetitive actions are required to complete tasks. The increase of functions will also lead to the increase of cognitive load [32]. For example, when multiple tasks need to be performed, it is not ideal to use the commonly used tracking controller [33]; therefore, it is crucial to study the most appropriate interaction metaphor and its applicability in VR. In addition, several studies have experimentally evaluated how different virtual hand representations affect user performance [34–36]. Lougiakis et al. [37] discussed the effects of different virtual hand representations in terms of interactivity and the user's sense of embodiment when using controllers, and no significant differences were identified in the sense of agency. Ismail et al. [38] highlighted that enabling speech with gestures for interaction can improve the speed of task completion.

As presented, the exploration of these techniques is still limited. Most gesture studies focus on specific applications and are not systematic or universally applicable. If each device requires a different gesture, the user may need to be familiar with too many gestures. In addition, to the best of our knowledge, most of the gestures proposed and researched to date were in an HMD environment, and few were in a large-scale environment, where users benefit more from stereoscopic vision [39].

In this paper, we define a relatively complete set of gestures that can not only be used in an HMD but also in a large-screen environment, fully considering utility and reliability and minimizing the cognitive burden on users. We then present our experiment to evaluate the performance and usability in an uncontrolled VR application scenario.

3 Methodology: design of freehand gestures

The goal of this study was to develop a universal set of gestures that could replace the handle for most tasks. This research involved several steps. First, we gathered statistics and studied the requirements of gesture operation in the field of HCI. Then, we defined gestures and developed programs based on the principles of gesture design. Next, we conducted an in-lab experiment and analyzed all the proposed techniques and their task performance in the VE. The whole design process is shown in Fig. 1. In this section, these steps are described in detail.

3.1 Gesture design requirements

With the development of MR technology, we must consider which problems should be solved by gesture interaction in AR/VR. By referring to the interaction design guidelines

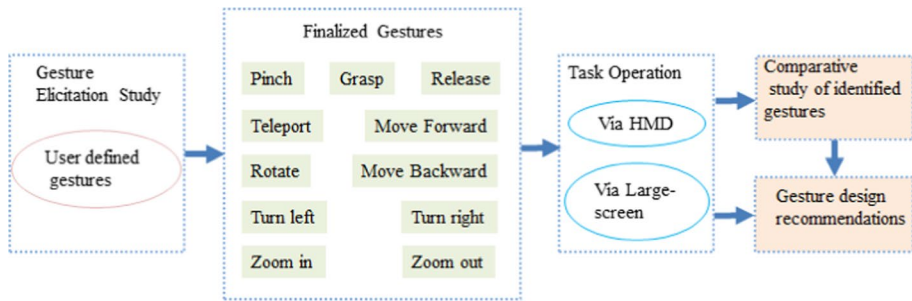


Fig. 1 Experimental design process

of Quest [40], the literature, and a user survey [41], we believe that gesture interaction needs to solve at least the following problems:

- (1) How to select a virtual object. Selecting a virtual object constitutes the essential basic operations, such as move and rotate. Distant pinch and template grab gestures are often used to select objects.
- (2) How to control and move an object. The ability to manipulate virtual objects is a key feature while interacting within VEs. It is still difficult to place a virtual object in the desired place with a high degree of accuracy.
- (3) How to rotate an object naturally and effectively. It would be interesting to implement 6DOF object rotation with a single hand.
- (4) How to scale a 3D object. Developing more convenient scaling manipulations for spatial operation.
- (5) VR locomotion. Allowing users to control their viewpoint motion in a three-dimensional environment is a crucial element in establishing a sense of immersion or presence. Teleportation and walking-based approaches are widely used for VR locomotion, while locomotion solutions based on classic input devices may break the illusion of interacting with the virtual world directly and may not be that attractive. Freehand techniques allowing a deeper immersion in the VE have begun to be more widely used. However, these techniques are challenging to design and optimize.
- (6) System functions (such as home button or menu): Through a menu, more interactive functions can be achieved but with the same effect as mouse operation.

3.2 Gesture design and connotations

Gesture design should follow sign language recognition design theory and principles [42]. Gesture recognition mainly consists of data acquisition, feature representation, and classification. Common feature extraction is divided into trajectory and hand shape features. Trajectory features are mainly used to collect information such as hand angle, acceleration, and position, which is followed by gesture feature extraction and gesture recognition. The principles of gesture design mainly include user aspect, interactive process, and interactive system. Through the research on gesture theory and the principles of gesture design, gestures were defined.

- (1) **Gesture design principles.** The principles of gesture design are mainly reflected in the user, interactive process, interactive system, and timely feedback.
 - (i) **Users.** The ultimate purpose of gesture design is to facilitate user operation. Therefore, the design of gestures needs to start from the perspective of user behavior research, conform to users' common habits, and ensure that the operation is simple, convenient, intuitive, and natural so that users can experience enjoyment in gesture operation.
 - (ii) **Interaction process.** The gesture design interaction must be smooth. Use of the dominant hand to complete a task should be prioritized, considering the historical and cultural connotations of gestures in specific regions to avoid conflicts.
 - (iii) **Interactive system.** We must simplify gesture design and emphasize systematic gesture design and unity of the design style, with as little functional crossover as possible.
 - (iv) **Timely feedback.** In the process of user operation, the computer system should give a variety of feedback prompts for the user's behavior in a timely manner so that the user can operate conveniently.
- (2) **Gesture definition.** According to common task requirements and gesture design principles, a specifically designed survey was first conducted. We invited ten researchers, teachers, and students in the field of VR, showed them the task, and asked them to perform the gestures they would use. We captured and recorded the process. Based on their proposed gestures and those already in the literature, we discussed and voted for universally applicable gestures and formed a gesture set. The gesture definitions are shown in Fig. 2, and their corresponding operation descriptions are given in Table 1. To test and compare the gestures we defined, all functions were also implemented in a controller. Most of our gestures are defined to use the right hand, except zoom and teleport. The zoom gesture requires two hands grasping the object to move closer or farther away. If the user's right hand is pinching or grasping an object without releasing it, then they can move it to the target location. The user's left hand performing the pinch gesture represents teleporting. The user needs to place the pinch gesture on top of the place they wish to reach. The visual feedback of the traveling direction is then displayed as a ray advancing from the center of the pinch down onto the VE terrain, where a placeholder indicates the position to reach. When the fingers are loosened, the virtual avatar arrives. This technique gives the user the ability to move in all directions, even walking backward and sideways, without rotating the head.


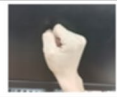





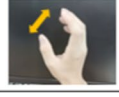
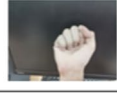
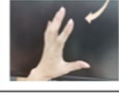

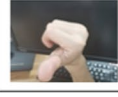


Pinch	Grasp	Teleport	Zoom in	Upward	Turn left	Controller
						
Release	Rotate		Zoom out	forward	Turn right	Controller
						

Fig. 2 Eleven basic gestures are defined, the same functions are implemented on the controller

Table 1 The corresponding relationship between gesture and operation type

Functions	Gestures for manipulations
Pinch	Point your finger at the target object then pinch
Release	Loosen the fingers
Grasp	When the virtual hand touches the virtual object, make a fist to grasp the object
Rotate	Grab the object and rotate the hand
Teleport	The left hand intersects the ground with making the pinch gesture
Zoom in	Grasp the object with two hands and make a fist to move closer
Zoom out	Grasp the object with two hands and make a fist to move apart
Forward	Thumbs up and forward and other fingers hold
Backward	Thumbs up to back and other fingers hold
Turn left	Thumbs up to the left and other fingers hold
Turn right	Thumbs up to the right and other fingers hold

3.3 Capturing static gestures

Leap Motion encapsulates a number of APIs that can be called on to retrieve relevant data. However, some simple function calls cannot guarantee the accuracy and realism of manipulating objects in the VE. We propose a gesture capturing and freehand interaction system that is implemented by detecting the finger state and palm direction in real time, using grasp decision rules based on collision detection and virtual gesture manipulation. This system can recognize gestures effectively.

The system monitors the state of the finger and distinguishes between two states: stretched or bent. Thus, the state of each finger forms a clear descriptor of the gesture; for example, a fist is recognized when all fingers are bent. Pointing gestures are detected when the index finger is stretched and the other fingers are curled. Using finger states as descriptors can detect certain hand postures, but not a variety of more complex gestures.

For example, the pinch gesture is judged by the pinching of the thumb and index finger, but the pinching surface also has an orientation. Thus, finger state descriptors alone are not sufficient. Hand orientation provides information about which way the hand is facing, as well as implicit information about which way each finger is pointing. To this end, we also use gesture direction as a descriptor for gesture recognition systems and find it very suitable for this purpose. Because the original orientation value of the hand is too limited, we activate a gesture by adding a tolerance value. By combining these two descriptors, the system recognizes multiple static gestures and is able to perform gesture capture events while the user is running, and quickly prototype any combination of gestures to control motion. Leap Motion has a wide range of recognition. When worn on the head, it can also be detected whether the wearer is standing up and the hands are hanging down at the sides, which may cause malfunction. To solve this problem, we set a range beyond which the operation remains untriggered even if the hand is detected.

3.4 Technical implementation

We implemented the interaction system based on the gesture design, which includes the following functions:

- (1) Recognition of pinch: Pinch gesture is used in many gesture-based HCI devices [43]. Because ray-casting techniques are convenient to reach an object, we set the condition that if the distance between the thumb and index finger is less than 1 cm when they make a kneading posture, a ray will be emitted from the center point between the thumb and index finger tips. The direction formed by the arm to the center point is the vector of the ray, and the point where the ray intersects the virtual object is the point to be selected. When the distance between the tip of the thumb and tip of the index finger is less than 0.1, the pinching gesture is considered to be triggered. This allows objects or menus to be selected, especially distant objects. Here, the dummy hand is a ball-and-stick model.
- (2) Grasp: When people want to operate on an object, it is natural to reach out and grasp the object and then carry out the corresponding operation, especially for close objects. Our approach is closer to commonly used metaphors; the difference in our design is that when the hand grabs the object, a menu pops up. We can not only grasp the object and rotate it naturally but can also navigate through the menu to operate it precisely.
- (3) Translation: When the user's hand is judged to be in the state of pinching or grasping, the user can move the object by moving their hand. By calculating the displacement vector of the hand, the position of the virtual object is synchronized.
- (4) Rotation: When pinching or grasping a virtual object, the user rotates the hand to make the object rotate. The key to the rotation gesture is to calculate the rotation angle α between adjacent frames. Because the accuracy of Leap Motion differs for different gestures, experimental comparison is needed to find the most suitable data to represent the direction of the hand. In this system, the coordinate of palm position is selected as the characteristic data of hand position, which is used to calculate the vector of hand displacement.
- (5) Zoom in and out: The object is held with two hands. When the hands are slowly approaching, it means shrinking. When the hands are far away, it means enlarging. When both palms are extended, the current state remains.
- (6) Teleport: Pointing arrows are used to provide visual feedback for the user to select the destination of the teleport. When the left hand is pinched, the center point of the fingertips of the thumb and index finger is connected with the shoulder to emit a ray, and the point where the ray intersects the ground is the position where the teleport arrives. The direction can be selected by moving the pinching hand. When the finger is released, the teleportation is triggered and the user's avatar is teleported to that position. In pilot tests, when players were interacting with buttons or looking around, they would accidentally make gestures that triggered teleportation commands. We addressed this issue by making it impossible to teleport when near interactive objects.
- (7) Steering-based walking: When users move around in VR, as is often discussed in the literature, a natural and effective way of controlling locomotion is still a general problem to be solved. In much of the existing literature [30], the palm is used as a navigation gesture, which requires the cooperation of two hands and is obviously not convenient. According to a study on human kinematics [44], it is comparatively easier to control the direction of the thumb muscle group. Different gesture types were investigated and

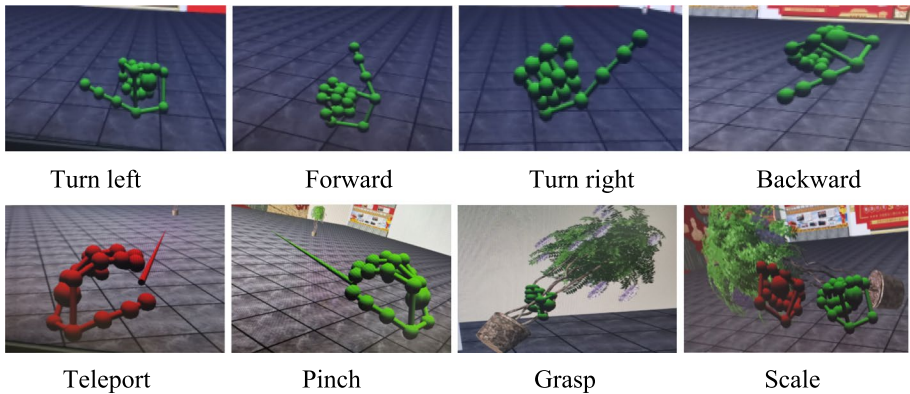


Fig. 3 Different gestures identified by Leap Motion are displayed in the scene

compared, and we finally decided to use the “thumbs up” gesture, using the thumb direction to determine the direction of movement. The locomotion is performed at a constant speed. Because the original orientation value of the hand was too restrictive, a tolerance value was added to allow the system to activate the gesture. The gesture operation is shown in Fig. 3.

The above functions were implemented in HMD mode. To fully test the usage experience of gestures, we developed a system based on a large operating screen, as shown in Fig. 4. This approach does not require an HMD and uses the device shown in Fig. 6b. We attached the Leap Motion sensor on the 3D glasses and adjusted the fixed angle between the sensor and the glasses through a self-made 3D printing bracket. At the same time, a gyroscope is used to locate the device, and the data are sent to a computer through a mobile phone connection for processing. Then, the running results are projected onto the screen.

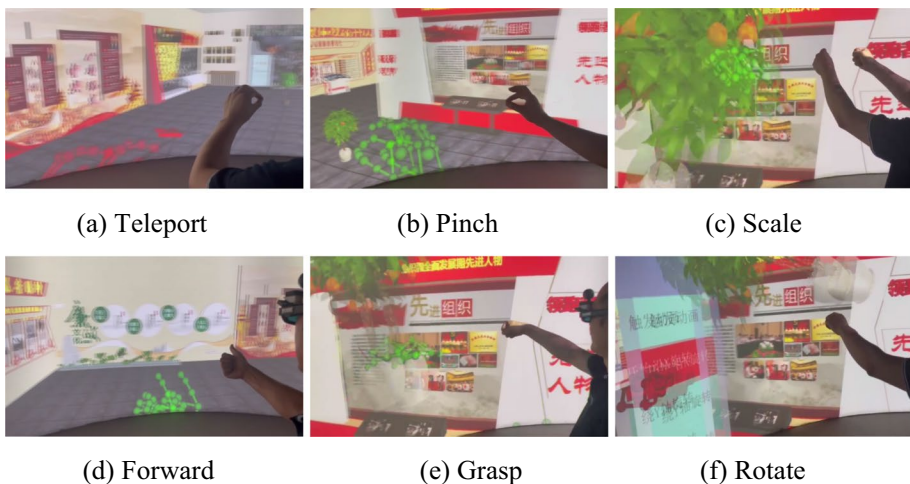


Fig. 4 Gesture operations in large-screen environment

In the large-screen environment, when the avatar moves, it is impossible to change the direction by turning the head as is the case when wearing an HMD. Therefore, we designed a direction arrow to appear at the intersection point between the virtual hand and ground during the teleportation operation. The direction can be adjusted by rotating the wrist, and when the thumb and index finger are released, a teleport is performed.

To enable gesture manipulation to support complex interactions better, we implemented two menus. Figure 5a shows that once the user grasps the object, a menu will pop up, from which they can choose to let the virtual object rotate around the X, Y, or Z axis through the pinch gesture. Moreover, they can rotate it just by rotating their wrist. Figure 5b shows the functional menu to extend the interactive function. As an example, we mainly implemented the following typical functions: add objects to the scene, change wallpaper or floor, edit objects in the scene, and copy and delete virtual objects. As there is no Home button for both hands, we designed a way for the user to summon a menu in the application experience. Users can select various menus, and when the thumb and forefinger are pinched together, the corresponding buttons will be executed. Figure 5b shows the user selecting the material change button to change the floor material to carpet. In Fig. 5c, the user selects the “add object” button, selects a sofa from the object library to add to the scene, and then scales and positions the sofa. In addition, all of these features were also developed and implemented with the controller. In a later experiment, we invited users to experience both and provide feedback.

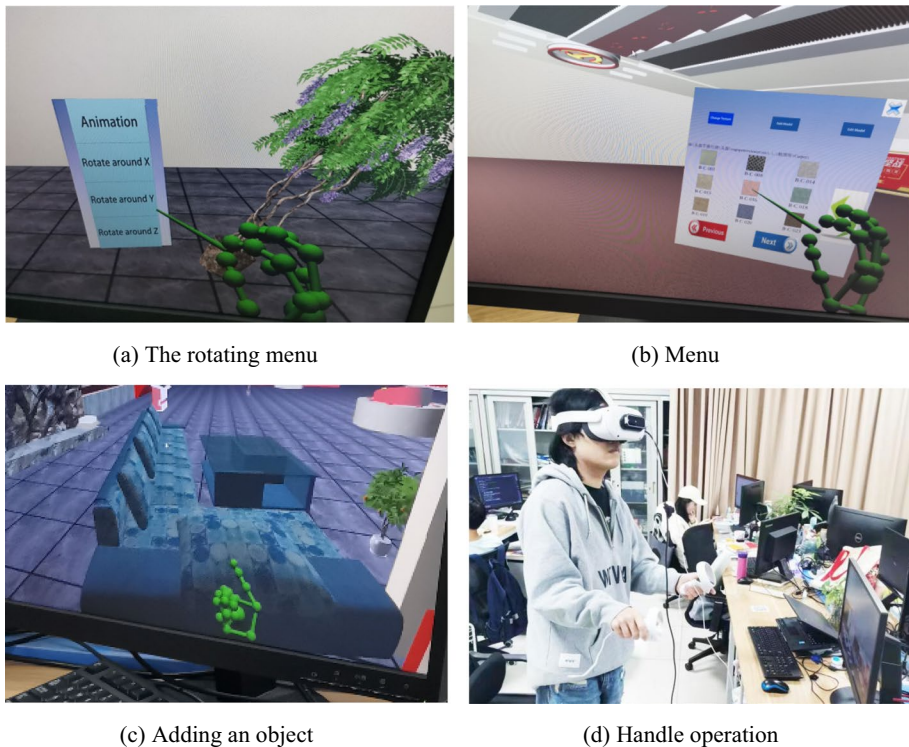


Fig. 5 Operation via menu and handle

3.5 Controller function development

The functions of the controller were defined as follows. The left joystick controls forward and backward direction, and pressing the left trigger button continues the forward movement. The right controller emits a ray in the scene. If the ray intersects an object, the object can be selected by pressing the trigger button of the right controller. If the ray is intersecting the ground, the trigger button performs the teleportation action. While selecting the object, press X or Y on the left controller can scale the object. When the object is selected, the wrist can be rotated to realize rotation of the object, and releasing the trigger button drops the object.

4 User study

To comprehensively evaluate the proposed strategy for using freehand interaction in VR and compare it with controller manipulation, we designed an empirical experiment to compare the performance of all proposed techniques performing the same task in the same VE. In addition, these methods were evaluated, considering both quantitative and qualitative measures. In this section, the experiments are described and the results are discussed.

4.1 Objectives and hypotheses

Our objective was to measure different freehand motion techniques by evaluating efficiency and effectiveness and associated user satisfaction [45], and compare them with baseline techniques using controllers. Several studies have evaluated the efficiency and effectiveness of controller operation. According to ISO's generalized definition of usability [46], our efficiency measures task completion times, while effectiveness measures error rates and solution quality. The above objective indicators, efficiency and effectiveness, were used as performance indicators, that is, to measure the results of user interactions. In addition, to complete the usability and reliability analysis, we considered subjective indicators to explore emotional, usability, physical, and cognitive needs and preferences. Based on previous work and our experience, we built the following three hypotheses for experiment:

- H1: Hand-direct operation without relying on controllers has no effect on interaction performance.
- H2: Freehand gestural interaction has a higher user immersion than the controller-based interaction, and the former results in lower perceived exertion and task difficulty.
- H3: Freehand gestural interaction has advantages over controller-based interaction in terms of user navigation efficiency, naturalness, and convenience, which leads to its preference over freehand interaction.

We assessed the opinions of the subjects participating in the study, particularly user perceived exertion (UPE), user perceived task completion difficulty (UPTD) and user perceived system usability (UPSU), through a subjective assessment scale. These three subjective assessment scales were specifically designed based on the Borg15 Scale [47], NASA-TLX [48], and Brooke System Usability Scale (SUS) [49], respectively. The UPE assessment based on the Borg15 Scale requires no additional equipment but asks the

participant to assess their perceived rate of physical exertion from light to strenuous in the specific physical activity. It classifies physical exertion in a range between 6.0 and 20.0, where a higher quantitative value represents a stronger perception of the physical exertion. The UPTD assessment method was derived from the NASA TLX scale developed by Hart. It was initially developed for measuring the operators' perceived task load at NASA, but nowadays, it is widely used and regarded as the gold standard for measuring subjective workload across a wide range of industries. It contains 6 measurements: (1) mental demand, (2) physical demand, (3) temporal demand, (4) overall performance, (5) effort, and (6) frustration, which are specifically defined to assess the mental and perceptual demand level, physical efforts needed, time pressure perceived, user's overall satisfaction with the task performance, mental and physical difficulties encountered, and stress and irritations experienced in completing the required task. The UPSU assessment based on the Brooke SUS provides a quick technical assessment and useful score that simplifies comparisons with other methods. The overall goal is to determine whether the free-hand method is better in terms of efficiency, usability, and reliability than the controller-based method and to determine which gestures are more natural to the majority of users.

4.2 Participants

For the study, we recruited 40 unpaid volunteers (23 Male, 17 Female). The average age was 24.6 years ($SD=3.78$), and all participants were right-handed. Among all the participants, 22 were undergraduate and graduate students of the department of digital media and the other 18 were undergraduate students from the department of art and design from a local university. All of the participants rated themselves as having good knowledge of software and computers, with fifteen participants mentioning that they had used a VR headset before and the others having no previous hand tracking experience.

4.3 Apparatus

The computer used for the user study was a VR Ready computer running 64-bit Windows 10, Intel Core i7 8 GB, and NVIDIA RTX 3070. Oculus with a Leap Motion device was used in this experiment, as shown in Fig. 6.



(a) Leap Motion with HMD



(b) Leap Motion with 3D glasses

Fig. 6 Device used in the experiment: the left is Leap Motion mount on the Oculus Quest, the right is attached to the glasses with a locator

Oculus Quest 2 is an all-in-one VR machine that Facebook officially released in 2020. It has a pure white body design and weighs 503 g. It uses an LCD screen, runs on a Qualcomm Snapdragon XR2 processor, and comes with 6 GB of running memory. With monocular 1832×1920 and binocular 3664×1920 display resolutions, it supports 72 and 90 Hz refresh rates. The Oculus Quest 2 comes with a pair of 6DOF controllers and supports hand tracking.

We use the Leap sensor to achieve freehand interactive motion. Equipped with two gray scale camera sensors and three infrared LED light sources, the Leap Motion has a field of view of up to 180 degrees. It can detect and track the user's hand and fingers with high accuracy and tracking frequency and return the traced hand model for developers to use. Although Leap Motion's redeveloped operation is more playable, the definition of various gestures and the detection of complex movements require further research and development. The Leap Motion controller can be installed on any HMD device, including Oculus Quest.

This study used Open Scene Graph (OSG) for comprehensive development. OSG is a powerful and efficient 3D graphics rendering engine, which is completely based on C++ and Open GL graphics underlying development languages and can be applied to different development platforms, including Microsoft Windows system, Mac OSX, and Linux.

4.4 Experimental tasks and design

The virtual environment in this experiment, which was developed by OSG, was a large room scene having a few furniture items such as sofa, bed, lamp and vases. This virtual environment was designed for personalized customization of interior arrangement: through virtual interaction, users can design their own home and visualize the decoration effect in real time. The participants were required to navigate through a specifically designed home decoration virtual scene and complete various tasks.

These tasks were of five types:

- (1) Walking-based navigation task: the participant followed a designated path to visit the entire scene.
- (2) Rotation task: the participant teleported to a specific position, rotated the flowerpot to get a whole view of the flowerpot in 6DoF, and checked and reported the numbers marked on the back.
- (3) Translation task: the participant moved the flowerpot to the other side of the booth.
- (4) Scaling task: the participant zoomed the flowerpot in and out to give it the same height as the booth.
- (5) Teleportation task: By teleporting to the front of the round table, the right hand does the pinch gesture and intersects the ground to bring up the menu. The participant clicked the Add Object button in the menu, added a sofa to the scene, shrunk it to the appropriate size, and moved it to the specified position and orientation.

Participants were initially shown all five tasks, so they were not required to remember the tasks and their order. After completing one task, they were reminded of the next, and once the participants had completed all tasks with one technique, they answered a questionnaire about the experience. The questionnaire included closed and open-ended questions. A 5-point Likert scale was used, from very dissatisfied (1) to very satisfied (5).

We used a 2×5 within-participant design to conduct the experiment, with 2 interaction methods and 5 tasks. The benefits and limitations of the freehand gestural interaction method were measured quantitatively and qualitatively. The former included task performance, such as execution time, accuracy, frequency of motion interruptions, and hand tracking losses, while the latter included UPE and UPTD on two interaction methods. In summary, the independent variables and dependent measurements in the experiment are as follows:

Independent variables:

- Interaction method (2): freehand gestures, controller;
- Experimental task (5): translation, walking-based travel, teleportation, rotation, scaling;

Dependent measurements:

- Task execution efficiency (measured in task completion time);
- Task completion accuracy;
- Motion interruption frequency;
- Hand tracking losing frequency;
- UPE;
- UPTD;

4.5 Procedure

After entering the testing laboratory, participants first read and signed the consent form. Then, the supervisor explained the overview of the experiment and how the hardware is used to each individual. Furthermore, a brief introduction about the task was given, and the range of the hand tracker was explained. Participants were told they could stop the experiment at any time or take plenty of rest, especially if there were physical or mental problems at the time of the experiment.

The participants were asked to perform all tasks using hand gestures or the controller. To prevent learning and fatigue effects, each volunteer used a different ordering of the locomotion techniques, based on a Latin square. Subjects were allowed to try each method for 3 min before the actual task (as many had no experience with either VR or hand tracking). When all techniques had been tried, a thorough survey was completed by the users.

A total of 1600 interactive task tests were completed, with $40 \text{ users} \times 2 \text{ input devices} \times 5 \text{ tasks} \times 2 \times \text{Times} \times 2 \text{ rounds} = 1600$ interactive tasks. Before starting the test, the user was first introduced to the interaction map for each input device. After completing all individual tasks, users needed to give their preferences and reasons for each input device. For complex integrated tasks, there was no time limit, and the main measurement was the smoothness of the operation and overall completion. At the same time, subjective evaluations, such as user interactions and preferences, were recorded.

5 Analyses and results

Each participant completed two blocks; one used the freehand gestural interaction technique and the other used the controller to complete the tasks. After the experiment, a total of 80 task log files were collected, which recorded the task completion time, number of motion interruptions, hand tracking losses, and error frequency of operations.

These data were measured to be normally distributed; thus, they were analyzed by repeated-measures ANOVA and two-tailed dependent T-tests for paired comparisons.

Besides the task log files, there were also 120 subjective assessment scales collected following the experiment. For such subjective assessment scale statistics, they were analyzed by non-parametric multi-sample Kruskal–Wallis test and two-sample Mann–Whitney U test. All results reported below were significant at least at the $p < 0.001$ level.

5.1 Task completion time

The mean task execution time over the entire experiment was 106.95 s. There was a main effect of block ($F(1, 19) = 17.26, p < 0.001$). In completing the required task by the freehand interaction technique, the task execution time had minimal difference compared to that in the hand controller interaction ($M(\text{freehand}) = 105.31 \text{ s}, M(\text{hand controller}) = 108.58 \text{ s}, p = 0.724$). The average completion time for each task and each input device was compared, and the results are shown in Fig. 7. In translation tasks, the mean task time spent by the controller was 12.79 s, while that by freehand gestures was 11.73 s, which is an insignificant difference; in the tasks of teleportation and walking-based navigation, the two interaction methods resulted in similar efficiency in the task completion time without significant difference. However, in the rotation tasks, the new interaction by freehand gestures generated a significantly shorter task time than the controller-based interaction ($F(1, 19) = 8.12, p < 0.001$). In the scaling tasks, the freehand gestures were found to be faster than the controller-based interaction ($M(\text{freehand}) = 7.89 \text{ s}, M(\text{freehand}) = 8.95 \text{ s}, F(1, 19) = 10.31, p < 0.001$).

5.2 Loss frequency of hand tracking

During the system running, log files automatically record the process information. If the gesture node is not tracked for more than 2 s during the gesture operation, a trace loss is recorded. In this study, the count of the hand tracking loss was aimed to evaluate the stability of gesture tracking. We ensured that the gestures all use one-handed methods, except the rotation gesture that requires both hands. A lower tracking loss indicates better overall usability of the system [50]. According to statistics, the average number of one-handed tracking loss was 8 times and two-handed tracking was 10 times. Hand tracking failure

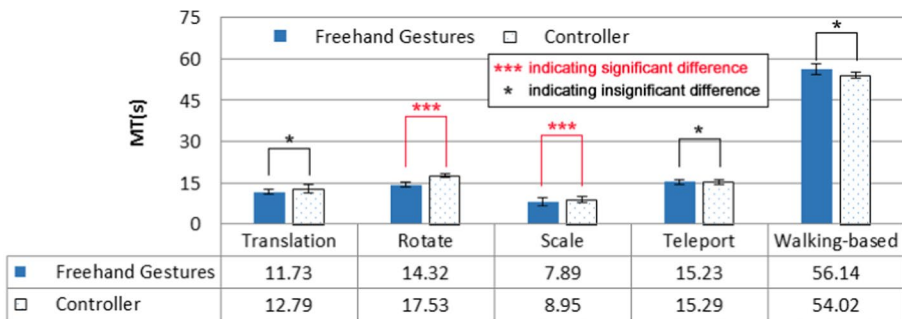


Fig. 7 The comparisons of task execution time in different tasks between freehand gestures- and controller-based interaction methods

happens in two cases: (1) the participant moved the hand suddenly at a high speed, or (2) the hand was moved outside the Leap Motion's sensing field.

5.3 Gesture recognition accuracy

We calculated statistics on the accuracy of gesture recognition. The recognition rate of mobile gestures was 100%. The recognition accuracy of selection, magnification, rotation, and teleportation was also high. The recognition accuracy of the gestures of “backward,” “turn left,” and “turn right” was relatively lower than others. When making the selection gesture, the side pinched by the thumb and index finger should be perpendicular to the line of sight. During the experiment, it was found that if the posture was standard akin to that of experienced VR players, the accuracy of this gesture was 100%, while users who had no VR experience would have inaccurate recognition when performing this operation.

Table 2 shows the recognition accuracy statistics of eleven types of gestures. The probability of a gesture being recognized once is very high, and generally speaking, it is very accurate.

We found that all gestures had a high recognition accuracy of $\geq 90\%$, except for the “backward” gesture. The main reason is that the backward gesture requires the user to make a fist while pointing their thumb towards themselves. This action is easy to understand, but if users put their hands slightly low, it led to movement beyond the detection range of Leap Motion or misidentification, resulting in lower accuracy.

5.4 User Perceived Exertion (UPE)

As we interpreted in Section 4.1, the participants were required to complete three scales to assess each of UPE, UPTD, and UPSU. In the UPE assessment scale, UPE for the freehand gestural and controller-based interaction techniques were rated 6.0 and 20.0, respectively. Figure 8 shows a statistical graph of the UPE comparison between the two interaction techniques. It can be seen that in the freehand gestural interaction, the UPE was lower than that in the baseline controller-based interaction; the Mann–Whitney U test on the UPE proved that this difference is statistically significant. There was also a significant interaction effect of *Gender* \times *Technique* on the UPE results ($\chi^2_2 = 11.98, p < 0.001$), as shown in the right graph of Fig. 8. For the male participants, using the freehand gestural interaction or controller technique, there was little difference in the UPE, but for the female participants, the freehand gestural interaction technique resulted in an obviously lower UPE than the controller interaction technique. Table 3 presents a summary of the statistical analyses of the UPE result.

5.5 User Perceived Task Difficulty (UPTD)

Given the 40 UPTD assessment scales, UPTD in different genders and techniques was also compared. As we described previously, the UPTDs consisted of 6 measurements derived from the NASA TLX scale and 2 other questions. The 6 measurements were rated on a 7-point Likert scale, with 1 representing “strongly disagree” and 7 representing “absolutely agree,” i.e., a higher rating represents a higher degree of agreement on the statement of the specific measurement. As shown in Table 4, a Kruskal–Wallis test of *Gender* (2) \times *Technique* (2) proved that the male and female participants had significantly different UPTDs

Table 2 Gesture recognition accuracy

	Pinch	Grasp	Move	Rotate	Zoom out	Zoom in	Teleport	Upward	Backward	Turn left	Turn right
Test number	400	300	200	200	200	200	200	400	200	200	200
Accuracy	97.8%	96.5%	100%	98.9%	99.1%	99.1%	95.2%	98.3%	88.7%	90.1%	91.2%

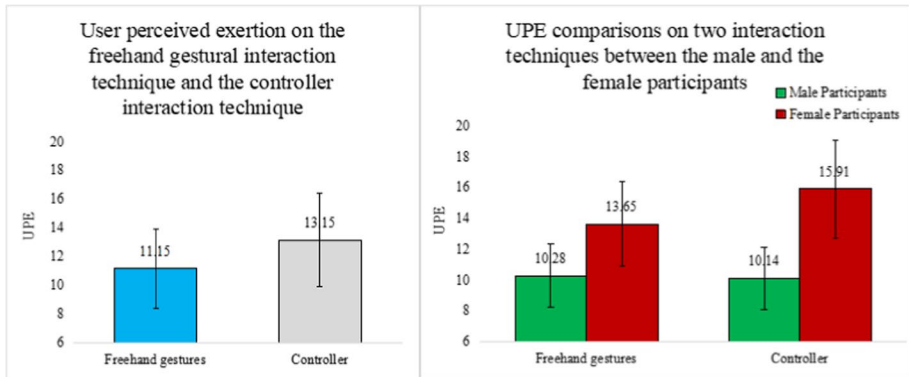


Fig. 8 UPE comparisons between two techniques across two genders

for the two interaction techniques ($U=38.97$, $Z = -4.106$, $p < 0.001$); more specifically, the female participants had a higher overall UPTD than the male participants. In addition, there was a more obvious difference in UPTD between the freehand gestural and baseline controller-based interaction technique ($U=99.20$, $Z = -7.920$, $p < 0.001$). As shown in Fig. 9, in separate measurements, the newly proposed freehand gestural interaction technique was assessed to have a lower mental demand than the controller interaction; it also had a lower perceived physical demand than the controller interaction. In the perceived performance, the former also led to a lower dissatisfaction with the task performance than the latter; in general, the freehand gestural interaction was assessed to experience less frustration than the controller interaction.

5.6 Summary of findings

In this study, a specifically designed empirical experiment was conducted to evaluate the effectiveness and benefits of the newly proposed freehand interaction technique and its hand gestures. It was found that freehand operations without relying on controllers had no reduction in interaction task performance against the traditional controller-based approach. Therefore, the first hypothesis (H1) is completely supported. It was also found that freehand gestural interaction had a higher user immersion than controller-based interaction, and the former resulted in lower perceived exertion and task difficulty, which confirmed the second hypothesis (H2). In the debriefing interviews, users commented that freehand gestural interaction had advantages of efficiency, naturalness, and convenience against the controller-based interaction, and thus, the freehand gestures were preferred by users, which supports the third hypothesis (H3).

Specifically, there were five main findings of the experiment:

- (1) The Leap Motion-based freehand gestural interaction technique was verified to be sufficiently robust for application in a VR interactive environment.
- (2) Compared to the more conventional hand controller interaction technique, the newly developed freehand gestural interaction technique resulted in a similar task performance in terms of task executive time and user perceived satisfaction.

Table 3 Statistical analyses and significance reports on the UPE

Independent variable	Mean UPE	Median	Min-Max	Sig.	Post-hoc comparisons
Gender	(1) Male (N=24)	10	8-14	M-W test, U = 39.32, Z = -4.270, $p < 0.001$	
	(2) Female (N=16)	14	10-16		
Technique	(1) Freehand gestures	12	8-16	M-W test, U = 22.87, Z = -3.600, $p < 0.001$	
	(2) Controller	14	10-18		
Gender× Technique	(1) Male × Freehand gestures	10.28	8-14	K-W test, $\chi^2_2 = 11.98$, $p < 0.001$	
	(2) Male × Controller	10.14	8-14		
	(3) Female × Freehand gestures	13.65	10-16		
	(4) Femal × Controller	15.91	10-16		

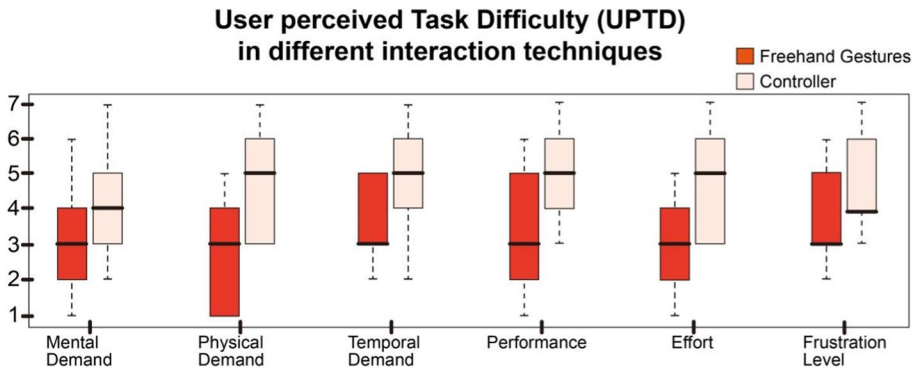


Fig. 9 UPTD comparisons between two techniques

- (3) In specific interactive tasks or operations, such as object rotations and zooming-in and zooming-out, the freehand gestural interaction technique was found to have a higher operating efficiency than the baseline technique.
- (4) In the accuracy evaluation of the freehand interaction technique, the 8 specifically defined gestures had a generally satisfying recognition and tracking accuracy. The gestures of “backward” and “turn left” had a relatively lower recognition accuracy in comparison to the other 6 gestures.
- (5) In terms of UPE and UPTD, the naturally performed freehand interaction was perceived to generate less physical and mental load than the conventional controller interaction; the former was also assessed to be easier to use and become accustomed to, which resulted in a reduced interaction difficulty than the latter and gained more user preferences.
- (6) Based on the above findings (1)–(4), the first hypothesis (H1) was supported; based on the findings of (5), the second and the third hypotheses (H2, H3) were also supported. Our findings indicate that the objective and subjective measures are relatively consistent with the proposed hypotheses, which comprehensively reflect the users’ behavioral and psychological performance during task completion.

6 Discussion

6.1 Specificity and advantages

In the experiment, we compared the proposed gesture set with a virtual handheld controller, which is commonly used in commercial immersive VR systems. The experimental results indicate that user-defined gestures allow users to interact with the VR environment easily and intuitively as well as offer improved user experience and user satisfaction. The findings during our quantitative and qualitative evaluation show promising results for all presented gesture interactions. The gestures designed and implemented using the proposed system scored high in the SUS, indicating a general good usability for all techniques. However, we did not find a significant difference in the proposed techniques in terms of effectiveness. Hence, hypothesis H1 can be partially accepted; hand tracking loss is minimal. The recognition rate of gestures reached over 88%. Further examination shows that in the

Table 4 Statistical analyses and significance reports on the UPTD

Independent variable	Mean score of the 6 measurements						Mean:Total score (S)	Total score comparison and statistical significance
	Mental demand	Physical demand	Temporal demand	Overall performance	Effort	Frustration		
Gender								
(1) Male (N=24)	3.19	2.78	3.05	3.19	3.08	2.78	18.07	M-W test, U=38.97, Z=-4.106, p<0.001
(2) Female (N=16)	4.08	3.55	3.90	4.08	3.93	3.55	23.09	
Technique								
(3) Freehand gesture	3.33	2.58	3.37	3.13	2.89	3.29	18.60	M-W test, U=99.20, Z=-7.920, p<0.001;
(4) Controller	4.38	4.68	4.81	5.23	4.92	4.12	28.14	

freehand gestural interaction, the UPE was lower than that in the baseline controller-based interaction, and a Mann–Whitney U test on the UPE proved that the above difference was statistically significant. For the male participants, using the freehand gestural or controller technique, there was little difference in the UPE, but for the female participants, the freehand gestural interaction technique resulted in an obviously lower UPE, which confirms hypothesis H2.

In addition, there was a more obvious difference in UPTD between the freehand gestural and baseline controller-based interaction technique. In separate measurements, the newly proposed freehand gestural interaction technique was assessed to have a lower mental demand than the controller interaction; it also had a lower perceived physical demand than the controller interaction; in the perceived performance, the former also led to a lower dissatisfaction in the task performance; in general, the freehand gestural interaction was assessed to experience less frustration than the controller interaction because of its intuitiveness and familiarity. We also noticed that the freehand gestures outperform the controller interaction in all criteria of the UPTD measurement. To confirm the accuracy of the results, we conducted interviews with the participants, and they reported that freehand interaction led to a better experience, particularly in scenes like museums, home furnishing, and production workshops, where gesture interaction can bring a more natural and interesting experience. It was generally accepted that the gestures are easier to remember and use than the controller-based interaction.

It is worth noting that most of the participants recruited in our study are students from different majors, and half of them had little experience with VR. The age of the participants in the experiment ranged from nineteen to thirty-eight; thus, results in the present study could represent the majority of the population who are potential users in real society. Through user interviews and questionnaires, we learned that some users have VR experience. It was found in the experiment that these users have relatively high proficiency and task completion rate. To reduce the difference caused by familiarity rather than the gesture itself, more time for practice is needed in future studies before the experiment and to take these factors into account when designing the experiment. Furthermore, among these participants, females had a higher overall UPTD than males, and the females had to compensate for this situation by exerting more effort to maintain task performance. We speculate that a possible reason is that a larger percentage of male users had video game experience than female users, so the former were more familiar with the VR tasks, which is beneficial for lowering their perceived task difficulty. This finding appears to suggest further research efforts may be required to examine the underlying reasons. In future research, to mitigate differences between genders in perceived difficulty and effort, different tasks can be assigned according to gender.

The contributions of the study contents, experimental findings, and implications are manifold. *From a technical development and application perspective*, this study introduced a new set of freehand gestures that can be prevalently used in 3D interactive environments, e.g., motion-sensing computing and VR interactive applications. Through an empirically comparative experiment, the new freehand gestural interaction technique based on the 8 basic operational gestures was verified to be valid and effective in an actual application. This freehand gestural interaction technique was designed and developed based on off-the-shelf technologies of a Leap Motion hand tracking sensor and Oculus Quest VR HMD, indicating that the proposed interaction technique can benefit the majority of VR applications. *From an HCI technique evaluation and ergonomic measurement perspective*, this study presented an example of user interaction evaluation in specific tasks in a VR application, in terms of quantitative measurements. In particular, the study proved that the

naturally performed freehand gestural interaction technique resulted in a satisfactory task performance, which was similar to that of the conventional hand controller interaction. In some operations, such as object rotation and zooming-in and zooming-out, the hand gestural interaction technique was found to be superior to the controller interaction technique. Besides the quantitative and qualitative task performance, subjective evaluations of UPE and UPTD, based on the NASA-TLX and Borg15 scales, respectively, were conducted, and the advantages of the hand gestural interaction technique were further verified. The results from the user study suggest that users tend to focus on tasks more when using direct manual input than using controller. Hand tracking gestures appear to have a higher level of naturalness and perceptual realism.

Can the observed differences be attributed to implementation choices? In order to eliminate the impact of setup difference on the freehand gestural interaction performance, we used two methods: large screen and HMD. As stated previously, for large-screen interaction, the screen is fixed, and the picture displayed by the helmet changes with the rotation of the head, always facing the front of the viewpoint in the helmet. Therefore, helmet-based gesture design may not be fully applicable to large-screen environments. For example, when we design teleportation, in the helmet environment, the point where the ray emitted by the gesture intersects with the ground is the point to be reached, and the line of sight is always facing forward. In the large-screen environment, we also need to make a direction arrow to choose the direction after we reach the destination point. As a result, the diverse nature of interactions has produced many research gaps and open areas for exploration and experimentation.

Furthermore, our study used self-made glasses with Leap Motion placed on them. Thus, compared with desktop applications, our gesture recognition has fewer occlusion problems, the range of motion is wider, and the user can walk around freely. Meanwhile, unlike the Leap Motion installed on the helmet horizontally and parallel to the helmet, in our self-made glasses, the Leap Motion is adjusted downward at a certain angle and attached to the 3D glasses, so it can be recognized even if the hand is hanging down without being lifted up, so the gesture recognition and tracking loss is decreased, which is also reflected in the log.

There are other more factors having potential effects on the freehand or controller based interaction in VR. These potential factors include but not limited to the type of the VR setup (e.g., large display and HMD), the scale of the VR scene, the posture and the state of the user (e.g., sitting, static standing and moving) and the user's familiarity with the hand gestures. The findings achieved in this study are based on a most universal and representative VR application in both a large display and a HMD -based environment, thus can benefit the majority of freehand gestural interaction applications in VR. But we have to acknowledge that for more specific task design in VR applications, other design factors such as the hand input modality, hand triggering mechanism and the hand interface geometry are also vital and should be specifically considered. Earlier research such as [17] had investigated on these factors and provided useful information about designing freehand mid-air drawing or writing tasks in VR. Task performance such as the hand operational speed, accuracy and the user perceived workload were also evaluated in their work. Combining all concerned factors and related findings from our study and [17], we can develop a more comprehensive insight for designing and evaluating freehand gestural interaction techniques in VR.

As a systematic and comprehensive evaluation, the present study presented an example of interaction measurement in a VR application. The proposed technique is also able to be utilized in AR or mixed reality (MR). AR and MR allow people to interact with virtual

objects and content overlaid on top of a world-tracking model. By incorporating the proposed gesture recognition system into AR and MR experiences, users can interact with virtual objects and interfaces using natural hand movements, thereby enhancing the overall user experience and immersion in these environments. Overall, gesture recognition is a valuable technology, but more extensive research is needed.

6.2 Limitations and future directions

This study had certain limitations that must be addressed. The first limitation is regarding the type of task proposed to the subjects. Our goal was to develop a set of common interaction methods that can conform to the user's usage habits and be easily accepted, so we did not set up a separate mini-task for each gesture to precisely test it. In addition, the participants did not have much knowledge of gesture recognition; therefore, they usually focused more on usability metrics, such as discoverability, learnability, and memorability, rather than identifiability and high recognition accuracy required for a gestural system in a gesture elicitation procedure.

The second limitation is regarding the apparatus used. Our goal was to empirically investigate the performance of our proposed gestures in VR. These gestures can be used in many apparatuses, such as various kinds of HMD. We used a Leap Motion placed on the user, which was appropriate for this purpose as it is widely used by the HCI community. However, the results may have been influenced by the sensor itself when there is occlusion or misidentification.

Among the freehand gestures, users agreed that zooming in and out and rotating are easier to achieve than with controllers. Some users also felt that remote pointing is not as useful as mouse pointing and is associated with greater levels of fatigue than mouse pointing. This fatigue may stem from the “Gorilla arm” effect, i.e., hanging and extending your arms to interact with a vertically oriented display. Gesture-based motion, on the other hand, has the advantage that it can be performed while the user is sitting or standing with little physical activity. Users felt that with joystick or trigger technology, continuous controlled walking is required, which is prone to visual fatigue and vertigo, while the teleportation operation can bring a more relaxed experience and the ability to move very large distances with minimal effort. Research suggests that our gesture design may contribute to this. There are also users who preferred a mix of gamepads and gestures.

7 Conclusions and future work

In this study, based on the practical user interaction requirements in general VR applications, a set of freehand gestures was designed to achieve the interactive operation of freehand technology in VR while considering the interaction problems caused by the limitations of current hand tracking sensors. Eleven interactive gestures and controller-based motion techniques were implemented. Designed for entering the scene and manipulating objects, these techniques were studied quantitatively and qualitatively by comparing different gestures and controller-based solutions. All participants were required to complete a variety of prescribed actions in an immersive environment. The study collected data related to performance measures, such as efficiency (i.e., completion time) and effectiveness (i.e., execution errors, motion interruptions, and tracking errors), as well as data related to assessing user perception using UPE, UPTD, and UPSU.

To fully study the gesture operation experience and exclude other interference factors, we developed systems based on a VR HMD and large screen. As the user will also change the viewpoint when turning the head from side to side while wearing the HMD, the use of a steering-based operation seems unnecessary in this case. Moreover, when the object is far away from the user and needs to be selected by ray intersection, perception in HMD may not be more real than that in front of the large screen. Therefore, a system based on a large screen was also developed to operate it. As the research focuses on the design of gesture operation, there is no comparison between HMD and large-screen operation. We discussed the application and comparison of gestures in common scenes, but for more general applications, further studies are required.

This study and the collected results will help to provide researchers and interaction experts with recommendations on the design of effective and efficient immersive VR motion technologies. In particular, the performance of freehand interaction technology and the issues related to tracking sensors will encourage further research in this field.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11042-023-17640-8>.

Acknowledgements We thank all reviewers for their insightful comments on this work. This research was supported by the National Natural Science Foundation of China under the grant number of “61902097”; the PRC Industry-University Collaborative Education Program under the grant number of “CES/Kingfar-202209RYJG15”; the “Natural Science Foundation of Zhejiang Province” under the grant number of “LQ19F020010”; and the Zhejiang Provincial Philosophy and Social Science Planning Project under the grant number of “19NDQN301YB”.

Data availability Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Conflict of interest The author has declared that there is no conflict of interest.

References

1. Achenbach J, Waltemate T, Latoschik ME, Botsch M (2017) Fast generation of realistic virtual humans. In: Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology (VRST), pp 1–10. <https://doi.org/10.1145/3139131.3139154>
2. Feng A, Suma E, Shapiro A (2017) Just-in-time, viable, 3d avatars from scans. In: ACM SIGGRAPH 2017 talks, Article no. 19. ACM, Los Angeles, pp 1–2. <https://doi.org/10.1145/3084363.3085045>
3. Vosinakis S, Koutsabasis P (2018) Valuation of visual feedback techniques for virtual grasping with bare hands using Leap Motion and Oculus Rift. *Virtual Real* 22(1):47–62. <https://doi.org/10.1007/s10055-017-0313-4>
4. Li Z, Jiang Y, Zhu Y et al (2022) Modeling the Noticeability of User-Avatar Movement Inconsistency for Sense of Body Ownership Intervention. *Proc ACM Interact Mob Wearable Ubiquitous Technol*: 12071. <https://doi.org/10.1145/3534590>
5. Wu H, Zhang S, Liu J, Qiu J, Zhang X (2019) The gesture disagreement problem in free-hand gesture Interaction. *Int J Human-Computer Interact* 35(12):1102–1114. <https://doi.org/10.1080/10447318.2018.1510607>
6. Sreenath S, Daniels DI, Ganesh ASD, Kuruganti YS, Chittawadigi RG (2021) Monocular tracking of human hand on a smart phone camera using mediapipe and its application in robotics. 2021 IEEE 9th Region 10 Humanitarian Technology Conference (R10-HTC), Bangalore, India pp 1–6. <https://doi.org/10.1109/R10-HTC53172.2021.9641542>
7. Reipschlgger P, Flemisch T, Dachselt R (2020) Personal augmented reality for information visualization on large interactive displays. *IEEE Trans Vis Comput Graph* 27(2):1182–1192

8. Guerra-Segura E, Travieso CM, Alonso JB (2017) Study of the variability of the leap motion's measures for its use to characterize air strokes. *Measurement* 105:87–97. <https://doi.org/10.1016/j.measurement.2017.04.016>
9. Fh A, Va A, Kai LB, Ch A, Bp A (2021) Mma B estimating depth information of vascular models: a comparative user study between a virtual reality and a desktop application. *Comput Graph* 98:210–217. <https://doi.org/10.1016/j.cag.2021.05.014>
10. Li B, Zhang C, Han C, Bai B (2019) Gesture recognition based on Kinect v2 and leap motion data fusion. *Int J Pattern Recognit Artif Intell* 33(5):1955005. <https://doi.org/10.1142/S021800141955005X>
11. Ogdon DC (2019) Hololens and vive pro: virtual reality headsets. *J Med Libr Assoc JMLA* 107(1):118–121. <https://doi.org/10.5195/JMLA.2019.602>
12. Facebook. Oculus Quest 2. Available online: <https://www.oculus.com/quest-2/>. Accessed 16 Mar 2021
13. Tian F, Lyu F, Zhang XL, Ren X, Wang H (2017) An empirical study on the interaction capability of arm stretching. *Int J Hum Comput Interact* 33(7–9):565–575. <https://doi.org/10.1080/10447318.2016.1265782>
14. Wu H, Luo W, Pan N, Nan S, Deng Y, Fu S et al (2019) Understanding freehand gestures: a study of freehand gestural interaction for immersive VR shopping applications. *Human-centric Comput Inform Sci* 9(1):1–26. <https://doi.org/10.1186/s13673-019-0204-7>
15. Masurovsky A, Chojeci P, Runde D, Lafci M, Przewozny D, Gaebler M (2020) Controller-free hand tracking for grab-and-place tasks in immersive virtual reality: design elements and their empirical study. *Multimodal Technol Interact* 4:91. <https://doi.org/10.3390/mti4040091>
16. Lee M, Kwahk J, Han SH, Lee H (2020) Relative pointing interface: a gesture interaction method based on the ability to divide space. *Int J Ind Ergon* 75:102878. <https://doi.org/10.1016/j.ergon.2019.102878>
17. Venkatakrisnan R, Venkatakrisnan R, Chung CH, Wang YS, Babu S (2022) Investigating a combination of input modalities, canvas geometries, and inking triggers on on-air handwriting in virtual reality. *ACM Trans Appl Percept* 19(4):1–19. <https://doi.org/10.1145/3560817>
18. Allgaier M, Amini A, Neyazi B, Sandalcioglu IE, Preim B, Saalfeld S (2021) Vr-based training of craniotomy for intracranial Aneurysm surgery. *Int J Comput Assist Radiol Surg* 17(3):449–456. <https://doi.org/10.1007/s11548-021-02538-3>
19. Adkins A, Lin L, Normoyle A, Canales R, Ye Y, Jrg S (2021) Evaluating grasping visualizations and control modes in a vr game. *ACM Trans Appl Percept* 18(4):19. <https://doi.org/10.1145/3486582>
20. Argelaguet F, Andujar C (2013) A survey of 3D object selection techniques for virtual environments. *Comput Graph* 37(3):121–136
21. Schfer A, Reis G, Stricker D (2022) Comparing controller with the hand gestures pinch and grab for picking up and placing virtual objects. 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops. <https://doi.org/10.1109/VRW55335.2022.00220>
22. Bowman DA, Hodges LF (1997) An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In: *Proceedings of the 1997 symposium on Interactive 3D graphics* 35
23. Su GE, Sunar MS, Ismail AW (2020) Device-based manipulation technique with separated control structures for 3d object translation and rotation in handheld mobile ar. *Int J Hum Comput Stud* 141:102433. <https://doi.org/10.1016/j.ijhcs.2020.102433>
24. Mendes D, Relvas F, Ferreira A, Jorge JA P (2016) The benefits of DOF separation in mid-air 3D object manipulation. *VRST '16-22nd ACM Conference on Virtual Reality Software and Technology*. ACM
25. Gerber D, Bechmann D. The spin menu: a menu system for virtual environments. *IEEE Conference on Virtual Reality*. In *Proceedings of the IEEE Conference on Virtual Reality*, Bonn, Germany, 12–16 March 2005: 271–272. <https://doi.org/10.1109/VR.2005.1492790>
26. Xia Z, Oyekoya O Hao Tang (2022) Effective gesture-based user interfaces on mobile mixed reality. *SUI '22* In: *Proceedings of the 2022 ACM Symposium on Spatial User Interaction*, pp 1–2. <https://doi.org/10.1145/3565970.3568189>
27. Al Zayer M, MacNeilage P, Folmer E (2020) Virtual locomotion a survey. *IEEE Trans Vis Comput Graph* 26(6):2315–2334. <https://doi.org/10.1109/TVCG.2018.2887379>
28. Paulo SF, Medeiros D, Lopes D, Jorge J (2022) Controlling camera movement in VR colonography. *Virtual Real* 26:1079–1088. <https://doi.org/10.1007/s10055-021-00620-4>
29. Zhang F, Chu S, Pan R, Ji N, Xi, L (2017) Double hand-gesture interaction for walk-through in VR environment. In *2017 IEEE/ACIS 16th international conference on computer and information science (icis)*, pp 539–544. Wuhan, China.
30. Caggianese G, Capece N, Erra U, Gallo L, Rinaldi M (2020) Freehand-steering locomotion techniques for immersive virtual environments: a comparative evaluation. *Int J Hum Comput Interact* 36:1734–1755. <https://doi.org/10.1080/10447318.2020.1785151>
31. Poupyrev I, Ichikawa T, Weghorst S, Billinghurst M (1998) Egocentric object manipulation in virtual environments: empirical evaluation of interaction techniques. *Comput Graphics Forum* 17(3):41–52. <https://doi.org/10.1111/1467-8659.00252>

32. Habgood M, Moore D, Wilson D, Alapont S. Rapid, continuous movement between nodes as an accessible virtual reality locomotion technique. In: Proceedings of the 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Tuebingen/Reutlingen, Germany, 18–22 March 2018: pp 371–378
33. Medeiros D, Sousa A, Raposo A, Jorge JM (2019) Magic carpet: Interaction fidelity for flying in VR. *IEEE Trans Vis Comput Graph* 26:2793–2804. <https://doi.org/10.1109/TVCG.2019.2905200>
34. Argelaguet F, Hoyet L, Trico M, Lécuyer A (2016) The role of interaction in virtual embodiment: effects of the virtual hand representation. In: *IEEE Virtual Reality*, Mar 2016, Greenville, United States. IEEE, pp 3–10. <https://doi.org/10.1109/VR.2016.7504682>
35. Tran TQ, Shin H, Stuerzlinger W, Han J (2017) Effects of virtual arm representations on interaction in virtual environments. In: *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology* pp 1–9
36. Venkatakrishnan R, Raveendranath B, Pagano CC, Robb AC, Lin WC, Babu SV (2023) How virtual hand representations affect the perceptions of dynamic affordances in virtual reality. *IEEE Trans Vis Comput Graph* 29(5):2258–2268
37. Lougiakis C, Katifori A, Roussou M, Ioannidis IP (2020) Effects of virtual hand representation on interaction and embodiment in HMD-based virtual environments using controllers. In: *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, Atlanta, pp 510–518. <https://doi.org/10.1109/VR46266.2020.00072>
38. Ismail AW, Billinghamurst M, Sunar MS (2015) Vision-based technique and issues for multimodal interaction in augmented reality[C] *ACM*. In: *8th International Symposium on Visual Information Communication and Interaction (VINCI)*. ACM, Tokyo, pp 75–82. <https://doi.org/10.1145/2801040.2801058>
39. Du MH, Cui H, Wang Y, Duh HBL (2023) Learning from deep stereoscopic attention for simulator sickness prediction. *IEEE Trans Vis Comput Graphics* 29(2):1415–1423. <https://doi.org/10.1109/TVCG.2021.3115901>
40. <https://developer.oculus.com/resources/hands-design-intro/>
41. Ankit C, Jagdish RL, Karen D, Sonia R (2011) Intelligent approaches to interact with machines using hand gesture recognition in natural way: a survey. *Int J Comput Sci Eng Survey* 2(1):122–133
42. Lou ZH, Yin JB (2018) Analysis of the gesture design principles in human-computer interaction. *Softw Guide* 17(4):19–24
43. Ni T, Bowman DA, North C, McMahan RP (2011) Design and evaluation of freehand menu selection interfaces using tilt and pinch gestures. *Int J Hum Comput Stud* 69(9):551–562
44. Ramadoss P, Rapetti L, Tirupachuri Y, Grieco R, Milani G, Valli E et al (2022) Whole-body human kinematics estimation using dynamical inverse kinematics and contact-aided lie group kalman filter. <https://doi.org/10.48550/arXiv.2205.07835>
45. Frøkjær E, Hertzum M, Kasper (2000) Measuring usability: are effectiveness, efficiency, and satisfaction really correlated? The SIGCHI conference. *ACM*, pp 345–352. <https://doi.org/10.1145/332040.332455>
46. ISO I. 9241--11 (1998) Ergonomic requirements for work with visual display terminals (VDTs) - Part 11: Guidance on usability. CEN, Brussels
47. Borg GA (1982) Psychophysical bases of perceived exertion. *Med Sci Sports Exerc* 14(5):377–381. <https://doi.org/10.1249/00005768-198205000-00012>
48. Hart SG, Staveland LE (1988) Development of NASA-TLX (task load index): results of empirical and theoretical research. In: *Advances in psychology* 52(6):139–183. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
49. Brooke JSUS (2013) A retrospective. *J Usability Stud* 8(2):29–40. <https://doi.org/10.1080/01440368008530710>
50. Wei LuZ, Tong ChuJ (2016) Dynamic hand gesture recognition with leap motion controller. *IEEE Signal Process Lett* 23(9):1188–1192. <https://doi.org/10.1109/LSP.2016.2590470>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.