




Virtual object sizes for efficient and convenient mid-air manipulation

Myoung Gon Kim¹ · JiSeok Ryu¹ · Jaemin Son¹ · JungHyun Han¹ 

Accepted: 21 May 2022 / Published online: 1 July 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

It has been taken for granted that the sizes of virtual objects affect the efficiency and convenience of mid-air manipulation in immersive virtual environments. If a virtual object is too small or too large, for example, manipulating it becomes a difficult task. Nevertheless, the virtual object sizes that are optimal and convenient have rarely been studied. In this paper, we select a virtual object with many distinct geometric features and conduct user studies via docking tasks. Through the user studies, the optimal and convenient sizes for mid-air manipulation are estimated. In order to verify the results, a proxy-based manipulation method is designed and implemented, where the proxy is created with the estimated optimal size. The test based on the method shows that the optimal-size proxy enables users to manipulate efficiently virtual objects and the estimated range of convenient sizes is also preferred by the users.

Keywords Mid-air manipulation · Virtual reality · Human–computer interaction

1 Introduction

In immersive virtual environments (IVE), *mid-air manipulation* has been a popular research topic. Its goal is to provide a natural way of handling 3D objects in IVE. There are many factors that affect the efficiency and convenience of mid-air manipulation. They include degrees of freedom [20], virtual object shapes [15] and sizes [30]. With respect to the object sizes, it has been discovered that the performance of manipulation tasks decreases significantly if virtual objects are too small or too large [8,35,38].

One of the techniques widely used to overcome the size problem is *zoom* [2,11,34], but the artifacts caused by zoom have also been recognized well. To list a few, they include mismatch between the physical and scaled virtual worlds [28], misperception of scale [9] and loss of perspective [40].

On the other hand, virtual objects can be manipulated in *proxy-based methods* [25,26], where the virtual objects are provided as proxies of “convenient working sizes.” Unfortunately, the sizes were selected heuristically in the previous works. To the best of the authors’ knowledge, no studies have been made on the convenient working sizes.

This paper reports the results of our efforts to investigate the object sizes that are optimal and convenient for mid-air manipulation. Using a well-known mid-air manipulation method, the optimal size of a reasonably complex object (Utah teapot) was estimated through an experiment. Through another experiment with zoom operation, a range of convenient working sizes was also estimated. Finally, the optimal and convenient working sizes were verified via an additional experiment based on a new proxy-based method. We believe our studies give a guideline for advanced studies on the object size problems in IVE.

✉ JungHyun Han
jhan@korea.ac.kr

Myoung Gon Kim
m_gon_kim@korea.ac.kr

JiSeok Ryu
roadmageb@korea.ac.kr

Jaemin Son
jasson31@korea.ac.kr

¹ Department of Computer Science and Engineering, Korea University, Seoul, South Korea

2 Related work

Various mid-air manipulation techniques have been proposed in order to manipulate virtual objects with accuracy and speed [18]. The most intuitive ones are the 6DOF techniques [3,12,16,36], since they mimic the way humans manipulate a single object in the real world. The PRISM technique [6,7] increases the accuracy of 6DOF manipulation by scaling down the translation and rotation of objects. In addi-

tion, symmetric-synchronous techniques such as Handle-bar metaphor [32] and Spindle \times Wheel [5] allow manipulating multiple objects by grabbing the objects with both hands. For example, the global translation/rotation of the objects is enabled by simultaneously moving both hands in the same/different direction.

On the other hand, some researchers focused on the comparisons between the proposed techniques as well as their evaluations. For example, Rodrigues et al. [31] compared and evaluated the 6DOF techniques and the Handle-bar metaphor, whereas Mendes et al. [19] compared different manipulation methods using a stereoscopic tabletop-based approach and mid-air bi-manual methods. They also evaluated the benefits of adopting DOF separation in mid-air manipulation against direct manipulation [20].

There have been several papers that address the manipulation of distant objects, as the visual size of the objects may vary in the perspective view. For example, Whitlock et al. [37] evaluated the effects of different modalities. The Go-Go technique [29] manipulates distant objects by introducing a nonlinear mapping between the physical hand and the virtual hand in order to extend the user's reach in the virtual environment. The hybrid interaction technique named as HOMER [3] uses ray-casting for ease of selecting and manipulating distant objects. Mine et al. [22] introduced the concept of scaled world grab, which scales up the world to deal with the relatively resized object.

In accordance with [30], sizes of virtual objects have significant effects on mid-air manipulation. For example, Scaled HOMER [38] revealed significant interactions between techniques and target sizes. Laviola et al. [14] have found that it is difficult to select and manipulate virtual objects of different sizes not only from multiple angles but also at different distances to the user's virtual avatar, which requires significant time and efforts on navigation. Selecting virtual objects in IVE is studied, and its elapsed time is also measured with varying conditions of occlusion, size and density [35]. In this study, it has been observed that the elapsed selecting time for small objects is usually doubled than that for large objects. PinNPivot [8] which constrains rotation of virtual objects has discovered that positional error of large object is greater than that of small object.

Zoom techniques, which resize a user [1], have been also widely studied. GulliVR [13], for example, adjusted the depth perception during scaling user by resizing IPD to match the player's size in order to prevent cybersickness. The technique which allows to rotate, scale and translate the virtual scene about the user reduces the need of physical locomotion [21]. The EiHCam [17] uses an extra-camera in immersive environments, which can be scaled by a pinch-gesture. Wang et al. [34] provided zoom using pinch gesture in complex level editing tasks. MakeVR [11] enabled users to translate, zoom and rotate in real space rather than on a

flat display. The comparison between a zooming interface and room-sized visualization has been conducted with and without an overview [40].

The methods of a proxy of virtual environments have been widely researched to provide interaction techniques such as object selecting, navigating, path planning, and visualization. Worlds-In Miniature [33] demonstrated the remote manipulation technique, providing small copies of the world held within the hands. Voodoo Dolls [26] creates an instant copy of the object, resizes the copy into comfortable working sizes, and attach the object to the user's hand. The image plane interaction technique [25] deals with 3D distant objects by interacting with 2D projections of the objects on the image plane. GARDEN [24] was designed to increase the accuracy of manipulation of distant objects in a shared environment by using sphere casting. Remote Collaboration presents a similar system which provides the remote user with the ability to produce and manipulate 3D objects to be assisted by communication [4,23]. Recently, Poros [27], a system that allows users to rearrange distant space after marking a portion of the space, has been proposed. Magnoramas [41] presents a solution for the simultaneous view on the virtual visualization of the real-world and a user-controlled copy of a region of interest.

3 Mid-air manipulation

Our study focuses on manipulating 3D rigid objects through translations and rotations in IVE. For experiments, we implemented a well-known manipulation technique named *6DOF method* [36].

3.1 6DOF method

As the name says itself, 6DOF method supports 6DOF motions which are comprised of 3DOF translations and 3DOF rotations. It mimics interactions with physical objects, i.e., it consists of grabbing an object, moving it to a new location, and releasing it. Dragging an object changes the object's position and the wrist's rotation changes its orientation. Throughout the entire manipulation, both translation and rotation are applied to the object at the same time. See Fig. 1. When the user touches an object, its wireframe bound-

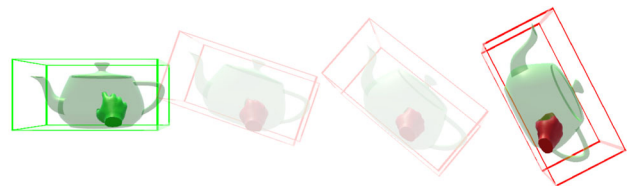


Fig. 1 Utah teapot manipulated with 6DOF method

ing box appears in green, and the virtual hand touching the object also turns into green. The colors turn into red when the user grabs the object. After being grabbed, the object follows directly the movement of the hand. Henceforth, the 6DOF method is denoted simply as **6DOF**.

3.2 Supportive operations

In our experiments, two operations are supported for mid-air manipulation.

3.2.1 Grab-navigation

In our experiments, subjects are often asked to manipulate a huge object. Unfortunately, it may be larger than a real-world room, making it impossible for the subjects to look around the object or look down on it from high in the air.

In order to tackle the problems, a supportive operation is provided, which we call *grab-navigation*. By grabbing an empty space with a single hand and pulling it along a direction, we can move along the opposite direction. This enables us to navigate the virtual space in any direction we want [10,39].

3.2.2 Zoom

This is used to scale up or down the whole virtual environment [40]. When the user grabs an empty space with both hands and squeezes it, the scene is scaled down so that the user would feel like a giant. In contrast, by stretching the empty space, the scene is scaled up.

4 Experiment setup

We conducted three experiments: (1) pilot study, (2) main experiment, (3) additional experiment for verification. The pilot study was conducted with a small number of participants for the purpose of designing the main experiment. The main experiment was conducted with a larger number of participants. Its goal was to discover the object sizes that are optimal and convenient for mid-air manipulation. Then, using a new proxy-based manipulation method, an additional experiment was made to verify the discovered optimal and convenient sizes.

Oculus Quest 2 was used for all experiments. Its wireless HMD and controllers communicated with a desktop PC over 5G Wi-Fi network. Through the PC screen, the experiment supervisor shared the view displayed at the participants' HMD. In the controllers, only trigger and grip buttons were used. All experiments were pre-approved by the authors' institutional review board.

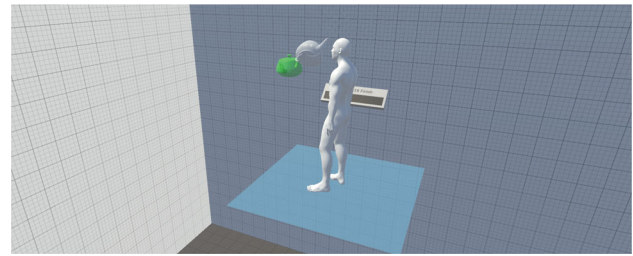


Fig. 2 Virtual workspace used for all experiments

4.1 Virtual workspace

The virtual workspace used for all experiments is shown in Fig. 2. It is a 30 m \times 30 m \times 30 m cubical space, at the center of which a translucent blue square (2 m \times 2 m) mat is located. The subjects stand on the mat. It corresponds to a 'clear' real-world floor. As the subjects are asked not to go out of the mat, it helps them avoid collision with the real-world obstacles.

In order to provide the visual effect of grab-navigation and zoom operations presented in Sect. 3.2, six faces of the cube are all textured with grid patterns, where parallel lines are placed at intervals of 1 m.

4.2 Docking tasks

In all experiments, the subjects are asked to perform *docking tasks* between the same size objects. As shown in Fig. 3, we use Utah teapots. The *source* teapot (in green) should be moved to fit to the *target* teapot (in silver) that is fixed and immovable in the space. The Utah teapot is selected due to its distinct geometric features, such as spout, handle and lid. As discussed by Martin-Gomez et al. [15], such distinct features provide useful cues for orientation.

In the docking tasks, different scales of the teapot were used. See Fig. 4. In terms of the length of the line connecting the spout and handle, the sizes of 0.01 m, 0.05 m, 0.1 m, 0.5 m, 1 m, 3 m, 5 m and 10 m (in increasing order) were used in the pilot study. The minimum, 0.01 m, and the maximum, 10 m, were heuristically determined. While designing the pilot

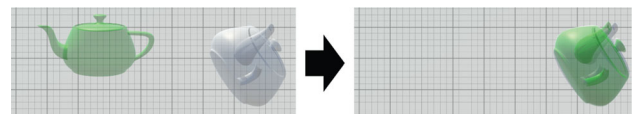


Fig. 3 Docking task between green source and silver target

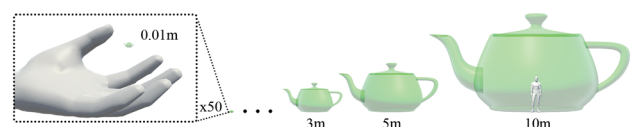


Fig. 4 Eight object scales from 0.01 to 10 m

study, we figured out that objects that are smaller than 0.01 m or larger than 10 m are very difficult or almost impossible to manipulate.

4.3 Object placements

In the docking tasks, the target objects are placed in front of the subject “one at a time” with different poses. Figure 5a shows an imaginary cone, the apex of which is in the middle of the shoulder. The apex angle is 120° . The boundary of the cone’s base is regularly sampled at 45° intervals to make eight sample points.

A target is placed per sample point. Figure 5b shows the steps to place a target. Initially, the target’s center is located at a sample point. It is rotated through Euler transform. Each of the Euler angles, θ_x , θ_y and θ_z , is either $+45^\circ$ or -45° so that the targets have eight distinct orientations in total. Shown in the middle of Fig. 5b is one of them. Then, the target is translated horizontally such that it contacts but does

not penetrate the cone’s base. Each subject’s shoulder height and arm length are measured before the experiments, and all targets are placed within arm’s reach regardless of their sizes.

The source object is in front of the subject at the shoulder’s height. See the green teapot in Fig. 5c. Whereas the source’s size changes over docking tasks, so that it is always the same as the target’s, its orientation is fixed. The source’s position is adjusted to make its surface within arm’s reach. Figure 5d, e shows additional source–target examples.

The source and target appear in front of the subjects with the ‘initial’ size of 0.2 m, as shown in Fig. 6a, and then, they are scaled up or down to the pre-specified scale, e.g., 0.01 m shown in Fig. 6b. (The initial size, 0.2 m, is determined heuristically.) The scaling animation takes 1 s. After scaling is completed, two red circles pop up and disappear for the subjects not to miss the scaled objects. See Fig. 6c.

5 Pilot study

5.1 Participants and procedure

The pilot study was made with five participants, including three authors of this paper. Each subject performed 16 docking tasks, which were grouped into two experimental blocks. In a block, the eight targets were distinct in every aspect of size, position and orientation, as shown in Fig. 5c through Fig. 5e. To mitigate possible bias, the order in which the targets appear was given by Latin squares. By using the pattern of a balanced Latin square, the ordering can be counterbalanced, and the order-effect can be reduced [42].

As can be found in Fig. 6, there exists a button beside the subject. If a subject is satisfied with the result of a docking task, the subject can proceed to the next task by touching the button. We call it ‘finish’ button. In order to avoid unwanted finish that can be made by accidentally touching the button, subjects were asked to keep touching the button for 1 s. See Fig. 7.

5.2 Results and analysis

In each docking task, we measured the *completion time* consumed until the finish button was touched. The statistics for eight different scales are depicted in Fig. 8. At the stage of designing the pilot study, longer completion times were expected for small and large objects, and shorter times were for medium-size ones. However, Fig. 8 does not exactly match the expectation. During the pilot study, we observed that many subjects consumed unnecessarily longer time to achieve higher precision even after the source was almost fit to the target. This suggested the need for establishing a certain error *threshold* such that the docking task can be judged

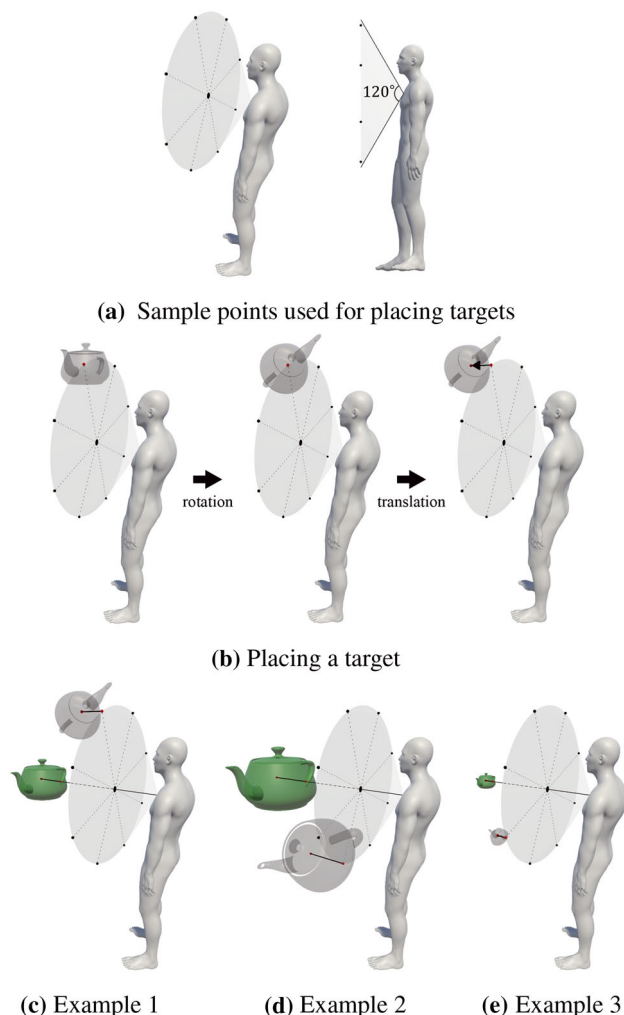
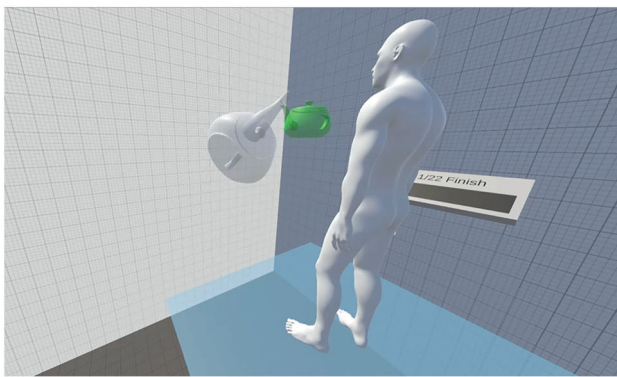


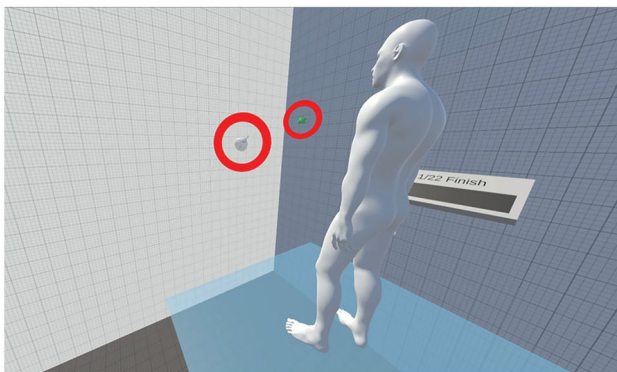
Fig. 5 Poses of source and target objects



(a) Source and target objects appearing in front of the subject have the initial size of 0.2m.



(b) In this example, the objects shrink to 0.01m.



(c) Two red circles indicate the locations of the objects.

Fig. 6 The locations of the source and target objects are highlighted before the docking task is started

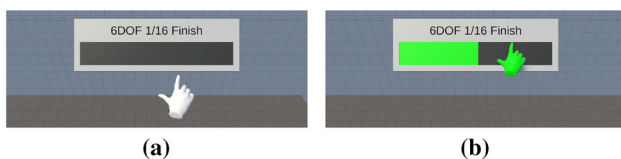


Fig. 7 Finish button. **a** Until touching the button, it remains black. **b** If the subject's hand stays touching the button for a second, the green progress bar starts to grow. It is 100% filled after 1 s. Then, the next docking task is presented to the subject

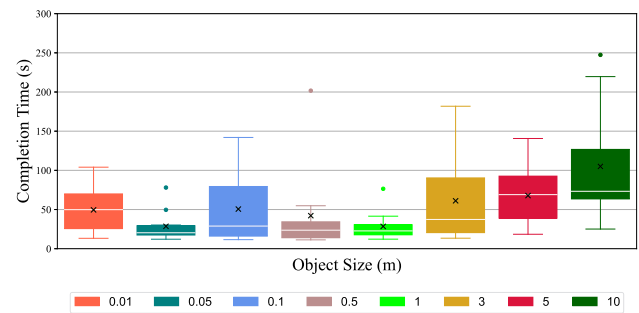
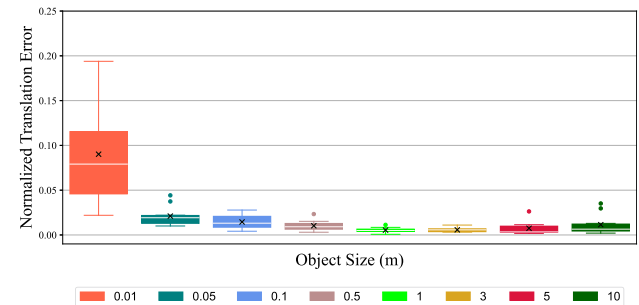
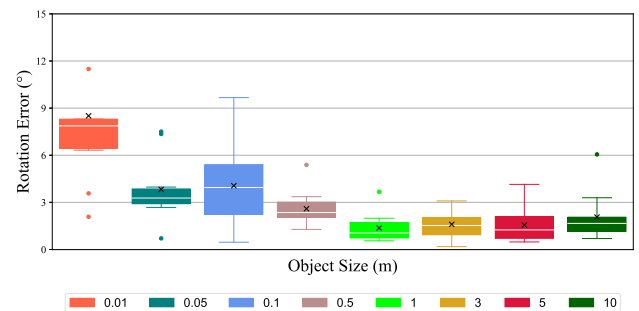


Fig. 8 Completion times in pilot study



(a) Normalized translation errors (NTE)



(b) Rotation errors (RE)

Fig. 9 Normalized translation error and rotation error in pilot study

to be completed once the error between source and target falls below the threshold.

During the experiment, we measured translation and rotation errors. The translation error is defined as the Euclidean distance between the centers of source and target. It is normalized so that for example, 1 m error between 10 m-long objects is identical to 0.5 m error between 5 m-long objects, i.e., the errors equal 0.1. This normalized translation error is denoted as **NTE**. On the other hand, the rotation error denoted as **RE** is defined as the angle between two unit quaternions that represent the orientations of source and target. The measured errors are depicted in Fig. 9.

Through the regression analysis, the ratio between **NTE** and **RE** was computed as 0.008:1. We defined the total error

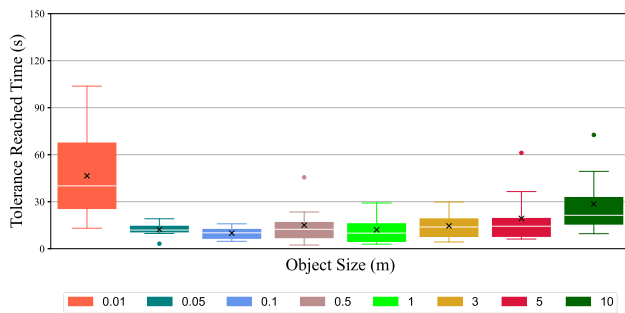


Fig. 10 Completion times recomputed using a threshold

as $\text{NTE} + 0.008 \times \text{RE}$. Then, we heuristically set the threshold of the total error to 0.1.

During the docking tasks, the subjects keep repeating grab and release actions presented in Sect. 3.1. We measured the pose of the source whenever it was released. Using the logged data, the completion time is recomputed, i.e., as soon as the pose error falls below the threshold, the docking task is judged to be completed and the elapsed time is recorded. Figure 10 shows the recomputed completion times. The results are largely compatible with our expectation, showing that the threshold is reasonably set.

Additional observations can be made in Fig. 10. First of all, most docking tasks were completed in 60 s. This led us to introduce the *time limit* for a task in the main experiment, i.e., if a subject fails to complete a docking task in 60 s, the task will be finished, and the completion time is set to 60 s. In the pilot study, there was no time limit, and we observed that subjects often got exhausted at the end of the experiment and tended to finish the tasks even though the objects were not sufficiently fit. This also supported the introduction of time limit.

Secondly, the statistics differ significantly between 0.01 and 0.05 m in Fig. 10. The same can be found in Fig. 9a (for translation errors) and Fig. 9b (for rotation errors). This led us to sample the interval between 0.01 and 0.05 m. We add three more scales, 0.02 m, 0.03 m and 0.04 m, in the main experiment.

In summary, the pilot study contributed to the following features of the main experiment: (1) error threshold, (2) time limit (60 s), (3) additional scales of 0.02 m, 0.03 m and 0.04 m.

6 Main experiment

In the main experiment, the participants performed docking tasks not only with **6DOF** but also with **6DOF (Z)**, which is **6DOF** augmented with the *zoom* operation presented in Sect. 3.2.2. With **6DOF**, the *optimal* object size is identified, which minimizes the completion time. It is the size for

efficiency. In contrast, with **6DOF (Z)**, the *convenient working sizes* (henceforth, simply *convenient sizes*) are identified, which are selected by the participants. All subjects performed **6DOF** ahead of **6DOF (Z)** so that the completion times measured with **6DOF** are free from learning effect.

6.1 Participants and procedure

In the experiment, thirty-three subjects participated: Seven were females and twenty six were males. Their ages ranged from 18 to 35 ($M = 23.12$, $SD = 3.15$). All subjects had normal or corrected-to-normal vision. Only two subjects used the left hand as the dominant hand. Ten subjects did not have experiences of playing 3D games. Ten subjects had no experiences of wearing VR devices such as HMD, and sixteen subjects have not used gesture-based devices or controllers in IVEs.

During the time of this study, COVID-19 spread rampantly. Upon arrival at the building, all participants took their temperatures at the lobby. They were allowed to enter the building only if they did not have a fever. HMD and controllers were disinfected with a sanitizer before and after the experiment.

Before the experiment began, the interpupillary distance (IPD) of each subject was measured for HMD setup. Then, each subject tested **6DOF**. The experimental procedure for **6DOF** is presented in Table 1 with the elapsed time on each step. After watching a tutorial video, the subject had a 10-min practice using three objects shown in Fig. 11. (Neither the tutorial nor the practice was for docking tasks. They were about general 6DOF manipulation.) Then, the rubric for the docking tasks was presented, as shown in Fig. 12. Finally, the subject started the docking tasks by touching the button. It worked in the same way as the ‘finish’ button used for the pilot study (Fig. 7), but it was used for ‘start’ in the main experiment.

As presented in Sect. 5.2, three more scales (0.02 m, 0.03 m and 0.04 m) were added to the eight scales used for the pilot study. For each method, a subject performed 22 docking tasks, which were grouped into two experimental blocks. The orders of size, position and orientation of the targets were given by Latin squares.

Table 1 Experiment procedure for each method

Step	Time (min)
Video tutorial	5
Practice in IVE	10
Rubric presentation	< 1
Docking tasks	*

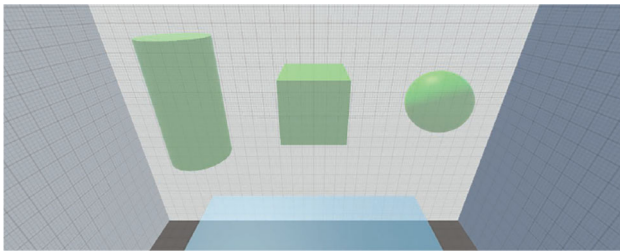


Fig. 11 The objects used for practice in the main experiment

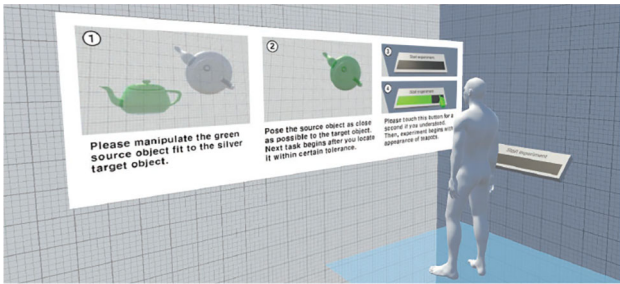


Fig. 12 The rubric presented in the main experiment

Each docking task was finished either when the source was fit to the target within the tolerance or after 60 s passed. The 22 docking tasks may take up to 22 min, as each task is assigned 1 min at maximum. However, the entire docking tasks took 10.44 min on average.

After taking a 10-min break, the procedure presented in Table 1 was resumed with the next method, **6DOF (Z)**, where the 22 docking tasks took 3.91 min on average. The whole experiment with two methods took about an hour, and a subject was offered 30 USD for participation.

6.2 Results and analysis

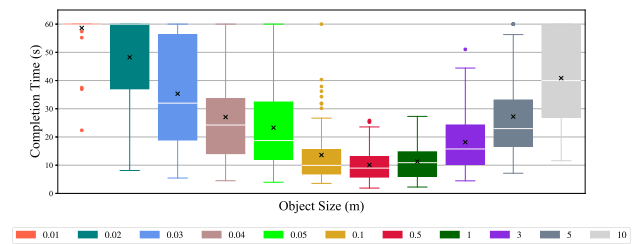
6.2.1 Optimal size

For **6DOF**, the completion times consumed by a subject for finishing 22 docking tasks were measured. Figure 13a shows the completion times for eleven scales in **6DOF**.

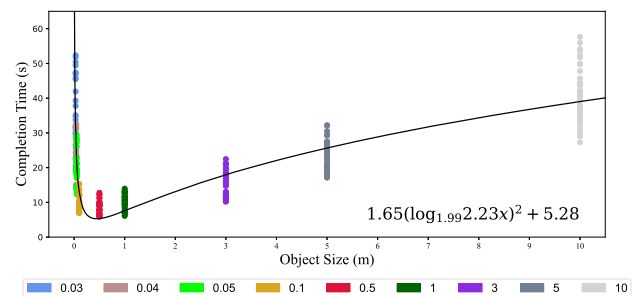
In order to remove outliers, only the data between the first quartile (Q1) and the third quartile (Q3) are considered in each scale, i.e., Q1 and Q3 are the upper and lower bounds of the completion times, respectively.

We need to fit the data between Q1 and Q3 with a function so that its argmin (argument of minimum) is taken as the optimal value. Before fitting, the data are post-processed. For example, many values for 0.01 m and 0.02 m in **6DOF** would be greater than 60 s without the time limit. Such data are unreliable. We abandon the data, Q3 of which reached 60 s.

The remaining reliable data are plotted in Fig. 13b. Note that the completion times are dramatically high for small



(a) Box plots of completion times in **6DOF**



(b) Valid data and curve fitting in **6DOF**

Fig. 13 Completion times of **6DOF** in the main experiment

objects, while they are moderately high for large objects. A one-way analysis of variance (ANOVA) test revealed that the differences between the object sizes in terms of completion times are significant ($F = 42.1$, $p < 0.001$). For the post hoc analysis, we conducted multiple F -tests between the adjacent object sizes. Once an F -test revealed the inequality of variances, a Welch's t -test (unequal variance t -test) was conducted. Otherwise, a t -test assuming equal variances was conducted. The results are presented in Table 2, where asterisks indicate significant differences ($*p < 0.05$, $**p < 0.01$, $***p < 0.001$). Based on this observation, we selected “square of log” function to fit the data:

$$a(\log_b(cx))^2 + d \quad (1)$$

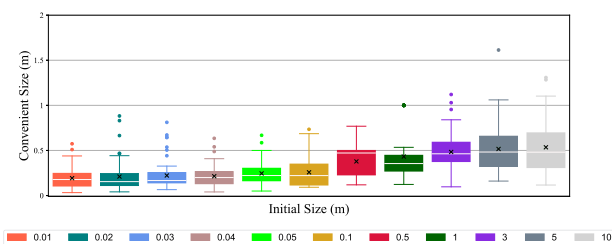
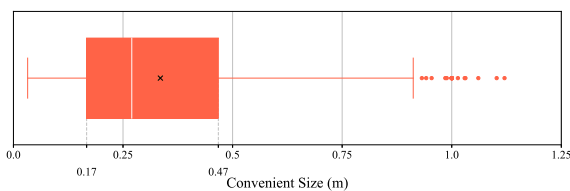
The fitted graph is overlaid on the data in Fig. 13b. It shows that the argmin in **6DOF** is 0.45 m where the minimum completion time is 5.28 s. We take the argmin as the optimal size in **6DOF**.

6.2.2 Convenient size

In **6DOF (Z)**, a subject is allowed to scale up and down the virtual experiment using *zoom* to obtain the “convenient size” of the source and target. If a subject takes the given size as sufficiently convenient, however, no zoom operation will be made. We take the convenient size as the object size at the

Table 2 Results of post hoc analysis between the adjacent object sizes in terms of completion times

	Object sizes							
	0.03–0.04	0.04–0.05	0.05–0.1	0.1–0.5	0.5–1	1–3	3–5	5–10
<i>F</i> -test								
*			*	***		***	**	
<i>F</i>	1.28	1.20	2.03	3.12	1.08	0.33	0.50	0.86
<i>p</i>	> 0.05	> 0.05	< 0.05	< 0.001	> 0.05	< 0.001	< 0.01	> 0.05
<i>t</i> -test								
*	*		***	**		***	***	***
<i>t</i>	2.65	1.35	4.21	2.32	1.19	4.72	4.09	5.23
<i>p</i>	< 0.01	> 0.05	< 0.001	< 0.05	> 0.05	< 0.001	< 0.001	< 0.001

**(a)** Box plots of convenient sizes for each object scale**(b)** A box plot of convenient sizes for all object scales**Fig. 14** Convenient sizes for **6DOF (Z)** in the main experiment

time when a docking task is completed. It may or may not be a scaled one.

Figure 14a shows the convenient sizes for each object scale. If the given size is small, subjects tended to scale it up. If the given size is large, subjects tended to scale it down.

Figure 14b depicts the statistics of *all* convenient sizes. We take the data between Q1 (0.17 m) and Q3 (0.47 m) as the range of convenient sizes.

Note that the optimal size, 0.45 m, discovered in Sect. 6.2.1 lies in this range. Interestingly, it is closer to the upper bound of the range, 0.47 m, than to the lower bound, 0.17 m. This implies that convenience is rather distinguished from efficiency that is related with the optimal size.

In order to verify the optimal size and the range of convenient sizes identified in the main experiment, we developed a proxy-based metaphor, named **Diorama**, in which a proxy of an object is given “with the optimal size.” **Diorama** also supports *zoom* (presented in Sect. 3.2.2). We hypothesized that

providing such a proxy helped users complete the docking tasks by minimizing the use of zoom operations.

7 Application and additional experiment

7.1 Diorama—proxy-based manipulation

Figure 15 shows a scenario where a docking task is completed with **Diorama**. In addition to the green source and silver target, the *diorama* is presented to the user. As shown in Fig. 15a, it is composed of three handles (named *crop-handle*, *torus-handle* and *stick-handle*) and what we name *proxy-box*, in which objects are manipulated.

When the user touches the source and then selects it (by pressing both trigger and grip buttons), its wireframe bounding box becomes red and its proxy appears in proxy-box (Fig. 15b). The environment surrounding the selected object is also copied into proxy-box. Observe that part of the silver target is visible at the lower-left corner of proxy-box.

By grabbing and dragging stick-handle, the user translates proxy-box to a more convenient location to work (Fig. 15c). Then, in order to secure a better view to work, the user rotates proxy-box by grabbing and rotating torus-handle (Fig. 15d). While the user moves proxy-box, the proxies are accordingly moved. However, the original source and target are fixed in space and do not move at all.

The user considers proxy-box too small and so scales it up by grabbing and pulling crop-handle (Fig. 15e). By doing so, the target in proxy-box is fully visible to the user. Then, the user starts the docking task inside proxy-box (Fig. 15f). This time, the original source is accordingly manipulated.

In Fig. 15f, the proxies are not located in the middle of proxy-box. Feeling uncomfortable with that, the user performs *grab-navigation* (presented in Sect. 3.2.1) by grabbing and pulling the empty space of proxy-box so that the proxies are located in the middle (Fig. 15g). Then, in order to be able to manipulate the source more accurately, as shown in Fig. 15h, the user performs *zoom* by grabbing and stretching

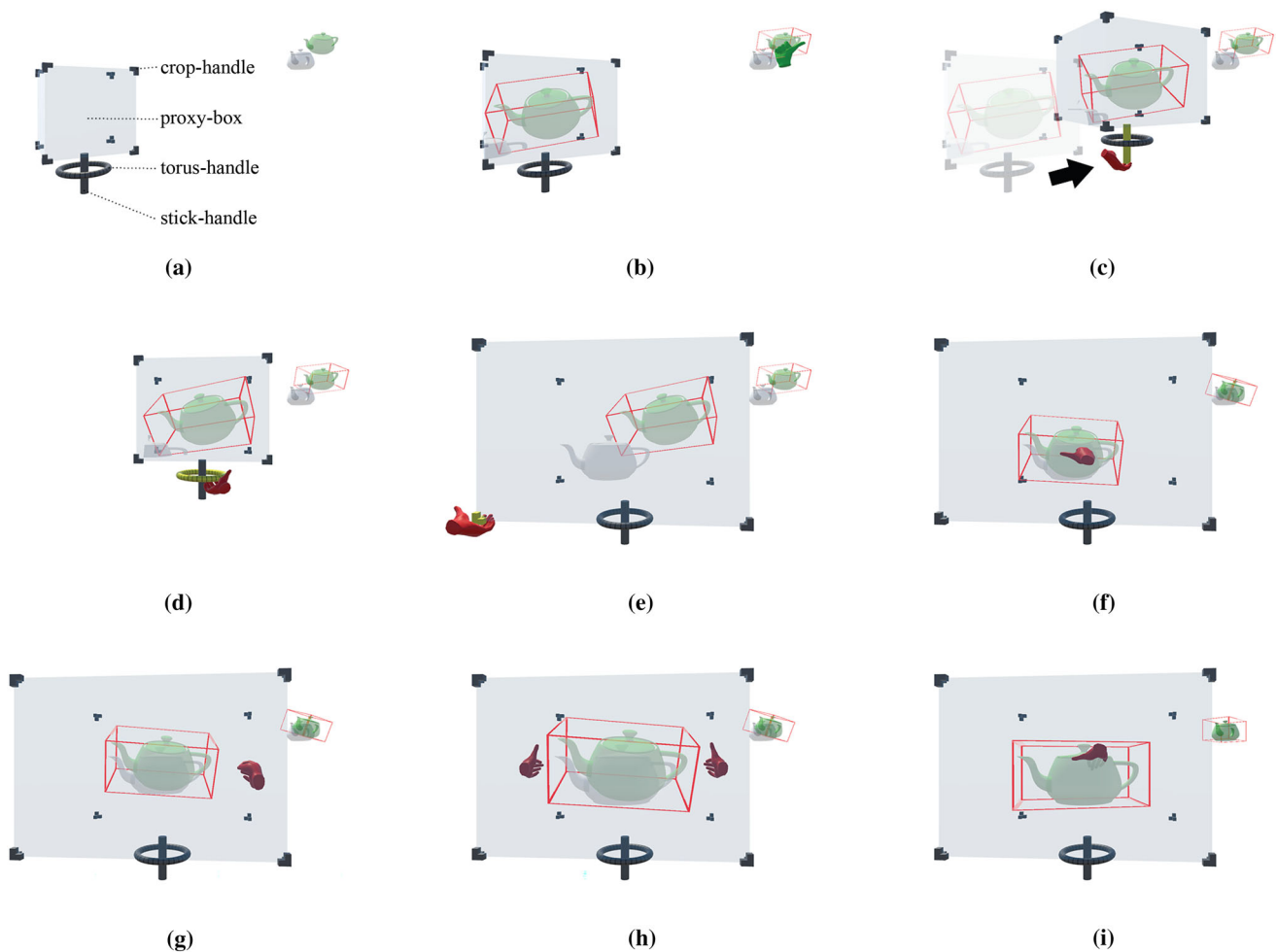


Fig. 15 A docking task completed with **Diorama**

the empty space of proxy-box. (*Zoom* is not available outside proxy-box.) Neither grab-navigation nor zoom alters the original source. Figure 15i shows the state where the docking task is completed.

Suppose that a user judges that the given size of an object is appropriate enough to be manipulated. Then, the user is allowed to manipulate it directly without making a proxy. This implies that ‘diorama’ is not used at all, but the user simply performs **6DOF**.

Following the procedure presented in Table 1, the completion times were measured in **Diorama**. The experiment was made one week after the main experiment, with the same (thirty-three) participants of the main experiment.

7.2 Results and analysis

As in the main experiment, eleven scales were used. Each scale was given twice, making a subject perform 22 docking tasks. In total, 726 docking tasks were performed by thirty-three subjects. Among them, three tasks were not completed

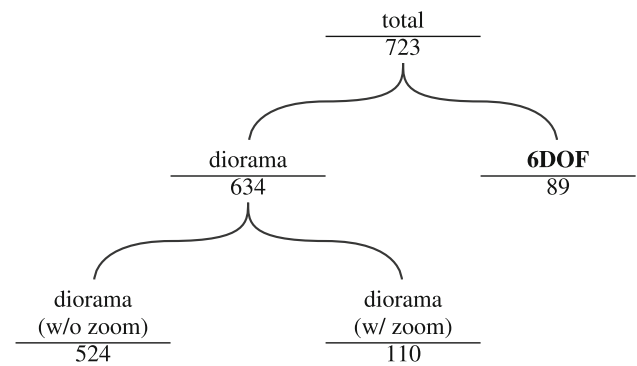


Fig. 16 Classification of docking tasks in **Diorama**

in 60 s. Analysis is made with the remaining 723 tasks. Among them, 634 tasks were performed by using the diorama, whereas 89 tasks were performed without it, i.e., just with **6DOF**. See Fig. 16. Among the 634 tasks completed by using the diorama, zoom was not used in 524 tasks. It was used in 110 tasks only. This fits our hypothesis. Providing the optimal-size objects minimized the use of zoom operations.

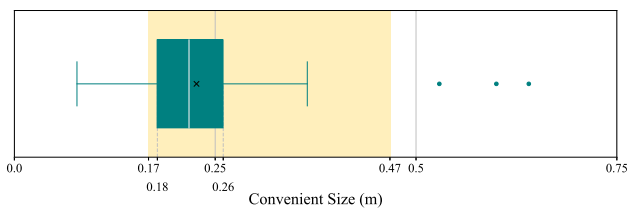


Fig. 17 Convenient sizes for “diorama (w/ zoom).”

The mean of the completion time for “diorama (w/o zoom)” is 12.73 s, whereas it was 18.32 s for “diorama (w/ zoom).” Since the F -test detects inequality of variances between them, Welch’s t -test is performed. It reveals that there exists significant difference between them in terms of completion time ($t = 5.27$, $p < 0.001$). This shows that manipulating an optimal-size proxy is more efficient than using *zoom*.

For the 110 tasks of “diorama (w/ zoom),” we measured the convenient sizes. As depicted in Fig. 17, Q1 of the data is 0.18 m and Q3 is 0.26 m. They fall into the convenient size range, 0.17 to 0.47 m, which is identified in the main experiment and colored in yellow in Fig. 17.

8 Conclusion and future work

This paper reported the results of a series of experiments made for investigating the object sizes that are optimal and convenient for mid-air manipulation in immersive virtual environments. We used a well-known mid-air manipulation method, **6DOF**, and the Utah teapot, which has many distinct geometric features. Through an additional experiment based on a new proxy-based method, **Diorama**, the optimal and convenient sizes were verified. We believe that our studies can give a guideline for advanced studies on the size problems for mid-air manipulation. If fast and accurate mid-air manipulation is required, for example, a proxy-based method is suggested, where the proxy is given with the optimal size. If speed and accuracy are not a critical issue, however, we would suggest to provide the virtual objects in the range of convenient sizes.

However, there exist some limitations in our work. In the current study, a single method (**6DOF**) and a single object (Utah teapot) were used. There exist other mid-air manipulation methods such as Handle-bar metaphor [32], PRISM technique [6,7] and ray-casting method [3]. It has also been reported that 3D objects can be classified by their shapes [43]. Therefore, it is worth investigating the optimal and convenient sizes with multiple manipulation methods and a set of classified objects. On the other hand, it would be also worth investigating if **Diorama** excels different proxy-based methods, where the proxy is given with no optimal sizes. In

addition, recruiting subjects with more diverse natures (in terms of dominant hand, gender, and age) would help the study to be more generalized. These will make up our future research.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00371-022-02555-6>.

Funding This research was supported by the Ministry of Science and ICT, Korea, under the ICT Creative Consilience Program (IITP-2022-2020-0-01819), ITRC (Information Technology Research Center) Support Program (IITP-2022-2020-0-01460) and the grant No. 2020-0-00861, which are all supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation).

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

- Argelaguet, F., Maignant, M.: Giant: stereoscopic-compliant multi-scale navigation in VEs. In: Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology, pp. 269–277 (2016)
- Beever, L., Pop, S., John, N.W.: Leveled VR: a virtual reality level editor and workflow for virtual reality level design. In: 2020 IEEE Conference on Games (CoG), pp. 136–143. IEEE (2020)
- Bowman, D.A., Hodges, L.F.: An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In: Proceedings of the 1997 Symposium on Interactive 3D Graphics, pp. 35–ff (1997)
- Elvezio, C., Sukan, M., Oda, O., Feiner, S., Tversky, B.: Remote collaboration in AR and VR using virtual replicas. In: ACM SIGGRAPH 2017 VR Village, pp. 1–2 (2017)
- Feng, J., Cho, I., Wartell, Z.: Comparison of device-based, one and two-handed 7DOF manipulation techniques. In: Proceedings of the 3rd ACM Symposium on Spatial User Interaction, pp. 2–9 (2015)
- Frees, S., Kessler, G.D.: Precise and rapid interaction through scaled manipulation in immersive virtual environments. In: IEEE Proceedings. VR 2005. Virtual Reality, 2005, pp. 99–106. IEEE (2005)
- Frees, S., Kessler, G.D., Kay, E.: Prism interaction for enhancing control in immersive virtual environments. *ACM Trans. Comput. Hum. Interact. (TOCHI)* **14**(1), 2-es (2007)
- Gloumeau, P.C., Stuerzlinger, W., Han, J.: Pinnpivot: object manipulation using pins in immersive virtual environments. *IEEE Trans. Vis. Comput. Graph.* **27**(4), 2488–2494 (2020)
- Hartman, N., Delahaye, M., Decroix, H., Herbelin, B., Boulic, R.: Does scaling player size skew one’s ability to correctly evaluate object sizes in a virtual environment? In: Motion, Interaction and Games, pp. 1–6 (2020)
- Homan, R.: SmartScene: an immersive, realtime, assembly, verification and training application (1997)
- Jerald, J., Mlyniec, P., Yoganandan, A., Rubin, A., Paullus, D., Solotko, S.: MakeVR: a 3D world-building interface. In: 2013 IEEE Symposium on 3D User Interfaces (3DUI), pp. 197–198. IEEE (2013)

12. Kim, T., Park, J.: 3D object manipulation using virtual handles with a grabbing metaphor. *IEEE Comput. Graph. Appl.* **34**(3), 30–38 (2014)
13. Krekhov, A., Cmentowski, S., Emmerich, K., Masuch, M., Krüger, J.: GulliVR: a walking-oriented technique for navigation in virtual reality games based on virtual body resizing. In: *Proceedings of the 2018 Annual Symposium on Computer–Human Interaction in Play*, pp. 243–256 (2018)
14. LaViola, J.J., Jr., Kruijff, E., McMahan, R.P., Bowman, D., Poupyrev, I.P.: *3D User Interfaces: Theory and Practice*. Addison-Wesley Professional, Boston (2017)
15. Martin-Gomez, A., Fotouhi, J., Eck, U., Navab, N.: Gain a new perspective: towards exploring multi-view alignment in mixed reality. In: *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 207–216. IEEE (2020)
16. Masliah, M.R., Milgram, P.: Measuring the allocation of control in a 6 degree-of-freedom docking experiment. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 25–32 (2000)
17. Medeiros, D., Carvalho, F., Teixeira, L., Braz, P., Raposo, A., Santos, I.: Proposal and evaluation of a tablet-based tool for 3D virtual environments. *SBC J. Interact. Syst.* **4**(2), 30–42 (2013)
18. Mendes, D., Caputo, F.M., Giachetti, A., Ferreira, A., Jorge, J.: A survey on 3D virtual object manipulation: from the desktop to immersive virtual environments. In: *Computer Graphics Forum*, vol. 38, pp. 21–45. Wiley Online Library (2019)
19. Mendes, D., Fonseca, F., Araujo, B., Ferreira, A., Jorge, J.: Mid-air interactions above stereoscopic interactive tables. In: *2014 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 3–10. IEEE (2014)
20. Mendes, D., Relvas, F., Ferreira, A., Jorge, J.: The benefits of DOF separation in mid-air 3D object manipulation. In: *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, pp. 261–268 (2016)
21. Mine, M., Yoganandan, A., Coffey, D.: Principles, interactions and devices for real-world immersive modeling. *Comput. Graph.* **48**, 84–98 (2015)
22. Mine, M.R., Brooks Jr, F.P., Sequin, C.H.: Moving objects in space: exploiting proprioception in virtual-environment interaction. In: *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 19–26 (1997)
23. Oda, O., Elvezio, C., Sukan, M., Feiner, S., Tversky, B.: Virtual replicas for remote assistance in virtual and augmented reality. In: *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pp. 405–415 (2015)
24. Oda, O., Feiner, S.: 3D referencing techniques for physical objects in shared augmented reality. In: *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 207–215. IEEE (2012)
25. Pierce, J.S., Forsberg, A.S., Conway, M.J., Hong, S., Zeleznik, R.C., Mine, M.R.: Image plane interaction techniques in 3D immersive environments. In: *Proceedings of the 1997 Symposium on Interactive 3D Graphics*, pp. 39–ff (1997)
26. Pierce, J.S., Stearns, B.C., Pausch, R.: Voodoo dolls: seamless interaction at multiple scales in virtual environments. In: *Proceedings of the 1999 Symposium on Interactive 3D Graphics*, pp. 141–145 (1999)
27. Pohl, H., Lilija, K., McIntosh, J., Hornbæk, K.: Poros: configurable proxies for distant interactions in VR. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–12 (2021)
28. Pouke, M., Mimnaugh, K.J., Chambers, A.P., Ojala, T., LaValle, S.M.: The plausibility paradox for resized users in virtual environments. *Front. Virtual Reality* **2**, 48 (2021)
29. Poupyrev, I., Billinghurst, M., Weghorst, S., Ichikawa, T.: The go-go interaction technique: non-linear mapping for direct manipulation in VR. In: *Proceedings of the 9th Annual ACM Symposium on User Interface Software and Technology*, pp. 79–80 (1996)
30. Poupyrev, I., Weghorst, S., Billinghurst, M., Ichikawa, T.: A study of techniques for selecting and positioning objects in immersive VEs: effects of distance, size, and visual feedback. In: *Proceedings of ACM CHI*, vol. 98 (1998)
31. Rodrigues, V., Mendes, D., Ferreira, A., Jorge, J.: Mid-air manipulation of 3D models in (semi-) immersive virtual environments. *Atas do Encontro português de Computação Gráfica e Interação-SciTecIn/EPCGI 2015*, pp. 100–108 (2015)
32. Song, P., Goh, W.B., Hutama, W., Fu, C.W., Liu, X.: 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. 1297–1306 (2012)
33. Stoakley, R., Conway, M.J., Pausch, R.: Virtual reality on a WIM: interactive worlds in miniature. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 265–272 (1995)
34. Wang, J., Lindeman, R.: Coordinated hybrid virtual environments: seamless interaction contexts for effective virtual reality. *Comput. Graph.* **48**, 71–83 (2015)
35. Wang, M., Ye, Z.M., Shi, J.C., Yang, Y.L.: Scene-context-aware indoor object selection and movement in VR. In: *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 235–244. IEEE (2021)
36. Wang, R., Paris, S., Popović, J.: 6D hands: markerless hand-tracking for computer aided design. In: *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, pp. 549–558 (2011)
37. Whitlock, M., Harnner, E., Brubaker, J.R., Kane, S., Szafir, D.A.: Interacting with distant objects in augmented reality. In: *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 41–48. IEEE (2018)
38. Wilkes, C., Bowman, D.A.: Advantages of velocity-based scaling for distant 3D manipulation. In: *Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology*, pp. 23–29 (2008)
39. Xia, H., Herscher, S., Perlin, K., Wigdor, D.: Spacetime: enabling fluid individual and collaborative editing in virtual reality. In: *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, pp. 853–866 (2018)
40. Yang, Y., Cordeil, M., Beyer, J., Dwyer, T., Marriott, K., Pfister, H.: Embodied navigation in immersive abstract data visualization: is overview+ detail or zooming better for 3D scatterplots? *IEEE Trans. Vis. Comput. Graph.* **27**(2), 1214–1224 (2020)
41. Yu, K., Winkler, A., Pankratz, F., Lazarovici, M., Wilhelm, D., Eck, U., Roth, D., Navab, N.: Magnoramas: magnifying dioramas for precise annotations in asymmetric 3D teleconsultation. In: *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 392–401. IEEE (2021)
42. Zeelenberg, R., Pecher, D.: A method for simultaneously counterbalancing condition order and assignment of stimulus materials to conditions. *Behav. Res. Methods* **47**(1), 127–133 (2015)
43. Zingg, T.: *Beitrag zur schotteranalyse*. Ph.D. thesis, ETH Zurich (1935)

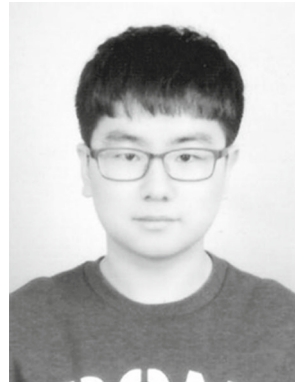


Myoung Gon Kim is currently a Ph.D. student in the Interactive 3D Media Lab directed by Prof. JungHyun Han at Korea University, Seoul, South Korea. He received a B.S. degree and an M.Eng. degree from the Department of Computer Science and Engineering at Korea University. He worked as an engineer at Navigation group, Robot team, Research Center of Samsung Electronics Co., Ltd. for 3 years. His research interests lie in

simultaneous localization and mapping in virtual and augmented environments.



JiSeok Ryu received a B.S. degree in computer science and engineering from the Department of Computer Science and Engineering at Korea University. He is currently pursuing an M.Eng. degree in the Interactive 3D Media Lab directed by Prof. JungHyun Han at Korea University. His research interest is human–computer interaction in virtual reality, such as media manipulation and text entry in virtual environments.



Jaemin Son received a B.S. degree in computer science and engineering from the Department of Computer Science and Engineering at Korea University. He is currently pursuing an M.Eng. degree in the Interactive 3D Media Lab directed by Prof. JungHyun Han at Korea University. His research is mainly focused on the human–computer interaction in augmented reality and virtual environments.



JungHyun Han is a professor of the Department of Computer Science and Engineering at Korea University, where he directs the Interactive 3D Media Laboratory. Prior to joining Korea University, he was a faculty member of Sungkyunkwan University in Korea and a researcher of US Department of Commerce National Institute of Standards and Technology (NIST). His research interests lie in 3D graphics and game. Dr. Han received a B.S. from Seoul National University,

an M.S. degree from the University of Cincinnati, and a Ph.D. degree from the University of Southern California. His degrees are all in Computer Science.