

RESEARCH

Open Access



Understanding freehand gestures: a study of freehand gestural interaction for immersive VR shopping applications

Huiyue Wu^{1,2*} , Weizhou Luo¹, Neng Pan³, Shenghuan Nan¹, Yanyi Deng¹, Shengqian Fu¹ and Liuqingqing Yang¹

*Correspondence:

wuhuiyue@mail.sysu.edu.cn

¹The School of Communication and Design, Sun Yat-sen University, Guangzhou, China
Full list of author information is available at the end of the article

Abstract

Unlike retail stores, in which the user is forced to be physically present and active during restricted opening hours, online shops may be more convenient, functional and efficient. However, traditional online shops often have a narrow bandwidth for product visualizations and interactive techniques and lack a compelling shopping context. In this paper, we report a study on eliciting user-defined gestures for shopping tasks in an immersive VR (virtual reality) environment. We made a methodological contribution by providing a varied practice for producing more usable freehand gestures than traditional elicitation studies. Using our method, we developed a gesture taxonomy and generated a user-defined gesture set. To validate the usability of the derived gesture set, we conducted a comparative study and answered questions related to the performance, error count, user preference and effort required from end-users to use freehand gestures compared with traditional immersive VR interaction techniques, such as the *virtual handle controller* and *ray-casting* techniques. Experimental results show that the freehand-gesture-based interaction technique was rated to be the best in terms of task load, user experience, and presence without the loss of performance (i.e., speed and error count). Based on our findings, we also developed several design guidelines for gestural interaction.

Keywords: Immersive virtual reality, Elicitation study, User-defined gestures, Online shopping, User experience

Introduction

Compared to brick-and-mortar retail stores, online shopping has many advantages, such as unrestricted shopping hours as well as a greater focus on functionality and more efficient information retrieval. However, the current online shopping systems only present products using text and images, and they cannot provide end-users with an immersive shopping experience [1–4]. For end-users, the product representations in the form of images and text in scrollable lists are difficult to understand, i.e., end-users cannot obtain a clear sense of the size, weight, and shape of the product. In addition, the unnatural interaction techniques, such as scrolling a list or navigating through product

information pages, also increase the workload of the end-users (e.g., effort and frustration) as well as subsequently reduce their shopping experience (e.g., presence, immersion, and attractiveness).

With the rapid development of computer graphics and data visualization techniques, most VR shopping systems can simulate traditional brick-and-mortar retail stores. With products represented in the form of 3D models in VR shopping applications, end-users can view products from different perspectives and showcase the details of items (e.g., to show the material and texture). Therefore, VR shopping has emerged as a new trend and has been widely applied in shopping experiences [1–10]. Outside the scope of VR shopping applications in academia, commercial solutions have also emerged on the market, such as ShelfZone VR,¹ eBay,² and Macy's VR.³

However, traditional VR shopping applications merely render the digital representation of a brick-and-mortar retail store and have lacked more natural interactive techniques. For example, end-users are equipped with the conventional mouse [6, 7, 9], joystick [8] or handle controller [4, 11] to interact with the virtual environment. Compared to the two-handed interaction mode in real stores, the limited interactivity of such input techniques may impair the workload of end-users and the user experience.

In this paper, we conducted two user studies to understand gesture design and application for immersive VR shopping environments. The main contributions of this paper are the following: (1) we propose a more practical approach to derive user-defined gestures than traditional elicitation studies and/or user-centered design methods; (2) we present the quantitative and qualitative characterizations of user-defined gestures for shopping tasks in an immersive VR environment, including a gesture taxonomic analysis of the gestures defined by the users and (3) we contribute to the existing body of knowledge on immersive VR by empirically demonstrating the performance benefits and user preference for using freehand gestures for VR shopping tasks compared to current commonly used VR interactive techniques, such as the *virtual handle controller* and *ray-casting*. We hope that our work will lay a theoretical foundation for gestural interaction for immersive VR environments.

Related work

In this section, we review prior research related to gestural interactions in VR shopping environments and elicitation studies of freehand gestures.

Gestural interaction in VR shopping applications

The freehand-gesture technique is a direct mapping of the user's hand motions in the physical world to the affected motions in a computer system. With the rapid development of computer vision techniques, sensor technologies, and human-computer interaction techniques, freehand gestures have been applied in many VR applications for tasks such as object manipulation [12–17], navigation [18–20], and system control [21, 22].

¹ <https://invrsion.com/shelfzone>

² <https://vr.ebay.com.au>.

³ <https://goo.gl/h22ezQ>.

Beyond the above-mentioned applications, some researchers have also explored gestural interaction for VR shopping. Previous research has shown that VR outperforms traditional 2D e-commerce systems due to the improved shopping experience [1, 2, 23]. With the increased user satisfaction provided by VR shopping systems, a user tends to make not only more purchases at one time, but also make repeat purchases [3, 4]. However, most prior VR shopping systems have merely virtualized and digitized physical brick-and-mortar stores and have lacked more natural interactive techniques. In those systems, the users were equipped with the conventional mouse [6, 7, 9], joystick [8] or handheld controller [4, 11] to select and manipulate virtual products, while in the physical world, people usually work with their hands. Therefore, freehand-gesture-based VR shopping can relate the physical manipulation of our physical world to provide virtual control of information space and leads to improved shopping experiences that combine the advantages of shopping online and offline.

The benefits of gesture-based online shopping have been demonstrated by many researchers. Badju et al. [24], for example, elicited a set of freehand gestures from end-users for such tasks as object manipulation and system control in an online shopping system. Similarly, Altarteer et al. [25] explored the feasibility of freehand gestures for interacting with a luxury brand online store. Their studies indicate that gestural interaction can substantially improve the users' shopping experiences by enabling them to perform a variety of shopping tasks, such as trying on new clothes or mixing and matching accessories without being physically present in a real shopping mall. Similar to our study, Verhulst et al. [20] conducted an experiment to compare the performance and user preferences for body gestures and a traditional game-pad. The experimental results showed that although body gestures are slower than the traditional game-pad, they are more natural and enjoyable when collecting various products in an immersive virtual supermarket system.

Gesture elicitation study

Although free-hand gestures have attracted worldwide attention in recent years, most of the gesture-based applications mentioned above were designed by professional system developers. End-users usually have few opportunities to participate in gesture design. In some cases, the usability of gestures may be overlooked by system designers in pursuit of high recognition performance and/or ease of implementation [26, 27]. As a result, the gesture disagreement problem [28] may occur between "good" gestures imagined by system designers and "good" gestures chosen by end-users. Similar to the vocabulary problem proposed by Furnas et al. [29] for information retrieval systems, the gesture disagreement problem may lead to a decrease in system usability and user satisfaction for gesture-based applications.

To design for the increasing number of freehand-gesture-based applications, we must understand how to design and identify "good" gestures that are discoverable, learnable, memorable, and easy to use in a human-computer interaction context. To address those issues, researchers have proposed a gesture elicitation method, in which the target users of a gestural system are invited to participate in the gesture design processes. In a standard elicitation study, end-users are first shown the initial and final states of a target task and then are required to design the best gesture for that task. Then, the system designers

compile all the gesture candidates and assign the top gesture (i.e., the most commonly selected gesture) for that task. The benefits of gesture elicitation studies have been demonstrated by previous studies. Morris et al. [30], for example, found that user-defined gestures are easier to memorize and discover than those designed solely by professional system developers.

However, most standard gesture elicitation studies adopted a “1-to-1” experimental protocol, which means participants were required to design only a single gesture for each task [11, 31–36]. In practice, such approach inevitably encounters the legacy bias problem [37] because when end-users are involved in gesture design their gesture choices are often biased by their experience with prior interfaces (e.g., graphical user interfaces) and technologies (e.g., multitouch-based techniques) that have recently been standard on traditional personal computers (PCs) or mobile phones. In addition, due to the influence of such factors as time and experimental conditions, end-users may fail to design the most appropriate gestures for given tasks. As a result, the traditional gesture elicitation study may not necessarily generate the globally optimal gesture set for specified system tasks.

Recently, some researchers [26, 37–40] proposed a new “1-to-3” experimental protocol for gesture elicitation, in which participants were required to derive three gesture candidates for each target task. They speculated that such a method may prompt end-users to think more deeply about which gestures were most appropriate for specified tasks, rather than directly using those legacy-inspired gestures popped out of their minds easily. However, such an approach raises a new problem—participants had difficulty designing three gestures for each given target task, especially when they already had a “good” gesture in mind [26, 40].

To address this issue, Wu et al. [10] proposed a more practical “1-to-2” experimental protocol, i.e., eliciting two gestures for each task in gesture elicitation studies. They reported that this approach can effectively alleviate the legacy bias problem without imposing too much cognitive burden on participants. However, Wu et al. generated two sets of user-defined gestures for their system without further discussing which one to choose. As a result, they transferred the burden of experimental participants to the potential users of the final system who will have to memorize many gestures for a single system.

In general, the design of gestural interaction remains a challenge due to the lack of general design guidelines and established conventions. Traditional elicitation method used the frequency ratio to select top gestures derived from participants for system tasks. Although different protocols such as “1-to-1”, “1-to-2” and “1-to-3” have been proposed to optimize the guessability procedure, those approaches relied heavily on the participants’ proposals while ignoring the designers’ contributions. In addition, existing gesture elicitation studies mostly halted in the initial gesture design stage; it remains unclear whether those gestures derived directly from nonprofessional users would perform well in terms of system performance and user satisfaction.

Study 1

To design more natural and user-friendly gestural VR shopping applications, we need to understand how to involve end-users more effectively in the gesture elicitation process and analyze their gesture proposals more comprehensively. Due to the open-ended

Table 1 Ten essential gestural tasks for a VR shopping application

No	Task name	Description	Frequency
1	Select	Select an object	30 (100%)
2	Rotate	Rotate an object	30 (100%)
3	Try on	Try on clothes	30 (100%)
4	Change color	Change to next/previous color	30 (100%)
5	Change size	Change to a larger/smaller size	30 (100%)
6	Enlarge	Enlarge an object	27 (90%)
7	Shrink	Shrink an object	27 (90%)
8	View	View product details	25 (83%)
9	Add	Add to a shopping cart	25 (83%)
10	Close	Close the current window	20 (67%)

nature of the gesture elicitation method, we decide to conduct two independent gesture elicitation studies to address the risk that a standard gesture elicitation study may be trapped around the local optima of the objective vocabulary and fail to generate the best suitable gestures for the corresponding tasks. Then, system designers are asked to resolve the conflicts that might arise between the two studies by applying their professional experience.

Participants

Sixty participants (29 male and 31 female) aged between 22 and 37 ($M=27.94$, $SD=3.809$) were recruited in this study. The participants came from different professional backgrounds, including programmers, salesclerks, market analysts, and university students. Although all participants had at least 3 years of online shopping experience, they had never used gestures for immersive VR shopping.

Tasks

To determine the essential interaction tasks and guarantee the usability of gestures for a VR shopping application, we first collected requirements for VR shopping systems from both popular e-commerce platforms and previous literature on gesture-based VR interaction systems [2–4, 20, 23–25, 41]. In this manner, a total of 29 common tasks were collected. Next, we conducted a brainstorming session in which 30 of the participants were invited to vote and rank the tasks with a 5-point Likert scale according to their importance (1 = worst, 5 = best) in gestural interaction with such a system. As a result, we generated ten core tasks. We ranked the ten core tasks according to popularity (Table 1). A core task was selected only if a minimum of half of the 30 participants chose it and the average importance score for the task was greater than three.

Apparatus

We conducted this experiment in a usability lab. The lab configuration consisted of a PC, a depth sensor (Leap Motion), a commercial helmet-mounted display (HTC Vive), and two wireless handle controllers (Fig. 1). We developed an immersive VR shopping application based on Unity 3D, which was delivered through the PC to the HTC display. During the experiment, we used 5 web cameras from different perspectives to record the



participants' gestures and their soliloquies for latter data analysis. Following traditional elicitation protocol, we did not give participants any hints, in order to prevent bias.

Procedure

Before the experiment, we first introduced the purpose and tasks of the experiment to all 60 participants. Then, participants went through an informed consent process. After that, they were randomly assigned to two 30-person groups, i.e., Group 1 and Group 2. Our aim here was to verify the consistency of the resulting user-defined gestures produced by two independent gesture elicitation studies. During the experiment, we asked participants to use the provided VR shopping system to finish the ten tasks (Table 1). As soon as they heard the instructions for the task from the experimenter, participants were required to design the best gesture for this task. To prevent potential bias caused by the current gesture recognition techniques, we used a "Wizard-of-Oz" method rather than a real gesture-based interaction system. In this method, participants thought they were using a real gesture-based VR shopping system, but instead, the experimenter (Wizard) complete tasks according to participants' gestures using the HTC handle controller. We also used a "think-aloud" method to collect participants' design rationales. All participants were asked to say out loud the reason they designed a gesture for a particular task. After they finished the experiment, we asked participants to answer a short questionnaire about their demographic data, including age, gender, professional background, and their suggestions for this study. A Latin square was used to counterbalance the possible order effects of the ten target tasks. Each experiment lasted approximately 40 to 80 min.

Results

In this section, we present the gesture taxonomies developed based on the user-defined gestures created by the two user groups, the agreement scores, and the selected top gestures according to a professional designer's suggestions.

Data analysis

With two user groups, 30 participants in each group, and the ten essential VR shopping tasks, we collected a total of 600 ($2 \times 30 \times 10$) freehand gestures. Then, three professional designers were invited to a brainstorming session to discuss how to group and merge similar gestures for the corresponding task. All three designers had at least 5 years of experience in gesture-based interface design. Gestures with the same features will be grouped into a single gesture, while gestures with similar features need to be discussed how to group according to the design rationales participants articulated during the experiment. For example, 15 *Grab* actions with different numbers of fingers can be merged into one group of identical gestures, as one participant stated:

I would like to use a Grab gesture for Task 1—Select an object in an immersive VR shopping environment. It doesn't matter whether I perform it with a whole hand or just with the thumb, index finger and middle finger.

After the grouping process is complete, we obtained 62 and 51 groups of identical gestures from Group 1 and Group 2, respectively.

Gesture taxonomy

Next, we manually classified the user-defined gestures produced by the two groups along the four dimensions of *nature*, *body parts*, *form*, and *viewpoint*. Each dimension is divided into multiple categories (Table 2). Different from the seminal work on gesture taxonomy by Wobbrock et al. [31] for surface computing, this taxonomy was adapted to match gestural interaction for immersive VR shopping environments.

Table 2 Taxonomy of freehand gestures for immersive VR shopping applications

Nature	
Physical	Gesture acts physically on an object
Symbolic	Gesture depicts a symbol
Metaphorical	Gesture indicates a metaphor
Abstract	Gesture-task mapping is arbitrary
Body parts	
One hand	Gesture is performed with one hand
Two hands	Gesture is performed with two hands
Full body	Gesture involves movements of at least one other body part in addition to the hands
Form	
Static	Hand pose is held in one location
Dynamic	Gesture contains movements of one or more body parts during the stroke phase
Viewpoint	
Independent	Gesture is independent of the viewpoint
Object-centric	Gesture is performed from the object's point of view
User-centric	Gesture is performed from the user's point of view

The *Nature* dimension is divided into four categories of physical, symbolic, metaphorical, and abstract. Physical gestures usually act directly on virtual objects, e.g., grabbing an item. Symbolic gestures are visual depictions, e.g., drawing a “+” in the air to add the current item to the cart. Metaphorical gestures occur when a gesture acts on, with, or like something else, e.g., the user views the palm as a color palette to change the color of a selected item of clothes. Finally, abstract gestures have no physical, symbolic, or metaphorical connection to the corresponding tasks. The mapping between an abstract gesture and a target task is arbitrary.

The *Body parts* dimension refers to how many body parts are involved in a gesture. It distinguishes between one-handed gestures, two-handed gestures, and full-body gestures that involve at least one other body part.

The *Form* dimension distinguishes between static and dynamic gestures. Static gesture refers to hand shape or finger configuration while dynamic gesture involves the spatiotemporal movement of the hand [42].

The *Viewpoint* dimension describes the relative location where gestures are performed. Object-centric gestures act on specific virtual objects, e.g., rotating an item in the virtual shopping environment. User-centric gestures are performed from the user’s point of view, e.g., when the user is pointing to his/her own body, the selected clothes should move in the pointing direction and, therefore, to the user’s body. Independent gestures require no information about the world and can occur anywhere, e.g., crossing the arms to indicate closing the current window.

Using the abovementioned taxonomy, we present the breakdown of the user-defined gestures from the two user groups in Fig. 2.

Although the percentage of gestures in each taxonomy dimension from the two groups is slightly different, we can find the following common patterns: (1) half of the user-defined gestures were object-centric; (2) participants preferred dynamic gestures and performed them with one hand; (3) participants proposed a few more metaphorical gestures than physical, abstract, or symbolic gestures.

Agreement scores

To evaluate the consistency of gesture choices between participants, we calculated the agreement scores (AS) for the ten target tasks following the agreement formula (Eq. 1) developed by Vatavu et al. [43].

$$AR(r) = \frac{|P|}{|P| - 1} \sum_{P_i \subseteq P} \left(\frac{|P_i|}{|P|} \right)^2 - \frac{1}{|P| - 1} \quad (1)$$

where P is the set of all proposed gestures for task r , $|P|$ is the size of the set, and P_i represents subsets of identical gestures from P . The higher the agreement score is, the more likely the same gesture will be selected.

Figure 3 shows the agreement scores for the ten tasks and ranks them from large to small. As seen, the agreement scores of all tasks in the two groups are below 0.4. The average agreement scores of gestures for the ten target tasks in Group 1 and Group 2 are 0.190 ($SD=0.091$) and 0.225 ($SD=0.072$), respectively. According to Vatavu et al. [43], the average agreement scores for the ten tasks are medium (0.100–0.300) in magnitude.

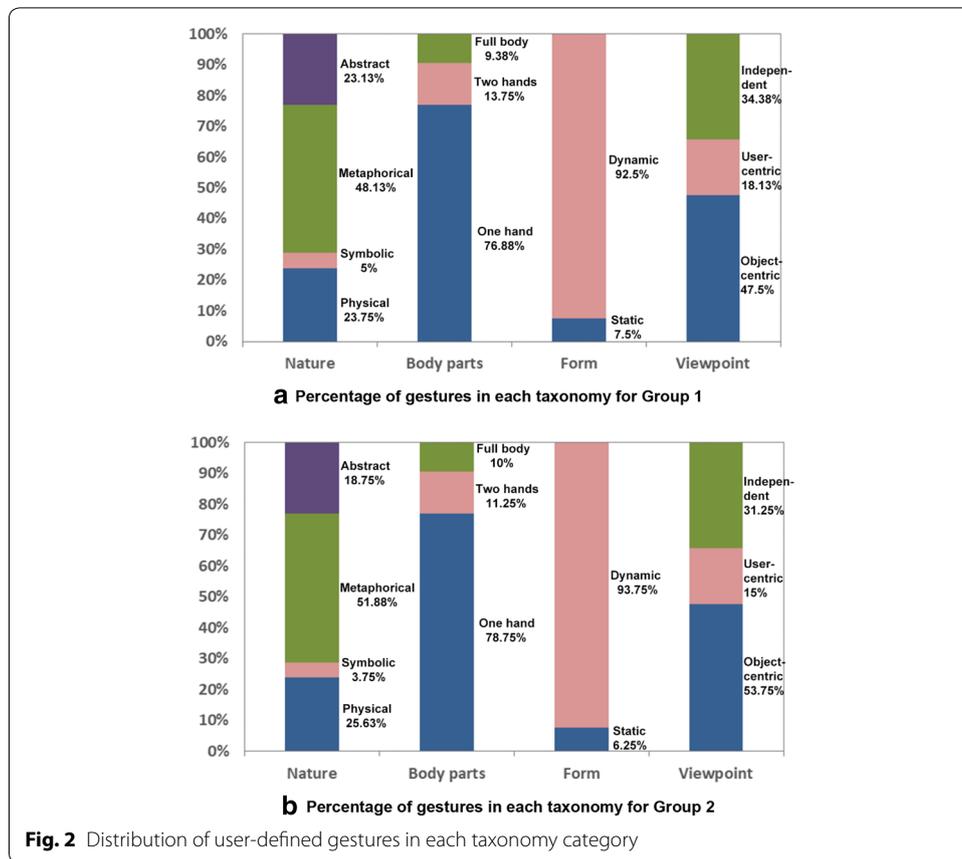


Table 3 shows the ten target tasks, the corresponding top gestures (i.e., the gestures with the highest frequency), and the agreement score for each task between the two participant groups.

From Table 3, we can see that although some similar patterns could be found between the two independent gesture elicitation studies (Figs. 2, 3), they produced some inconsistent top gestures. Half of the ten top gestures were different between the two groups, including top gestures for Tasks 4, 5, 6, 7, and 8. The disagreement rate for the top gestures between the two studies was 50%.

Conflict resolution from the designers’ perspective

At this point, we have obtained two different sets of user-defined gestures from two user groups who participated in two independent elicitation studies. Following the standard gesture elicitation study procedure, one might get confused and not know which gesture to choose and subsequently assign a gesture to a corresponding task (e.g., Tasks 4, 5, 6, 7, and 8) due to conflicts. Therefore, we invited five professional system designers to a brainstorming session to resolve conflicts between the two gesture sets. All five designers had more than 7 years of experience in developing gesture-based interactive systems.

As shown in Table 3, participants chose the same top gestures for Tasks 1, 2, 3, 9, and 10 between the two groups; considering their naturalness and legibility, the five designers immediately recommended the top five gestures for these five VR shopping tasks.

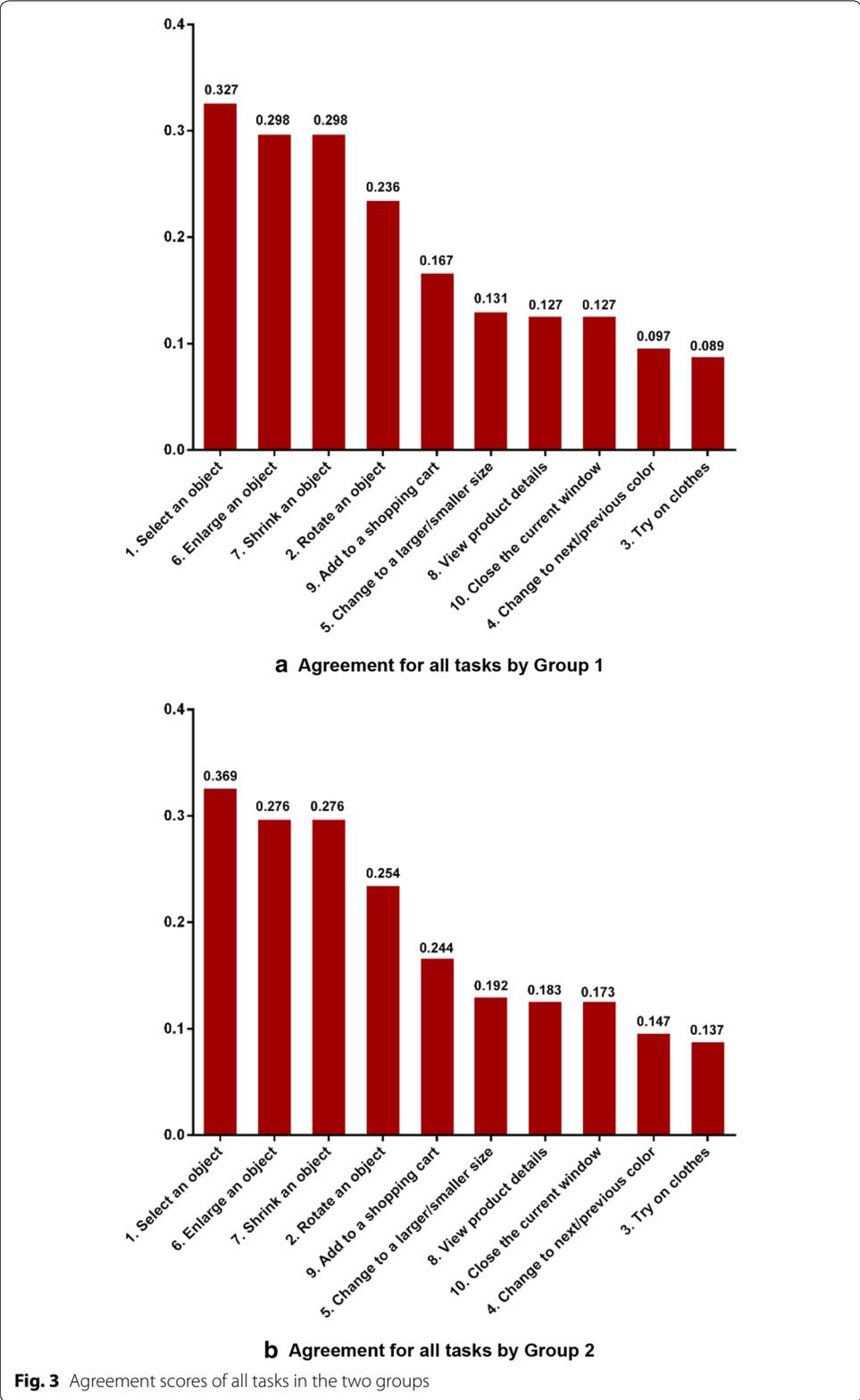


Table 3 Ten target tasks and the corresponding top gestures produced by two independent elicitation studies

Task	Group 1		Group 2	
	Top gesture	AS	Top gesture	AS
1. Select an object	Grab	0.327	Grab	0.369
2. Rotate an object	Twist	0.236	Twist	0.254
3. Try on clothes	Drag onto one's body	0.089	Drag onto one's body	0.147
4. Change to next/previous color	Swipe right/left	0.097	Tap on an imagined color palette	0.137
5. Change to a larger/smaller size	Swipe up/down	0.131	Tap on an imagined color palette	0.192
6. Enlarge an object	Both hands moving from the center middle to the outer left and right	0.276	Perform a pinch-out gesture with the thumb and index finger	0.298
7. Shrink an object	Both hands moving from the outer left and right to the center middle	0.276	Perform a pinch-in gesture with the thumb and index finger	0.298
8. View product details	Tap twice with the index finger	0.127	Open the fist	0.244
9. Add to a shopping cart	Drag onto an imagined shopping cart icon	0.167	Drag onto an imagined shopping cart icon	0.183
10. Close the current window	Swipe away	0.127	Swipe away	0.173

Next, the five designers were asked to evaluate the other top five gestures that differed between the two groups and to recommend the best gesture for each target task by considering both the performance and user preferences from the perspective of system designers.

For Task 4—Change to next/previous color and Task 5—Change to a larger/smaller size, participants proposed *Swipe right/left* and *Swipe up/down*, respectively, in Group 1. Compared to the gesture *Tap on an imagined color palette* chosen by participants in Group 2, these gestures require much more cognitive effort because participants have to remember the mapping relationships of different directions of hand movement and the corresponding interactive semantics. In addition, the interactive efficiency of these gestures would be greatly reduced if the user had to choose between many different colors or sizes in practice. In contrast, the user does not need to be concerned with these issues and can easily use the gesture *Tap on an imagined color palette* to perform such tasks. Therefore, the five designers recommended the use of the top gestures from Group 2.

For Task 6—Enlarge an object and Task 7—Shrink an object, participants chose *Both hands moving from the center middle to the outer left and right* and *Both hands moving from the outer left and right to the center middle*, respectively, in Group 1 instead of *Perform a pinch-out gesture with the thumb and index finger* and *Perform a pinch-in gesture with the thumb and index finger*, respectively, in Group 2. Although the two single-hand gestures from Group 2 are easier to perform and can reduce physical fatigue compared to the corresponding two-handed gestures from Group 1, they are prone to cause the “Midas Touch” problem in real-world scenarios [44], which refers to the phenomenon in which every “active” hand action from the user, even unintentional, could be interpreted as an interaction command by vision-based interfaces. Considering the performance problem in practice, the five designers recommended the top two gestures from Group 1 for Tasks 6 and 7.

For Task 8—View product details, participants preferred *Open the fist* to *Tap twice with the index finger* in Group 2. Similar to the action of double-clicking a mouse, the *Tap twice with the index finger* gesture designed by participants in Group 1 is a typical operation in Window, Icon, Menu, and Pointing (WIMP) device graphical user interfaces. However, this gesture contains two atomic actions in 3D space, and the frequency of the gesture varies widely among different users, which may lead to more recognition errors by vision-based gestural interactive systems. In addition, there is no strong semantic mapping between this gesture and Task 8. In contrast, the gesture *Open the fist* favored by participants in Group 2 had simple and clear movement and was in line with the user's mental model. Therefore, the five designers chose this gesture for Task 8.

Based on the above analysis, the five system designers recommended using the top five gestures produced by both groups (Tasks 1, 2, 3, 9, and 10), the top two gestures generated solely by Group 1 (Tasks 6 and 7), and three gestures by Group 2 (Tasks 4, 5, and 8).

User-defined gesture set

Based on the co-working of participants and professional designers, we derived a set of freehand-based gestures for immersive VR shopping environments (Fig. 4).

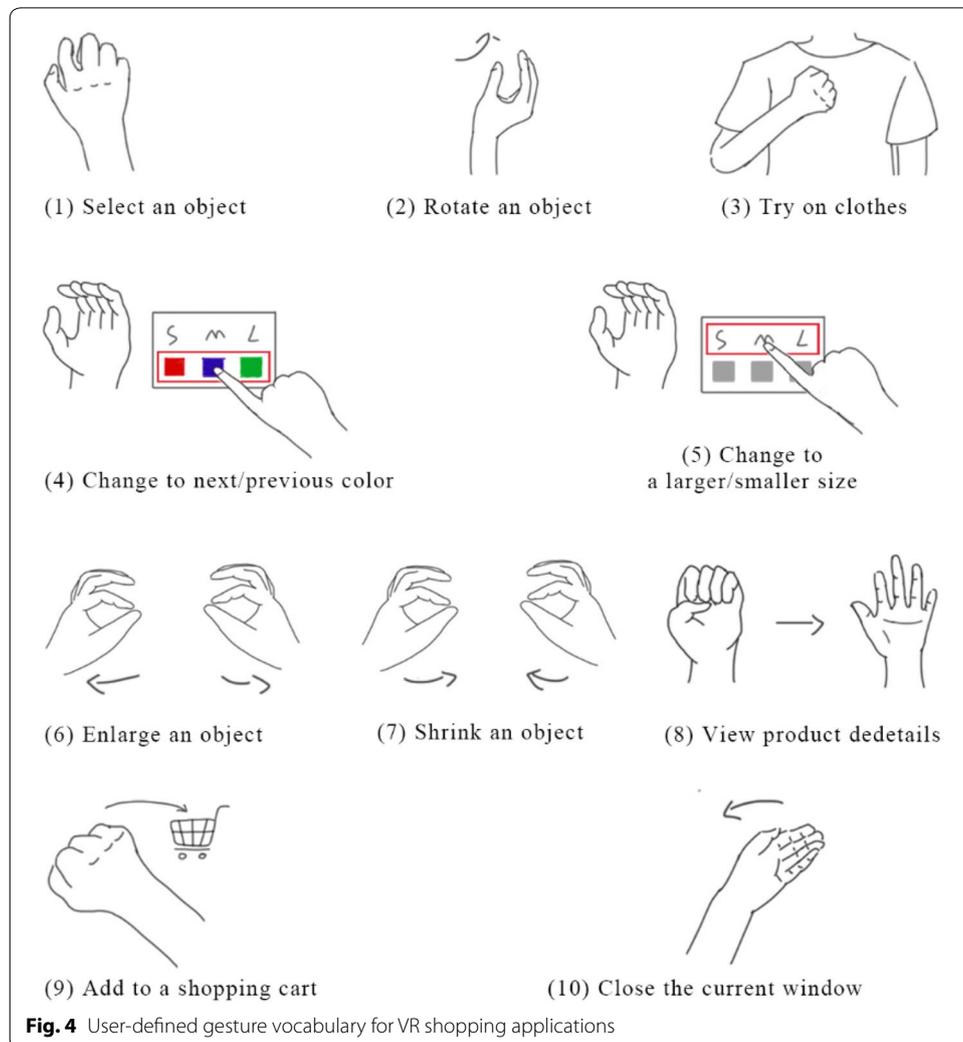
Discussion of Study 1

In this study, we learned about the essential target tasks and the most commonly used freehand gestures (i.e., the top gestures) for those tasks in a gestural VR shopping application. Different from standard gesture elicitation study procedure [31–36], we asked two groups of 30 participants to design gestures for the ten system tasks in two independent elicitation studies. The experimental results verified our hypothesis that the user-defined gesture set produced solely from a single elicitation procedure may be trapped in the local minima and fail to uncover gestures that may be better suited for given target tasks. As shown in Table 3, 50% of the top gestures changed between Group 1 and Group 2.

Fortunately, despite the disagreement concerning the top gestures among participants of the two groups, we found that participants showed consistency in the types of gestures they preferred for interaction with immersive VR shopping environments (Fig. 2). This information lays the foundation for the selection of gesture-recognition algorithms and interactive techniques in the latter stage of system development.

Figure 3 and Table 3 also suggest that for certain types of tasks such as Task 1—Select an object, one might obtain the same top gesture with a high agreement score by conducting two independent gesture elicitation studies. However, for other tasks, such as Task 6—Enlarge an object and Task 7—Shrink an object, one may obtain different top gestures in spite of the relatively high agreement scores in two independent gesture elicitation studies. In this case, the popular “frequency ratio” method and the “winner-take-all” strategy adopted by most standard elicitation studies [28, 31–36] may not be effective due to conflicts.

Based on the findings from Study 1, we suggest the need to involve professional designers to resolve conflicts and contribute to the gesture proposals by applying their professional skills.

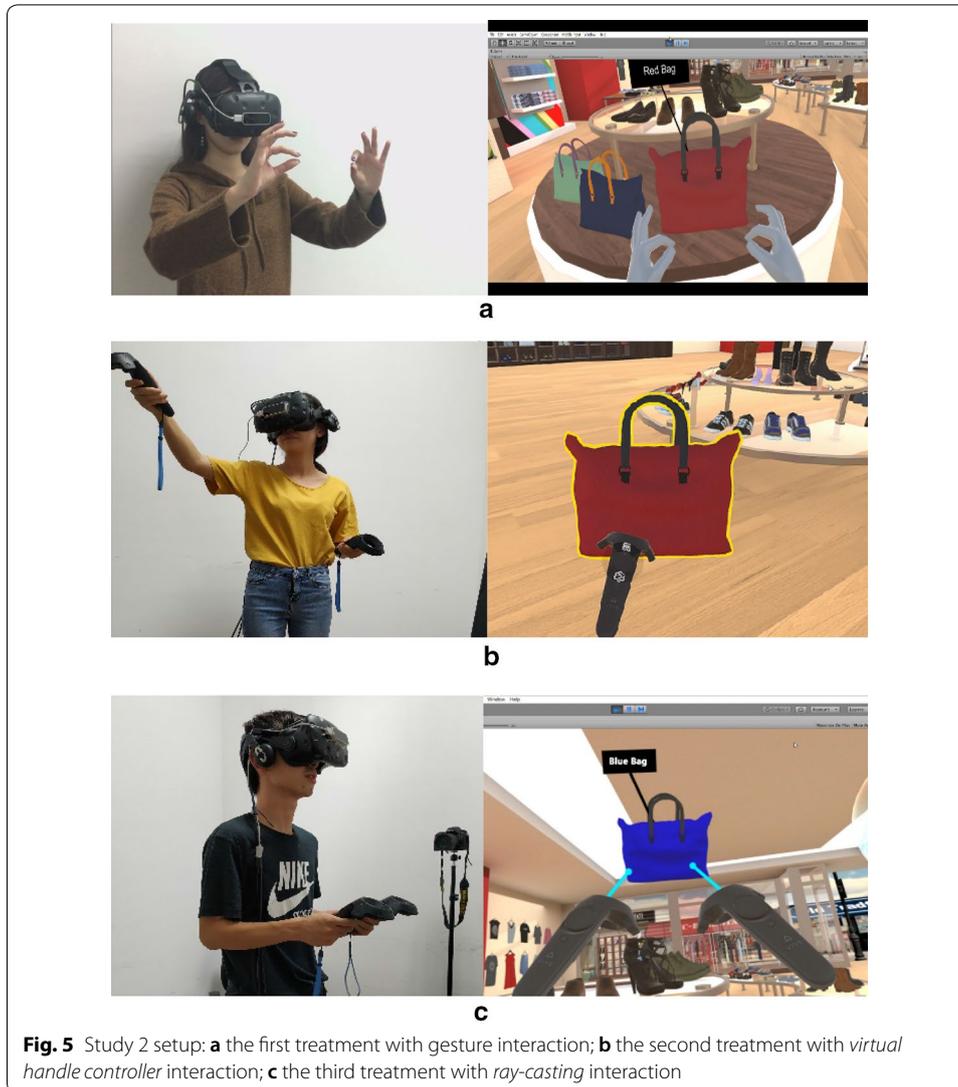


Study 2

In Study 1, we obtained a user-defined gesture vocabulary for shopping tasks in an immersive VR environment through the co-design of ordinary participants and professional system designers. To deepen our understanding of the usability and social acceptance [45] of the user-defined gestures, we developed a gesture-based VR shopping prototype. Based on the gestural system, we conducted a comparative study to investigate how participants perceived the benefits and shortcomings of using user-defined gestures as well as other traditional 3D interactive techniques to interact with a VR shopping application. We hope that this study will enable researchers to better understand the capabilities of gestural interaction techniques and consequently design and develop appropriate applications.

Experimental design

Most commercial immersive VR systems, such as the Oculus Rift and HTC Vive, provide two popular interactive techniques, the *virtual handle controller* and *ray-casting* (virtual



pointer). With the *virtual handle controller*, the user can grab and position virtual objects by “touching” and “picking” them with a virtual representation of the real handle controller. A typical *virtual handle controller* technique provides a one-to-one mapping between the real and virtual handle controllers. In comparison, *ray-casting* employs nonlinear mapping functions and a “supernatural” metaphor to extend the user’s area of reach by using a “laser ray”. Therefore, we compared the proposed gestural system with these two popular interactive techniques used in commercial systems.

This experiment included three treatments, each of which used a different input method for VR interaction. In the first treatment, participants were required to interact with the VR system through user-defined gestures (Fig. 5a). In the second treatment, participants used a virtual handle controller (Fig. 5b), which the user could use to press different buttons and/or different regions of the trackpad to perform different tasks, such as pressing the trigger button to select an object and then pressing the top region of the trackpad to produce a window with detailed information about the object. In the

third treatment, participants were provided with a “magic” laser beam (*ray-casting*) for object selection and manipulation (Fig. 5c). During the interaction process, two blue rays extending from the tips of the virtual handle controllers were activated and were used to manipulate the virtual objects. A finite-state machine (FSM) was designed to facilitate the user switch from one state (e.g., select an object) to another (e.g., enlarge a selected object) by following the transition rules we defined in advance.

Participants

In this experiment, we recruited 30 participants (14 male and 16 female) aged between 20 and 25 years ($M = 22.19$, $SD = 2.500$). They were pursuing different majors, including interaction design, journalism, atmospheric sciences, and computer science. None of the 30 participants had any experience with gestural interaction in an immersive VR shopping environment before this study. None of them participated in the previous study.

Apparatus

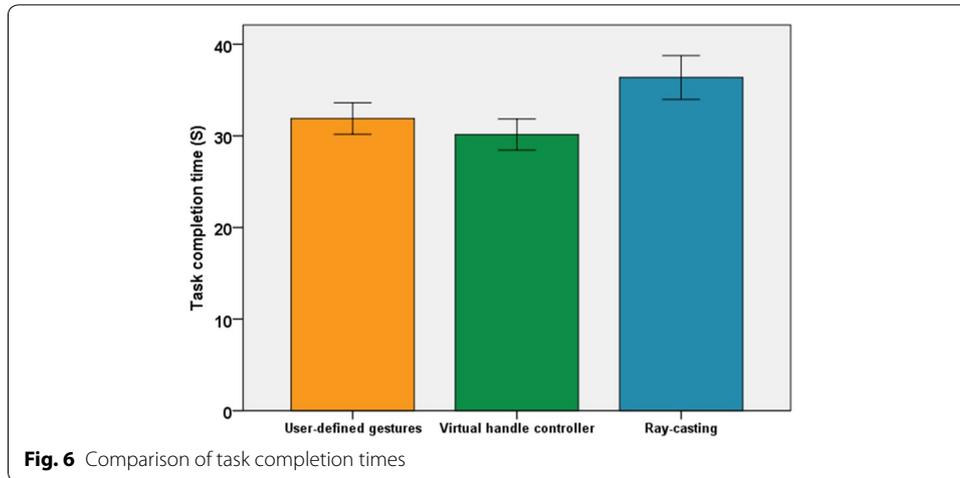
We conducted this experiment in a usability lab. The lab configuration consisted of a 15.6-inch HP laptop, an HTC Vive, two wireless handle controllers, and a Leap Motion sensor attached to the front panel of the HTC Vive (Fig. 5). To meet the requirements for the HTC Vive for VR scene rendering, the laptop used in this study had a 2.2GHz i7 CPU, 16G memory, and an 8GH GeForce GTX1070 graphics card. It also hosted a gestural VR shopping system that we developed in advance to process the user’s gesture inputs and deliver the virtual shopping environment to the HTC Vive. The gestural VR system was implemented based on the Leap Motion SDK, Virtual Reality Toolkit (VRTK) and the Unity game engine. Using our gesture recognition toolkit [42, 44, 46], the 10 user-defined gestures were tested on an average recognition rate of 98.6%. We used a web camera to capture the participants’ gesture behaviors and voices. We also used an iPad Pro to collect the participants’ answers to a short questionnaire after the experiment.

Task scenarios and procedures

We designed a set of typical VR shopping tasks for this experiment. The task set involved selecting a red bag, enlarging the red bag twice and then shrinking it to its normal size, rotating the red bag 180 degrees along the y -axis, viewing the detailed information for this bag (e.g., brand and price), invoking an attribute window to change its size from small to large as well as change its color from red to blue, closing the attribute window, selecting a T-shirt and trying it on, and finally putting the T-shirt into a shopping cart.

Participants were first introduced the experimental objective and requirements and then participated in an informed consent process. Next, they were allowed to practice until they completed a set of virtual object manipulation tasks similar to the real tasks by using the three treatments in a training scene.

Each participant was asked to complete the same shopping task set as quickly as possible using three different interactive techniques (treatments). Our experiment used a within-subject design. A Latin square was used to counterbalance the treatment orders. Participants were randomly assigned to these orders.



We collected data on user performance and user satisfaction with the three provided interactive techniques. The performance was measured by task completion time and error count. The completion time is straightforward; it was defined as the time interval between the moment a task started and the moment a participant correctly finished the task. The error count is the number of wrong attempts before a task was correctly finished.

After completing the set of shopping tasks, participants filled out a questionnaire about their opinions of the three different interactive techniques. The questionnaire included three parts. The NASA Task Load Index (NASA-TLX) [47] was used to measure the task load, the Usefulness, Satisfaction, and Ease of Use (UEQ) [48] was used to measure the user experience, and the Igroup Presence Questionnaire (IPQ) [49] was used to measure the sense of presence, i.e., the extent to which participants believed themselves to 'be there' in the immersive VR environments. The experiment lasted approximately 90 to 120 min.

Results

In this section, we report the experimental results in terms of task completion time, error count, task load, user experience, and presence.

Task completion time

Figure 6 compares the completion time for the three treatments. As shown, the average completion time for the ten target tasks by the 30 participants with user-defined gestures, the *virtual handle controller* technique, and the *ray-casting* technique are 31.9 s ($SD = 4.6$ s), 30.1 s ($SD = 4.5$ s), and 36.4 s ($SD = 6.4$ s), respectively. Using a one-way ANOVA test, we found that the differences in task completion time between the three treatments were significant ($F_{2,87} = 11.199$, $p = 0.000$). Post hoc analysis (Tukey's HSD) indicated that the *ray-casting* technique was significantly slower than the user-defined gestures ($p = 0.004$) and the *virtual handle controller* ($p = 0.000$). No significant difference was found between user-defined gestures and the *virtual handle controller* ($p = 0.405$).

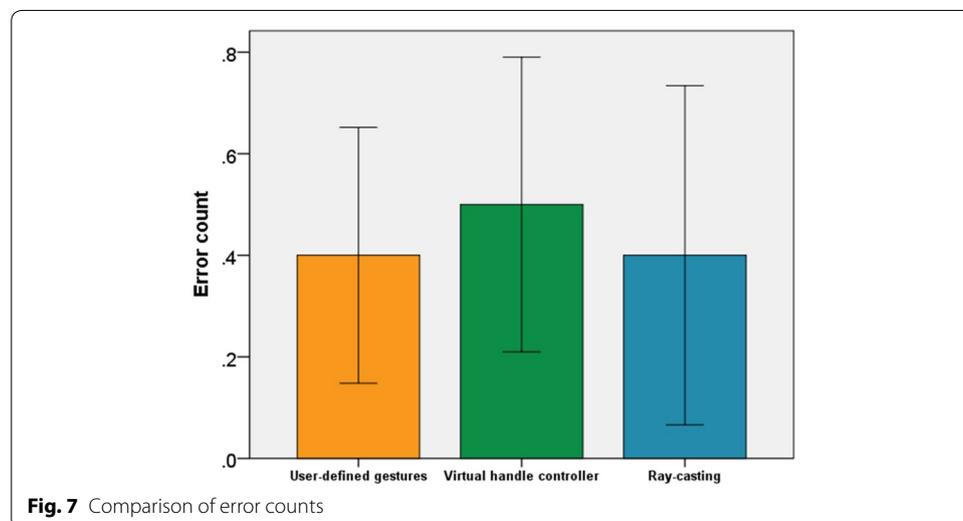
Next, we compare the average completion time for each single task. Using a one-way ANOVA test, we found that the differences in task completion time between the three treatments were significant for eight target tasks (80%). A post hoc analysis (Tukey's HSD) indicated the following: (1) user-defined gestures perform best for Task 2—Rotate an object ($F_{2,87}=26.511$, $p=0.000$), Task 3—Try on clothes ($F_{2,87}=10.403$, $p=0.000$), Task 8—View product details ($F_{2,87}=6.554$, $p=0.002$), Task 9—Add to a shopping cart ($F_{2,87}=8.922$, $p=0.000$), and Task 10—Close the current window ($F_{2,87}=3.749$, $p=0.027$); (2) the *ray-casting* technique performs best for Task 1—Select an object ($F_{2,87}=3.614$, $p=0.031$); and (3) the *virtual handle controller* technique performs best for Task 6—Enlarge an object ($F_{2,87}=5.804$, $p=0.004$) and Task 7—Shrink an object ($F_{2,87}=5.702$, $p=0.005$).

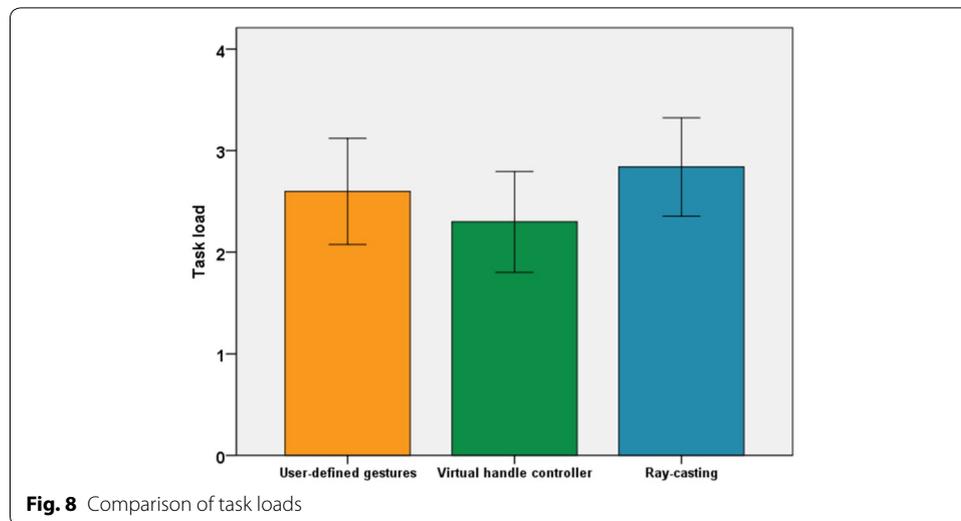
Error count

For each target task, we recorded the number of times the participants needed to use the different interactive techniques to complete the task. Because the total number of attempts is highly dependent on the participants' comfort levels and manipulation habits, we proposed an error count formula to calculate participants' total number of attempts:

$$\text{Error Count} = \begin{cases} 0, & \text{if the user finished a task on the first attempt} \\ 1, & \text{if the user finished a task after a second attempt} \\ 2, & \text{if the user finished a task after more than two attempts} \\ 3, & \text{if the user was unable to finish a task without explicit help from the experimenter} \end{cases}$$

Figure 7 compares the error counts for the three treatments. As shown, the average error count for the ten target tasks for the 30 participants with user-defined gestures, the *virtual handle controller*, and the *ray-casting* technique are 0.4 ($SD=0.675$), 0.5 ($SD=0.777$), and 0.4 ($SD=0.894$), respectively. Using the Friedman test, no significant difference was found in the error counts between the three treatments ($\chi^2(2)=1.560$, $p=0.458$).





Task Load

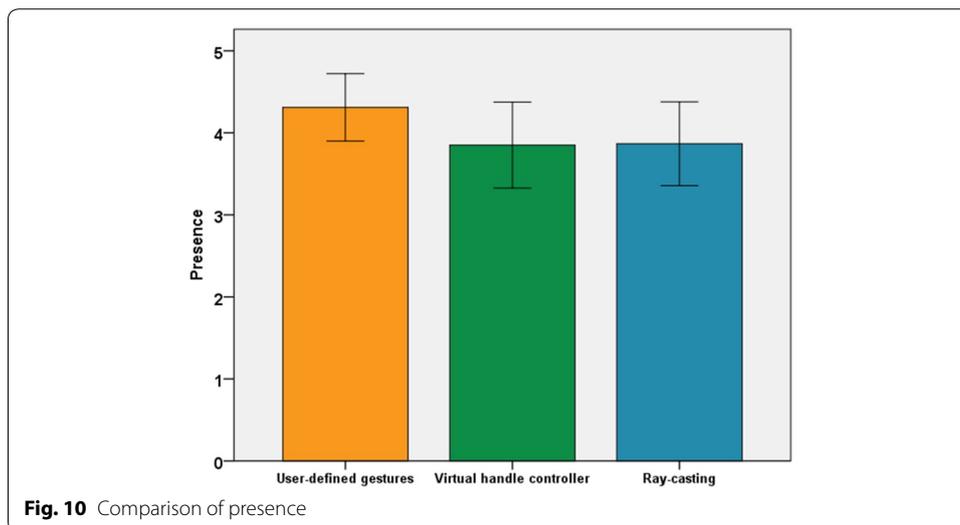
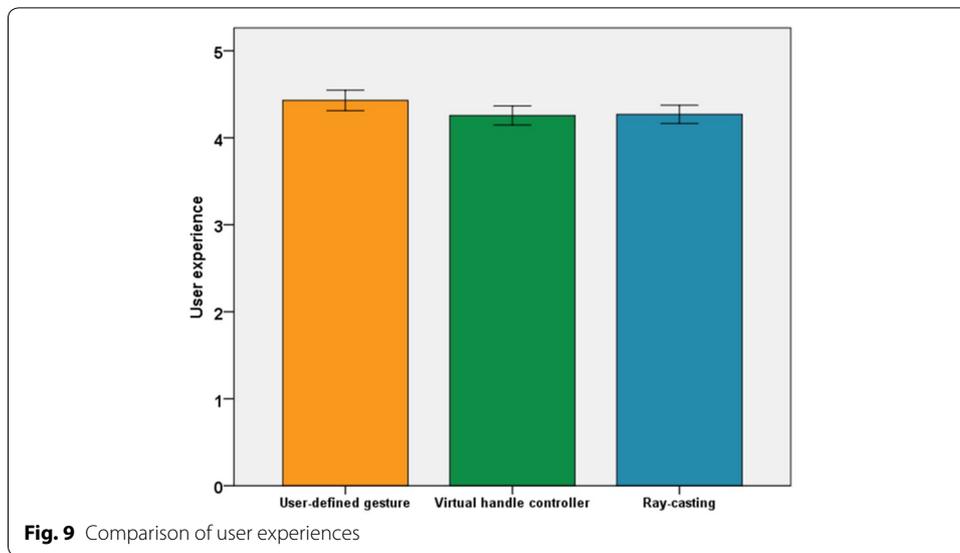
Figure 8 compares the task loads for the three treatments. As shown, the average scores for the task load for user-defined gestures, the *virtual handle controller*, and the *ray-casting* technique are 2.60 ($SD = 1.376$), 2.30 ($SD = 1.302$), and 2.84 ($SD = 1.274$), respectively. Using a Friedman test, we found that the differences in task loads between the three treatments were significant ($\chi^2(2, n = 30) = 10.807, p = 0.005$). In general, the *ray-casting* technique had a significantly heavier task load than user-defined gestures ($p = 0.026$) and the *virtual handle controller* ($p = 0.002$). No significant difference was found between user-defined gestures and the *virtual handle controller* ($p = 0.393$).

For mental demands, user-defined gestures, the *virtual handle controller*, and the *ray-casting* technique averaged 2.48 ($SD = 1.455$), 2.45 ($SD = 1.404$), and 3.41 ($SD = 1.593$), respectively, which indicated there were significant differences between the three techniques ($\chi^2(2, n = 30) = 13.802, p = 0.001$). The *ray-casting* technique requires much more mental and perceptual effort than user-defined gestures ($p = 0.015$) and the *virtual handle controller* ($p = 0.004$). No significant difference was found between user-defined gestures and the *virtual handle controller* ($p = 0.646$).

For physical demands, user-defined gestures, the *virtual handle controller*, and the *ray-casting* technique averaged 2.79 ($SD = 1.590$), 2.69 ($SD = 1.692$), and 3.34 ($SD = 1.495$), respectively, which indicated there were significant differences between the three techniques ($\chi^2(2, n = 30) = 7.386, p = 0.025$). The *ray-casting* technique requires much more physical effort than the *virtual handle controller* ($p = 0.022$). No significant difference was found between user-defined gestures and the *virtual handle controller* ($p = 0.511$) or between user-defined gestures and the *ray-casting* technique ($p = 0.101$).

User experience

Figure 9 compares the user experience for the three treatments. As shown, the average scores of user experience for user-defined gestures, the *virtual handle controller*, and the *ray-casting* technique are 4.43 ($SD = 0.309$), 4.26 ($SD = 0.289$), and 4.27 ($SD = 0.277$), respectively. Using a Friedman test, we found that the differences in user experience



between the three treatments were significant ($\chi^2(2, n = 30) = 7.649, p = 0.022$). In general, user-defined gestures provide a significantly better user experience than the *virtual handle controller* ($p = 0.005$) and the *ray-casting* technique ($p = 0.004$). No significant difference was found between the *virtual handle controller* and the *ray-casting* technique ($p = 0.681$).

Presence

Figure 10 compares the presence of the three treatments. As shown, the average scores for presence for user-defined gestures, the *virtual handle controller*, and the *ray-casting* technique are 4.31 ($SD = 1.080$), 3.85 ($SD = 1.377$), and 3.87 ($SD = 1.343$), respectively. Using a Friedman test, we found that the differences in presence between the three treatments were significant ($\chi^2(2, n = 30) = 11.065, p = 0.004$). In general, user-defined

Table 4 Comparison between Nanjappan et al. and Speicher et al. studies and our work

	Number of subjects	Number of tasks	Interactive technology	Development method	Speed (s)	Error counts
Speicher et al. [4]	16	1	Handle controller	Expert-defined	16.94 (beam) 17.61 (grab)	0 (sphere) 0.24 (basket)
Nanjappan et al. [11]	12	17	Handle controller	User-defined	N/A	N/A
Our work	30	10	Freehand gestures	User-defined	31.9 (gesture) 30.1(virtual handle controller) 36.4 (ray-casting)	0.4 (gesture) 0.5 (virtual handle controller) 0.4 (ray-casting)

gestures provide a significantly higher sense of presence than the *virtual handle controller* ($p=0.004$) and the *ray-casting* technique ($p=0.009$). No significant difference was found between the *virtual handle controller* and the *ray-casting* technique ($p=0.793$).

Comparison with previous work

We compared our work with two recent studies by Speicher et al. [4] and Nanjappan et al. [11]. We chose them because they focused on areas of particular relevance to our research. Table 4 shows the comparison results.

1. Speicher et al. [4] designed two interactive techniques (beam and grab) for participants to select 3D objects and then add to a shopping cart in a virtual shopping environment. They reported that participants could complete the specified task by using the two techniques with no significant difference. In contrast, the error rate varied significantly when the shopping cart represented as different forms (a basket or a sphere). However, Speicher et al.'s system was designed and developed from the perspective of professional designers, which may suffer from the disagreement problem [10, 29] and lead to a lower system usability and user satisfaction.
2. Compared with Speicher et al. [4], Nanjappan et al. [11] adopted an elicitation method to design dual-hand controller interactions from end-users for 17 tasks rather than simply using the technologies developed by professional system designers. They suggested that user-elicitation interactions are more natural and intuitive for manipulating 3D objects in a virtual environment. However, in Nanjappan et al. study, participants were still equipped with handle controllers in two hands. Compared to the freehand interaction mode in real stores, the limited interactivity of such input techniques may impair the workload of end-users and the user experience. In addition, Nanjappan et al. study halted at the stage of definition of dual-hand controllers and lack the further validation of the system performance and end-users' preferences in practice.
3. In contrast to Speicher et al. [4] and Nanjappan et al. [11], we used a human-centered method to derive freehand gestures from end-users. The user-defined gestures were used to interacting with an immersive VR shopping application directly. Experimental results show that the users can complete the specified tasks efficiently and accurately with higher satisfaction and lower cognitive load compared with traditional

handle controller techniques that were used in Speicher et al. [4] and Nanjappan et al. [11].

Discussion of Study 2

Traditional gesture elicitation studies mostly halted at the stage of defining the gestures for specific domains [10, 26, 31, 33, 37, 40]. Consequently, there is a lack of further validation of system performance and end-users' preferences in practice. In this experiment, we compared the proposed user-defined gesture set with two other popular input techniques commonly used in commercial immersive VR systems; i.e., the *virtual handle controller* and *ray-casting* techniques. The experimental results indicate that user-defined gestures allow users to interact with the VR shopping environment easily and intuitively as well as offer improved user experience and user satisfaction from several perspectives.

In general, the average time for user-defined gestures and the *virtual handle controller* to complete the ten common shopping tasks is significantly less than that for the *ray-casting* technique. The *ray-casting* technique outperforms the other two approaches only for Task 1—Select an object, because, with the “magic” laser beam, the *ray-casting* technique allows the user to select and manipulate virtual objects beyond their normal area of reach. In contrast, both user-defined gestures and the *virtual handle controller* technique require the user to adjust the distance and/or angle in the 3D virtual environment to accurately grasp a remote object. However, when interacting with high accuracy, e.g., rotating the red bag 180° along the *y*-axis, the *ray-casting* technique might be less efficient compared to user-defined gestures and the *virtual handle controller*. For object manipulation tasks, such as rotation, enlargement, and shrinking, user-defined gestures and the *virtual handle controller* are easier to perform than the *ray-casting* technique because both of these techniques were implemented based on the concept of a *virtual hand* metaphor [50] with virtual representations of their counterparts in the real world, and a user can efficiently manipulate virtual objects without too much cognitive effort.

The results of the error count metric indicated that no significant difference was found among the three techniques, and all 30 participants successfully finished the ten target tasks in the three treatments.

For the subjective rating, using user-defined gestures for shopping tasks in an immersive VR environment was rated as requiring significantly less mental and physical effort compared to the *ray-casting* technique. In addition, user-defined gestures were thought to provide better user experience and a higher sense of presence than the *virtual handle controller* and *ray-casting* technique. In general, higher user satisfaction maybe because we used an isomorphic metaphor in our gestural system in which the mapping between the user's real hand and the virtual hand is one-to-one and the movements of the virtual hand correspond to the real hand movements. Compared to the *virtual handle controller* and *ray-casting* technique, freehand gesture interaction is more natural and subsequently increases the feeling of presence because of its intuitiveness and familiarity.

In addition, the gestures tested in this study were derived from the previous elicitation study. The gestures invented by participants and designers involve no complex configurations or movements, and they are consistent with participants' mental models and

interactive habits developed in the physical world. Therefore, the naturalness, intuitiveness, and the originality of the idea for using freehand gestures to interact with immersive VR shopping applications may contribute to the perceived benefits.

Implications for freehand gesture design in immersive VR shopping environments

Combining the results of the two user studies, we derived several guidelines for gestural interaction for shopping tasks in immersive VR environments.

Different gesture vocabularies may be obtained by running independent elicitation studies

Involving end-users in the design of gesture-based interactions by analyzing the application's functionalities and users' requirements has become more common in gesture elicitation studies. However, designers should remember that in traditional gesture elicitation studies, the freedom end-users have to produce their gestures for a system inevitably results in certain challenging problems, such as gesture disagreement [28] and legacy bias [37]. These problems may cause different researchers to obtain different gesture vocabularies by running independent elicitation studies. For example, the experimental results of our first study indicated that, without any restriction, the chance for different groups of participants to produce the same top gestures for ten given shopping tasks is 50%. These findings imply that it is unrealistic to expect that one can obtain the same set of user-defined gestures for the same set of specified system tasks by running independent elicitation studies.

Eliciting gestures from ordinary end-users in the a priori stage and selecting gestures with professional designers in the a posteriori stage

Given limited time and experimental conditions, it is unrealistic to expect the participants to design the most appropriate gesture for a given task every time. In addition, the participants do not have professional knowledge of gesture-recognition performance; therefore, they usually focus more on usability metrics such as discoverability, learnability, and memorability rather than metrics such as identifiability and high recognition accuracy required for a gestural system in a gesture elicitation procedure. All of these factors may lead elicitation studies to become stuck in the local minima and fail to identify the most appropriate gestures for specified target tasks. Compared to standard gesture elicitation studies, we emphasize the co-design procedure for designers and end-users to refine and evaluate the resulting user-defined gestures in practice. According to our data, we suggest that an elicitation study is not about creating an absolute set of freehand gestures but to give system designers some ideas of potential good gestures. For those gestures with high agreement from the two elicitation studies such as *Grab*, *Twist*, *Drag onto one's body*, *Drag onto an imagined shopping cart icon*, and *Swipe away*, designers can immediately assign them to the corresponding tasks of Task 1—Select an object, Task 2—Rotate an object, Task 3—Try on clothes, Task 9—Add to a shopping cart, and Task 10—Close the current window, respectively. However, for the other tasks, such as Tasks 4, 5, 6, 7, and 8, designers should be involved to resolve conflicts based on their professional skills.

Including a practical evaluation of the user-defined gesture vocabulary in gesture elicitation studies may help users better understand the design space of gestural interaction

Most standard gesture elicitation studies halted at the stage of gesture design for specific target tasks [10, 26, 31, 33, 37, 40, 51]. However, the gestures derived from participants in those studies were often easy to recall, which does not necessarily guarantee their popularity and usability in practice. In this study, we suggest including a practical evaluation of the user-defined gesture vocabulary in gesture elicitation studies. In this manner, we hope to reduce the limitations of traditional gesture elicitation studies and help end-users better understand the design space of gestural interaction. Indeed, the results of the second experiment in our study indicate that the freehand-gesture-based interaction technique was considered to be the best regarding task load, user experience, and presence without the loss of performance (i.e., speed and error count) compared to traditional VR interaction techniques such as the *virtual handle controller* and *ray-casting* techniques.

Maximizing benefits and minimizing shortcomings in freehand gesture design for VR shopping

It is important to take full advantage and avoid the shortcomings of freehand-gesture-based interaction techniques in immersive VR shopping environments. According to the results of our second study, the main strength of the freehand-gesture-based interaction technique is the naturalness, intuitiveness, high efficiency, and multivariant characteristics in contrast to the *virtual handle controller* and *ray-casting* techniques. The freehand-gesture-based interaction technique provides a one-to-one mapping between real and virtual hands. In addition, it does not require any confirmation trigger or delimiter for virtual object manipulation like the *virtual handle controller* and *ray-casting* techniques. Therefore, participants can select and manipulate the target object in virtual environments using their experience of reaching for and manipulating an object in real life, and the approach has the potential to create a novel shopping experience that combines the advantages of e-commerce sites and conventional retail stores. In contrast, the *ray-casting* technique can overcome the limitations of the freehand-gesture-based technique in the tracking space or anatomical constraints as well as use a nonlinear mechanism to allow the user to select and manipulate virtual objects using a “supernatural” metaphor. As one participant stated:

I hope that the virtual arm can “grow” when I need to access and manipulate remote objects beyond my reach.

Therefore, the results suggest there is potential to explore possible ways to integrate various techniques into seamless and intuitive interaction dialogues by leveraging faster and more accurate object manipulation as well as isomorphic freehand gestures and the accuracy of remote object selection using nonisomorphic ray-casting, when necessary.

Conclusion

Freehand-gesture-based interfaces in interactive systems are becoming increasingly popular. Freehand-gesture-based interaction design allows end-users to directly control the information space in physical space with two hands, which provides end-users with more interaction freedoms, a larger interaction space, and more lifelike interactive

experiences. In this paper, we conducted a two-stage experimental study for exploring freehand gestural interaction in immersive VR shopping applications. The main contributions of our work include the following:

- The proposal of a more practical method for deriving more reliable gestures than traditional gesture elicitation studies;
- the quantitative and qualitative characterization of user-defined gestures for shopping tasks in an immersive VR environment, including a gesture taxonomy;
- New empirical evidence for the benefits of practices involving gestural interaction in immersive VR shopping systems;
- insight into end-users' mental models and shopping behaviors when making free-hand gestures; and
- An understanding of the implications of freehand-gesture-based interaction technology and user interface design.

There are some limitations to our study. One limitation is that our study only concerns virtual object manipulation tasks. To further generalize our findings, additional research is needed to investigate the usability of freehand gestures in other conditions for VR shopping tasks, e.g., the combination of manipulation and navigation tasks. Another interesting next step is to explore how users are affected by the increased user satisfaction provided by gesture-based immersive VR shopping and consequently tend to make more purchases at one time or make repeat purchases in practice.

Acknowledgements

The authors would like to thank the anonymous reviewers for their insightful comments.

Authors' contributions

HYW framed the ideas of the paper and the design of the study. WZL and NP developed the prototype and run the experiment to collect the data with the help of SHN, YYD, SQF, and LQQY. All authors analyzed the results. All authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China under Grant No. 61772564, 61202344 and the funding offered by the China Scholarship Council (CSC).

Competing interests

The authors declare that they have no competing interests.

Author details

¹The School of Communication and Design, Sun Yat-sen University, Guangzhou, China. ²Guangdong Key Laboratory for Big Data Analysis and Simulation of Public Opinion, Guangzhou, China. ³The School of Atmospheric Sciences, Sun Yat-sen University, Guangzhou, China.

Received: 31 July 2019 Accepted: 3 December 2019

Published online: 10 December 2019

References

1. Chen T, Pan ZG, Zheng JM (2008). EasyMall—an interactive virtual shopping system. In: 5th international conference on fuzzy systems and knowledge discovery. 4. pp. 669–673
2. Zhao L, Zhang N (2012). The virtual reality systems in electronic commerce. In: IEEE symposium on robotics and applications. pp. 833–835
3. Speicher M, Cucerca S, Krüger A (2017). VRShop: a mobile interactive virtual reality shopping environment combining the benefits of on- and offline shopping. In: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies. 1(3). pp. 1–31
4. Speicher M, Hell P, Daiber F, Simeone A, Krüger A (2018). A virtual reality shopping experience using the apartment metaphor. In: Proceedings of the international conference on advanced visual interfaces. pp. 1–9
5. Sanna A, Montrucchio B, Montuschi P, & Demartini C (2001). 3D-dvshop: a 3D dynamic virtual shop. In: Multimedia. pp. 33–42

6. Cardoso LS, da Costa RMEM, Piovesana A, Costa M, Penna L, Crispin A, Carvalho J, Ferreira H, Lopes M, Brandao G, Mouta R (2006). Using virtual environments for stroke rehabilitation. In: International workshop on virtual rehabilitation. pp. 1–5
7. Josman N, Hof E, Klinger E, Marié RM, Goldenberg K, Weiss PL, Kizony R (2006). Performance within a virtual supermarket and its relationship to executive functions in post-stroke patients. In: International workshop on virtual rehabilitation. pp. 106–109
8. Carelli L, Morganti F, Weiss P, Kizony R, Riva G (2008). A virtual reality paradigm for the assessment and rehabilitation of executive function deficits post stroke: feasibility study. In: IEEE virtual rehabilitation. pp. 99–104
9. Josman N, Kizony R, Hof E, Goldenberg K, Weiss P, Klinger E (2014) Using the virtual action planning-supermarket for evaluating executive functions in people with stroke. *J Stroke Cerebrovasc Dis* 23(5):879–887
10. Wu HY, Wang Y, Qiu JL, Liu JY, Zhang XL (2018) User-defined gesture interaction for immersive VR shopping applications. *Behav Inf Technol*. <https://doi.org/10.1080/0144929X.2018.1552313>
11. Nanjappan V, Liang HN, Lu FY, Papangelis K, Yue Y, Man KL (2018) User-elicited dual-hand interactions for manipulating 3D objects in virtual reality environments. *Hum Comput Inf Sci* 8(31):1–16
12. Song P, Goh WB, Hutama W, Fu CW, Liu XP (2012). A handle bar metaphor for virtual object manipulation with mid-air interaction. In: Proceedings of the SIGCHI conference on human factors in computing systems. pp. 1297–1236
13. Ren G, O'Neill E (2013) 3D selection with freehand gesture. *Comput Graph* 37:101–120
14. Feng ZQ, Yang B, Li Y, Zheng YW, Zhao XY, Yin JQ, Meng QF (2013) Real-time oriented behavior-driven 3D freehand tracking for direct interaction. *Pattern Recogn* 46:590–608
15. Alkemade R, Verbeek FJ, Lukosch SG (2017) On the efficiency of a VR hand gesture-based interface for 3D object manipulations in conceptual design. *Int J Hum-Comput Int* 33(11):882–901
16. Cui J, Sourin A (2018) Mid-air interaction with optical tracking for 3D modeling. *Comput Graph* 74:1–11
17. Figueiredo L, Rodrigues E, Teixeira J, Teichrieb V (2018) A comparative evaluation of direct hand and wand interactions on consumer devices. *Comput Graph* 77:108–121
18. Tollmar K, Demirdjian D, Darrell T (2004). Navigating in virtual environments using a vision-based interface. In: Proceedings of the third Nordic conference on Human-computer interaction. pp. 113–120
19. Sherstyuk A, Vincent D, Lui JH, Connolly KK (2007). Design and development of a pose-based command language for triage training in virtual reality. In: IEEE symposium on 3D user interfaces. pp. 33–40
20. Verhulst E, Richard P, Richard E, Allain P, Nolin P (2016). 3D interaction techniques for virtual shopping: design and preliminary study. In: International conference on computer graphics theory and applications. pp. 271–279
21. Kölsch M, Turk M, Höllerer T (2004). Vision-based interfaces for mobility. In: Mobile and ubiquitous systems: networking and services. pp. 86–94
22. Colaco A, Kirmani A, Yang HS, Gong NW, Schmandt C, Goyal VK (2013). Mine: compact, low-power 3D gesture sensing interaction with head-mounted displays. In: Proceedings of the 26th annual ACM symposium on user interface software and technology. pp. 227–236
23. Ohta M, Nagano S, Takahashi S, Abe H, Yamashita K (2015). Mixed-reality shopping system using HMD and smartwatch. In: Adjunct proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing and proceedings of the 2015 ACM international symposium on wearable computers. pp. 125–128
24. Badju A, Lundberg D (2015). Shopping using gesture driven interaction. Master's Thesis. Lund University. pp. 1–105
25. Altarteer S, Charissis V, Harrison D, Chan W (2017). Development and heuristic evaluation of semi-immersive hand-gestural virtual reality interface for luxury brands online stores. In: International conference on augmented reality, virtual reality and computer graphics. pp. 464–477
26. Chan E, Seyed T, Stuerzlinger W, Yang XD, Maurer F (2016). User elicitation on single-hand microgestures. In: Proceedings of the SIGCHI conference on human factors in computing systems. pp. 3403–3411
27. Choi S (2016) Understanding people with human activities and social interactions for human-centered computing. *Hum Comput Inf Sci* 6(9):1–10
28. Wu HY, Zhang SK, Qiu JL, Liu JY, Zhang XL (2018) The gesture disagreement problem in freehand gesture interaction. *Int J Hum-Comput Inter*. <https://doi.org/10.1080/10447318.2018.1510607>
29. Furnas GW, Landauer TK, Gomez LM, Dumais ST (1987) The vocabulary problem in human-system communication. *Commun ACM* 30(11):964–971
30. Morris MR, Wobbrock JO, Wilson AD (2010). Understanding users' preferences for surface gestures. In: Proceedings of graphics interface. pp. 261–268
31. Wobbrock JO, Morris MR, Wilson AD (2009). User-defined gestures for surface computing. In: Proceedings of the SIGCHI conference on human factors in computing systems. pp. 1083–1092
32. Kray C, Nesbitt D, Rohs M (2010). User-defined gestures for connecting mobile phones, public displays, and tabletops. In: Proceedings of the 12th international conference on human computer interaction with mobile devices and services. pp. 239–248
33. Ruiz J, Li Y, Lank E (2011). User-defined motion gestures for mobile interaction. In: Proceedings of the SIGCHI conference on human factors in computing systems. pp. 197–206
34. Shimon SSA, Lutton C, Xu ZC, Smith SM, Boucher C, Ruiz J (2016). Exploring non-touchscreen gestures for smartwatches. In: Proceedings of the SIGCHI conference on human factors in computing systems. pp. 3822–3833
35. Peshkova E, Hitz M, Ahlström D, Alexandrowicz RW, Kopper A (2017). Exploring intuitiveness of metaphor-based gestures for UAV navigation. In: 26th IEEE international symposium on robot and human interactive communication (RO-MAN). pp. 175–182
36. Gheran BF, Vanderdonckt J, Vatavu RD (2018). Gestures for smart rings: empirical results, insights, and design implications. In: ACM SIGCHI conference on designing interactive systems. pp. 623–635
37. Morris MR, Danielescu A, Drucker S, Fisher D, Lee B, Schraefel MC, Wobbrock JO (2014) Reducing legacy bias in gesture elicitation studies. *Interactions*. 21(3):40–45
38. Seyed T, Burns C, Sousa MC, Maurer F, Tang A (2012). Eliciting usable gestures for multi-display environments. In: Proceedings of the 2012 ACM international conference on interactive tabletops and surfaces. pp. 41–50

39. Tung YC, Hsu CY, Wang HY, Chyou S, Lin JW, Wu PJ, Valstar A, Chen MY (2015). User-defined game input for smart glasses in public space. In: Proceedings of the SIGCHI conference on human factors in computing systems. pp. 3327–3336
40. Hoff L, Hornecker E, Bertel S (2016). Modifying gesture elicitation: Do kinaesthetic priming and increased production reduce legacy bias? In: Proceedings of the tenth international conference on tangible, embedded, and embodied interaction. pp. 86–91
41. Jo D, Kim GJ (2019) lot + AR: pervasive and augmented environments for “Digi-log” shopping experience. *Hum Comput Inf Sci* 9(1):1–17
42. Wu HY, Wang JM, Zhang XL (2017) Combining hidden Markov model and fuzzy neural network for continuous recognition of complex dynamic gestures. *Visual Computer*. 33(10):1227–1263
43. Vatavu RD, Wobbrock JO (2015). Formalizing agreement analysis for elicitation studies: new measures, significance test, and toolkit. In: Proceedings of the SIGCHI conference on human factors in computing systems. pp. 1325–1334
44. Wu HY, Wang JM (2016) A visual attention-based method to address the Midas touch problem existing in gesture-based interaction. *Visual Computer*. 32(1):123–136
45. Montero CS, Alexander J, Marshall M, Subramanian S (2010). Would you do that?—Understanding social acceptance of gestural interfaces. In: Proceedings of the 12th international conference on human computer interaction with mobile devices and services. pp. 275–278
46. Wu HY, Wang JM, Zhang XL (2016) User-centered gesture development in TV viewing environment. *Multimedia Tools Appl* 75(2):733–760
47. Hart SG, Staveland LE (1988) Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. *Adv Psychol* 52:139–183
48. Lund AM (2001) Measuring usability with the USE Questionnaire. *SIG Newslett* 8:2
49. Regenbrecht H, Schubert T (2002) Real and illusory interaction enhance presence in virtual environments. *Pres Teleoper Virtual Environ* 11(4):425–434
50. Bowman DA, Kruijff E, LaViola J, Poupyrev I (2004) 3D user interfaces: theory and practice. Addison Wesley Longman Publishing Co., Inc, Redwood City
51. Chen Z, Ma XC, Peng ZY, Zhou Y, Yao MG, Ma Z, Wang C, Gao ZF, Shen MW (2018) User-defined gestures for gestural interaction: extending from hands to other body parts. *Int J Hum–Comput Inter* 34(3):238–250

Publisher’s Note

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

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
