The Benefits of Near-field Manipulation and Viewing to Distant Object Manipulation in VR

Sabarish V. Babu*
Clemson University

Wei-An Hsieh†
National Yang Ming Chiao Tung University

Jung-Hong Chuang‡
National Yang Ming Chiao Tung University

Abstract

In this contribution, we propose to enhance two distant object manipulation techniques, BMSR (Bimanual Near-Field Metaphor with Scaled Replica) and the classic Scaled HOMER (Scaled Hand-Centered Object Manipulation Extending Ray Casting), via near-field scaled replica manipulation and viewing. In the proposed Direct BMSR, context replicas are displayed so that the target replica can be manipulated relative to its context, allowing the user to directly manipulate the target replica in their arm's reach space. Some additional features were implemented to make Direct BMSR an effective interface for manipulating objects from a distance. We proposed Scaled HOMER+NFSRV, which augments Scaled HOMER with a near-field scaled replica view (NFSRV) of the target object and its context, enabling the user to observe how the target replica is manipulated in relation to its context in their arm's reach space while manipulating it from a distance. We conducted a between-subjects empirical evaluation of BMSR, Direct BMSR, Scaled HOMER, and Scaled HOMER+NFSRV. Our findings revealed that Direct BMSR and Scaled HOMER+NFSRV significantly outperformed BMSR and Scaled HOMER, respectively, in terms of accuracy. This finding highlights the advantages of adding near-field scaled replica viewing and manipulation with respect to distant object manipulation.

Keywords: Near-field scaled replica manipulation, Near-field scaled replica viewing, Distant object manipulation, Virtual object manipulation

Index Terms: Human-centered computing [Interaction design]: Interaction design process and methods—User interface design

1 Introduction

In recent years, the growth of virtual reality (VR) technology and its applications has been remarkable. This is due to the availability of cost-effective, accessible, and powerful hardware, as well as the progress made in rendering, tracking, and human interfaces. Manipulation of virtual objects in VR is essential for any kind of application, yet it is still considered a difficult task, despite being studied for many years. To manipulate objects located at a distance in VR, the user can move close to the target object and then directly manipulate it within their near field or arm reach space [4, 11, 15, 25]. Alternatively, the user can manipulate the object from a distance [5, 21, 29, 32, 33, 36].

There are two approaches to manipulating objects from a distance: directly manipulating the objects from a distance or indirectly manipulating scaled replicas of the objects in the user's arm-reach space. Distal manipulations such as ray casting [27] and HOMER (Hand-Centered Object Manipulation Extending Ray Casting) [5] may magnify or scale up the manipulation error due to hand jitter. Scaled HOMER [36] combines HOMER [5] with PRISM [14], with the aim of reducing the error scale-up by using velocity-based scaling. Although the scaling error can be reduced by using velocity scaling, Scaled HOMER still suffers from possible poor vision and impaired depth perception when manipulating distant objects [12]. In addition to velocity-based scaling, DOF separation transformation is another technique to reduce manipulation error [25]. However, none of the current distal techniques provide DOF separation transformations. For example, Scaled HOMER only offers 6DOF simultaneous translation and rotation, which is mainly suitable for coarse transformations [25]. Selecting an object from a distance and manipulating it in the arm's reach space would provide the advantages of both distal and direct near-field manipulation approaches, such as more precise motion control, sharp vision, and more accurate depth perception in the near field, without having to virtually move to the object. Various efforts have been made in the past, such as Worlds-In-Miniature (WIM) [33], Scaled-World Grab [28], and Voodoo Dolls [29]. However, these approaches do not prioritize manipulation accuracy and are not suitable for manipulation tasks that require accuracy. However, this method shows the highest potential for creating a precise interface by manipulating near-field replicas. This is due to the advantages that come with near-field manipulation.

We have enhanced BMSR [21] to to allow users to directly manipulate target replicas in their arm's reach space with satisfactory accuracy and ease, which is called *Direct BMSR*. Additionally, we have augmented Scaled HOMER [36] with a near-field view of the target object and its context, called Scaled HOMER with Near-Field Scaled Replicas Viewing or *Scaled HOMER+NFSRV*. We believe that no prior research has been conducted on the combination of near-field replica manipulation with DOF separation transformation on distant object manipulation, as well as on the augmentation of near-field viewing to a distal manipulation technique. In Direct BMSR, replicas of the context objects are displayed so that the target replica can be manipulated in relation to its context replicas in the user's arm's reach space. This allows the user to perform direct near-field manipulation for objects that are far away, taking advantage of the key benefits of near-field manipulation such as more precise motion control, sharper vision, and more accurate depth and relative size perception. In addition to providing DOF separation transformations such as 1D-3D translation, 1D-3D scaling, 1D rotation, and 6DOF simultaneous translation and rotation, Direct BMSR has been equipped with features to make it an effective and convenient tool for manipulating objects from a distance. In Scaled HOMER+NFSRV, we augment Scaled HOMER with a near-field view of scaled replicas of the target object and its context objects. This allows the user to observe the motion of the target replica and its context when manipulating it from a distance. We anticipate that this enhancement will significantly improve vision and depth perception, thus increasing the accuracy of manipulation. Finally, we comparatively evaluate BMSR, Direct BMSR, Scaled HOMER and Scaled HOMER+NFSRV on distant object manipulation effectiveness, efficiency, and user experience.

*e-mail: sbabu@clemson.edu
†e-mail: b04902106@ntu.edu.tw
‡e-mail: jhchuang@cs.nycu.edu.tw

2 Related Work

2.1 Direct Near-field Manipulation

Direct manipulation is the most popular and intuitive way of interaction with objects in VR, allowing users to directly manipulate objects within their near-field or arm's reach space. Examples include Simple Virtual Hand [4], Air-TRS [24], Spindle [22], Handle Bar [3, 32], Spindle+Wheel [9], and Crank Handle [3]. Some of these aim to improve accuracy, such as Grasping Object [3], 6-DOF Hand [24], 3-DOF Hand [24], Widgets [25], and PinNPivot [15]. With near-field manipulation, the user needs to walk or teleport towards target objects that are outside of user's arm's reach before manipulating them. Research has shown that repeated switching between object interaction and movement can break the rhythm of one's interaction with the VE [28]. However, when working in the near field, the user has finer motion control, a keen vision, a more accurate depth perception, and hence a greater sense of the object's position [2, 20, 28]. Moreover, with 6DOF simultaneous translation and rotation, direct near-field manipulation can mimic interactions in the real world, albeit with less precision [2, 4, 6].

Compared to manipulating distant objects from a distance, it is easier to develop DOF separation transformations for near-field manipulation. However, only a few metaphors offer DOF separation transformations. For example, Air-TRS [24], 3-DOF Hand [24], and Crank Handle [3] separate translation and rotation in 6DOF simultaneous translation and rotation, while Widgets [25] provide a 1DOF translation and a 1DOF rotation based on virtual handles. While most techniques include translation, rotation, and scaling, VR techniques that offer scaling only offer 3D uniform scaling [11, 24, 32]. More recently, PinNPivot has offered a richer set of transformations, including 3D translation, 1-3D rotation, and 6DOF simultaneous translation and rotation [15].

Widgets based on the bounding box of the target object have been used for object manipulation in interfaces for the mouse and keyboard [18], later for multitouch devices [10], and recently for VR devices [21]. The handle box is a bounding box of the object on which a lifting handle is used to move the object up and down and four rotation handles are used to rotate the object about its central axis [18]. The tBox widget consists of a wireframe box surrounding the target object, on which the user can drag an edge of the box to move the object along the axis containing the edge or drag a face of the box to rotate the object [10]. Recently, the bounding box widget in BMSR [21] allows the user to perform 1D, 2D, and 3D translations and scalings by dragging the bounding box's faces, edges, and vertices, respectively, and to do 1D rotation by grabbing a handlebar and dragging a box's edge. The bounding box widget in BMSR was developed for distant object manipulation but can also be applied to direct manipulation.

2.2 Distal Manipulation for Distant Objects

Go-Go [30] and ray casting [27] were proposed in the mid-1990s to interact with distant virtual objects. Go-Go uses the metaphor of extending the user's arm and employs a non-linear mapping for interacting and manipulating distant objects, while in ray-casting the user selects an object with a cast ray and manipulates the object that is stuck at the end of the ray. As indicated in [5], Go-Go, Stretch Go-Go, and ray casting had significant drawbacks. Bowman et al. combined Go-Go and ray casting into an out-of-reach interaction technique called HOMER (Hand-Centered Object Manipulation Extending Ray casting) [5], which provides better control ability than Go-Go or ray casting alone. With HOMER, however, the error due to hand jitter could be scaled up.

When manipulating an object, people tend to move relatively rapidly when precision is not a concern, but will slow down their hand movement when precision is a concern. On the basis of this observation, the PRISM mechanism [13] is a velocity-based scaling that scales the velocity of the object based on the velocity of the hand. To reduce the scale-up error of HOMER, Wilkes et al. applied the PRISM mechanism to HOMER and proposed Scaled HOMER [36], claiming that it performs significantly more accurately than HOMER. Although the scale-up error has been reduced, Scaled HOMER could suffer from problems such as poor vision and impaired depth perception. Moreover, it only supports 6DOF simultaneous translation and rotation, which is primarily suitable for coarse transformations [23]. In addition to velocity-based scaling, another approach proposed to increase manipulation precision is DOF separation [25, 35]. Mendes et al. compared Simple Virtual Hand with 6DOF translation and rotation, 6DOF translation and rotation with velocity-based scaling, and a widget for full DOF separation, and concluded that full DOF separation through widgets can lead to precision improvement at the cost of increased time for complex tasks [25]. However, it appears to be infeasible for the distal manipulation interface to perform DOF separation operations.

2.3 Near-field Replica for Distant Object Manipulation

The World-in-Miniature (WIM) [33] is a miniature model of the entire environment, providing users with a global view of the environment, an effective interface for object selection and manipulation, and teleportation. However, the accuracy of the manipulation is not its main concern. The Voodoo Dolls proposed by Pierce et al. [29] allow users to manipulate the doll of the target object held in their dominant hand relative to the dolls of the context objects held in their non-dominant hand. Voodoo dolls take advantage of the separation of labor between the dominant and non-dominant hands [16], in which the dominant hand of the user works within the reference frame defined by their non-dominant hand [16]. This provides a natural interface for object manipulation but may suffer from accuracy issues due to the instability of moving both hands and 6DOF simultaneous translation and rotation. The near-field metaphor with scaled replicas (BMSR) proposed in [21] aims to improve the accuracy of manipulating distant objects by two mechanisms. First, it manipulates the scaled replica of the target object in the arm's reach space instead of the object space, thus taking advantage of finer motion control. Second, based on a bounding box widget, it supports multilevel DOF separation, leading to more accurate manipulation [25] and more flexibility for complex tasks. However, without the display of replicas for the objects that are contextual to the target object, it still suffers from the error scale-up problem, as well as less keen vision and impaired depth perception.

3 Near-field Manipulation and Viewing

Bimanual Near-field Metaphor with Scaled Replica (BMSR) is an interaction technique that facilitates manipulation of distant objects in the user's near field or arm's reach space and can provide finer control with the capability of performing multilevel DOF separation operations for translation, rotation, and scaling [21]; see Fig. 2a. In BMSR, after the target object is selected, its scaled replica is placed in the arm's reach space of the user with a size of 20 centimeters for convenient near-field manipulation. The target replica is enclosed by a bounding box (Fig. 1). Interaction with the box's faces, edges, and vertices defines 1D, 2D, and 3D translation and scaling, respectively. For example, to translate the target object along one axis, the user grabs the face that is oriented on that axis and moves it (see Fig. 1a). The edges of the bounding box correspond to 2D translation in the plane orthogonal to the edge (see Fig. 1b). The vertices of the bounding box correspond to a 3D translation (see Fig. 1c). For uniform scaling, the user can grab a pair of opposite primitives on the bounding box and move both hands towards or away from each other. That is, by grabbing a face, edge, or vertex with one hand and grabbing the opposite face, edge, or vertex with the other hand, the user can perform uniform 1D, 2D, and 3D scaling, respectively. For anchored scaling, the hand that grabs the anchored primitive needs to remain almost stationary while the other hand moves. For 1D rotation, the user grabs the handlebar that appears at the center of a face and perpendicular to the rotation axis using their non-dominant hand, and then grabs and rotates an edge that is parallel to the rotation axis with dominant hand.

Scaled HOMER HOMER is a hybrid metaphor that integrates Go-Go [30] and ray casting [27] into a distal technique to manipulate distant objects [5]. It provides better control-ability than Go-Go or ray casting alone, but suffers from an error scale-up problem. To reduce error scaling, HOMER was later combined with the velocity-based scaling proposed in PRISM [14] to form Scaled HOMER [36]; see Fig. 4a. This method of manipulating objects from a distance is the most commonly used; however, it has some drawbacks, such as possible poor vision and impaired depth perception for distant objects.

(a) 1D: Grab face. (b) 2D: Grab edge. (c) 3D: Grab vertex.

Figure 1: 1D, 2D, and 3D translation in BMSR.

3.1 Direct BMSR

When replicas of objects that are contextual to the target object are not displayed in the user's arm's reach space, the target replica is manipulated without any reference. Consequently, with BMSR, the user indirectly manipulates the target object by manipulating its replica and must observe how the object has moved related to its context in the object space while manipulating the replica (see Fig. 2a). This approach has issues such as scaling errors, possible poor vision, and impaired depth and size perception for distant objects. To enhance BMSR, we display replicas of the target object and its context objects in the arm's reach space, allowing the user to perceive and manipulate the target replica relative to its context replicas (Fig. 2b). As a result, the user can manipulate the target replica directly in their arm's reach space when manipulating a distant object, which is why we refer to enhanced BMSR as *Direct BMSR*. This type of manipulation in the near field has the advantages of finer motion control, clearer near-field vision, improved depth perception, better perceptuo-motor coordination [20, 21] and additional contextual information, which will lead to more accurate and effective manipulation. Manipulating objects in the near field, Direct BMSR has the same characteristics as direct near-field manipulation, such as finer motion control, better near-field visual clarity, improved depth perception, perceptuo-motor coordination [20, 21] and additional contextual data, which will improve manipulation accuracy and efficiency. The context objects are those that intersect with a sphere centered at the center of the target object, with a radius of six times the maximum edge size of the target object's bounding box. The size of the target replica has been reduced from 20 centimeters, as in BMSR, to 15 centimeters to give more room for context replicas. It is assumed that the user and the virtual environment are in a 1:1 scale ratio.

Although Direct BMSR has finer motion control compared to distal object manipulation, it still has an issue with hand jitter when manipulating replicas. To reduce the manipulation error caused by hand jitter, we implemented PRISM [14], which reduces the movement of the target replica when the user's hand velocity falls below a certain threshold, making it less sensitive to hand jitter. Moreover, the accumulated offset representing the distance between the hand and the target replica can be decreased or restored by moving the hand back toward the target replica [14]. To reduce the issue of occlusion, each context replica is rendered in a translucent form and will become opaque when the target replica is close.

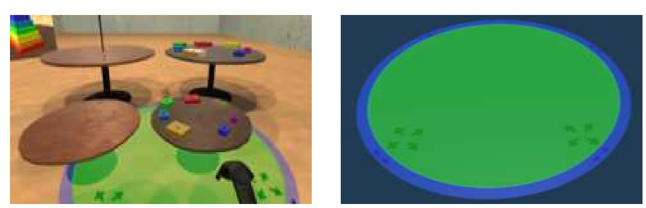
(a) BMSR (b) Direct BMSR

Figure 2: Scenarios for BMSR and Direct BMSR.

To facilitate effective object manipulation, we have added some features to Direct BMSR. The first is the ability to select a context replica as the next target for manipulation. In practice, after manipulating an object, the user often wants to adjust the position or orientation of some context objects. With this feature, the user can quickly and easily manipulate the context replica without having to go back into the object space, select the target context object, and then enter Direct BMSR again. In order to avoid a sudden shift in the display, when a context replica is selected as the next target to be manipulated, all replicas will not be reallocated.

As the second feature, we have incorporated a platform interface into Direct BMSR that enables the user to rotate or translate all replicas. This is especially useful when the target replica is blocked by other replicas, as the user can rotate them all so that the target replica can be identified, selected, and manipulated. A platform of this kind can be useful for selecting and manipulating an object that is hidden in the object space. To do this, the user can first select an object that is close to the target object and is blocking it from view, then rotate all the replicas in the near field so that the target replica is visible, and finally select and manipulate it. Due to the restricted workspace in the near field, for long-distance translation, Direct BMSR may require multiple cycles of selecting an object from the scene and performing a short-range translation in the near-field space. The platform allows the user to carry out a series of short-range translations for the target replica, followed by using the platform to move all replicas back to their front near-field space, all within the Direct BMSR.

Fig. 3a illustrates a platform interface with a circular table and a ring on it. The user can grab and move the table to do a 3D translation of all replicas or grab and rotate the ring to perform a 1D rotation of all replicas. Initially, the table and the ring are rendered with high transparency, but when the controller approaches them, the transparency decreases. Additionally, symbols on the table and ring serve as signifiers to indicate the types of operation the user can do intuitively, as shown in Fig. 3b.



(a) Direct BMSR with the platform (b) Platform interface

Figure 3: Platform interface in Direct BMSR.

3.2 Scaled HOMER+NFSRV

We propose an enhancement to Scaled HOMER to address the problem of poor vision and depth perception. This enhancement, called Scaled HOMER with Near-field Scaled Replicas Viewing (Scaled HOMER+NFSRV), allows the user to observe the motion of the target replica in relation to its context replicas in the near field while manipulating the distant target object; as illustrated in Fig. 4b.

In the Scaled HOMER+NFSRV, once an object is selected, scaled replicas of the object and its context objects are displayed in front of the user's near-field space. The context objects of the target object are determined, and the size of the replica for the target object is set to 15 centimeters, as in Direct BMSR. The size of the replica of an object is determined by the scaling factor, which is expressed as SF:

$$SF = 15\ cm/ObjectSize, \tag{1}$$

where *ObjectSize* is the maximum edge size (in centimeters) of the target object's bounding box. We determine the position of the replica of the context object i using a reference point in the scene, R_{SCENE}, and a reference point in the near-field space, R_{NEAR}, along with the scaling factor SF.

$$Replica_i = R_{NEAR} + SF(Object_i - R_{SCENE}), \tag{2}$$

where $Object_i$ and $Replica_i$ are centers of the object i and its replica, respectively.

Once the target object has been chosen, the center of the object is set to R_{SCENE} and R_{NEAR} to $NF_{Anchored}$, which is a point located beneath the line connecting the midpoint of the user's eyes and the center of the target object. This is defined as:

$$NF_{Anchored} = 60\ cm * V + (0, -50\ cm, 0), \tag{3}$$

where V is the unit vector (in centimeters) from the midpoint of the user's eyes to the center of the target object. In Eq. 3, the value of 60 *cm* is approximately two thirds of the arm length, and 50 *cm* was experimentally determined by piloting different distances. This was done to ensure a comfortable view of the replicas and to avoid too much obstruction of the scene objects. As we can see, at present the center of the target replica is at R_{NEAR}.

(a) Scaled HOMER (b) Scaled HOMER+NFSRV

Figure 4: Scenarios for Scaled HOMER and Scaled HOMER+NFSRV.

The next thing we need to address is how we place the replicas of context objects when the target replica is moved. The first option is to keep the context replicas in place when the target replica is moved, as is usually seen in reality. This can be achieved with fixed R_{SCENE} and R_{NEAR}. However, if the target object is moved a great distance, the target replica may be displayed at a location far from $NF_{Anchored}$, since the target replica is positioned based on SF and R_{SCENE} and R_{NEAR} remain stationary. An alternative is to reset R_{SCENE} and R_{NEAR} to the center of the target object and $NF_{Anchored}$ respectively in each frame. This will lead to a circumstance in which context replicas move in the opposite direction to the target replica's motion.

We proposed that when the angle between the direction vector from the user to the current center of the target object and the direction vector from the user to the last reset or the initial R_{SCENE} is greater than a certain angle threshold, R_{SCENE} should be reset to the center of the target object and R_{NEAR} to $NF_{Anchored}$. This scheme will cause the context replicas to remain in a fixed position between resets. When R_{SCENE} and R_{NEAR} are reset, all replicas will be relocated according to reset R_{SCENE} and R_{NEAR}, resulting in a sudden change in the replica display. The size of the threshold determines the magnitude of the sudden change, with larger thresholds leading to a greater sudden change in replica display, but fewer resets. In contrast, smaller thresholds result in less sudden change, but more frequent resets. It is important to note that the context replicas remain stationary between reset times.

We attempted to establish a fixed angle threshold and discovered that the sudden change in replica display usually takes place during a fine-manipulation stage, even with a large angle threshold. During a coarse or long-distance relocation of the target replica, we noticed that the velocity of movement is usually high and a static display of the context replicas is not usually required. When it comes to a fine-manipulation stage, the hand movement velocity decreases drastically and the target replica should move in relation to its context replicas. At this point, the context replicas should remain still. Keeping this in mind, we designed the following angle threshold setting.

$$Threshold_{angle} = \begin{cases} 0^\circ & \text{if the hand movement velocity} > SC \\ 90^\circ & \text{Otherwise,} \end{cases} \tag{4}$$

where SC is the Scaling Constant in [13], which is a relatively low velocity. When the velocity of hand movement is slower than the Scaling Constant (SC) of 0.15 m/sec, it is likely that the user intends to perform a fine manipulation [13]. The SC value of 0.15 m/sec for SC is considered a normal Scaling Constant value in [13]. The value of 90 degrees in Equation 4 was determined experimentally by piloting various values. We employ hysteresis boundaries to reduce dithering. When the hand moves from high to slow velocity, $Threshold_{angle}$ is altered from 90° to 0° if the hand movement velocity is less than 0.5 SC. On the contrary, when the hand moves from a slow to a high velocity, the $Threshold_{angle}$ is changed from 0° to 90° if the hand movement velocity is greater than 2.0 SC.

4 User Study

4.1 Research Questions and Hypotheses

The research question that we want to answer is: *To what extent do users' performance and preferences differ between the study conditions some of which feature near-field replica viewing and interaction for distant object manipulation?* For the research question, we have the following hypotheses:

H1: *Direct BMSR is expected to have higher accuracy and economy of movement than BMSR.*

H2: *Scaled HOMER+NFSRV is expected to have greater accuracy and economy of movement than Scaled HOMER.*

H3: *BMSR is expected to have higher accuracy and economy of movement than Scaled HOMER.*

H4: *Direct BMSR is expected to have higher accuracy and economy of movement than Scaled HOMER+NFSRV.*

H5: *Both Scaled HOMER and Scaled HOMER+NFSRV is expected to have higher efficiency than both BMSR and Direct BMSR.*

The rationale behind H1 and H2 is that near-field scaled replica manipulation provides direct manipulation with finer motion control, motion parallax, personal space viewing, and close-quarters visual feedback when manipulating distant objects. In addition, near-field scaled replica viewing may enable users to clearly see how the distant target object is moved relative to its context in near-field space. Furthermore, with enhanced clarity, context, and viewing distance, users may not need to try multiple times to move the target object to the desired position, which could result in a higher level of *economy of end-effector movement* [7,8]. For H3 and H4, DOF separation transformations have been shown to be more accurate than simultaneous 6DOF translation and rotation [25,35]. Moreover, the Control/Display ratio of Direct BMSR is much smaller than that of Scaled HOMER, resulting in a smaller error scale-up than Scaled HOMER and Scaled HOMER+NFSRV. Regarding H5, with 6DOF simultaneous translation and rotation, we expect that Scaled HOMER and Scaled HOMER+NFRV will be more efficient than BMSR and Direct BMSR.

4.2 User Study Design

Participants We conducted an a priori power analysis using G*Power to calculate the sample size. Based on an effect size of 0.30, an alpha level of 0.05, a beta level of 0.95, the number of conditions/groups of 4, the number of measurements of 2, the correlation among repeated measures of 0.5 and the non-spehricity correction of 1, revealed a minimum of 52 participants. Thus, we recruited a total of 64 participants, 16 in each condition, using our university's IRB approved Facebook recruiting page. The participants were randomly assigned to each of the 4 conditions. The ages of the participants ranged from 18 to 40 years and comprised 34 men and 30 women. Most of them play PC or smartphone games, and none of the participants had any VR experience.

Task Design Four tasks were chosen for the user study based on the relevant literature on user evaluation of interaction techniques for object manipulation in VR. These tasks that we replicated include *Rainbow Tower*, *Docking*, *Ring through Tube*, and *Pick-and-Place*, which have been used to evaluate interaction techniques for manipulation in the literature and are typically designed to be realistic, concrete, and ecologically valid [8,21,25]. In the Rainbow Tower task, there is a table on the left with a bar stand in the center and a table on the right with seven cuboids. The cuboids are of different colors and sizes and each has a pillar-shaped hole in the center, as shown in Figure 5a. The mission in this task is to move and pile up the cuboids through the bar stand as accurately as possible, as shown in the figure. As shown in Figure 5b, the Docking task is a room with some furniture, four of which are scattered on the floor. The participant is required to move each piece of the four furniture to its destination as accurately as possible, which is represented by the same piece of furniture in semi-transparency fashion. In the Ring Through Tube task, participants need to move the yellow ring, initially positioned on the left side, and pass it through the red M-shaped tube model without collisions and as accurately as possible, as shown in Figure 5c. This is similar to a toy for children to enhance perception-action coordination. The Pick-and-Place task is shown in Figure 5d. The objective of this task is to move a yellow cuboid, which is sitting on the right side of the table, through the bar stand, down to the table top, and fit it into a 3D semi-transparent cuboid on the table as accurately as possible, devoid of any collisions. During the process, they need to translate and rotate the cuboid to avoid the wooden blocks.

Study Procedure Participants first had to complete an informed consent form approved by the institutional review board. In the pre-experiment phase, the participants completed the demographic questionnaire and the SSQ questionnaire [19]. After that, participants were assigned to one interaction technique condition. Then, in a training phase, they watched a demonstration video of the condition and then practiced the condition on a simple fine motor training task in two practice trials. When they practiced, they were allowed to ask any questions about the assigned interaction condition. In the testing phase, the order of the tasks each participant performed was randomly presented using a balanced Latin square design, and for each task, the distance from the target (near = 3m or far = 10m) was presented in random order. We instructed the participants to complete each task as accurately as possible, with the least number of collisions possible. There was no time limit to complete the task. After the participants completed the VR testing phase, they completed the post-experiment questionnaires. After that, they were debriefed, compensated for their time and thanked for participating in the study.

(a) Rainbow Tower

(b) Docking

(c) Ring Through Tube

(d) Pick and Place

Figure 5: Four tasks for the user study.

Data Collection and Metrics The objective data we collected during the testing phase were as follows: 1. movement time; 2. number of attempts; 3. number of collisions; 4. path length; 5. total rotation; 6. position error; 7. rotation error; 8. proportion of time spent watching in the near-field or personal space.

Efficiency is related to movement time and the number of attempts. *Economy of movement* can be represented by hand/controller path length and total rotation. Finally, *accuracy* is calculated from the number of collisions, the position error, and the rotation error.

In the post-experiment phase, subjective quantitative data was collected from a series of questionnaires. Participants completed the Simulator Sickness Questionnaire (SSQ) [19], Workload Assessment Survey (NASA-TLX) [17], Presence Questionnaire (IPQ-Presence) [31], IBM System Usability Scale (SUS) [1], and a system performance questionnaire designed by the experimenters. The system performance questionnaire contained general questions that asked participants their opinions on translation, rotation, and simultaneous translation and rotation operations, as well as a set of condition-specific user impression questions.

Analysis Method On all quantitative objective data, parametric ANOVA analyzes were performed on the data after carefully verifying that the underlying assumptions were met, namely that the data in the samples were normally distributed and the error variance between the samples was equivalent. We ensured that Box's test of equality of covariance matrix was not significant. Levene's test was conducted to verify homogeneity of variance, and Mauchly's test of sphericity was conducted to ensure that the error variance in groups of samples was equivalent. After verifying that the assumptions were met, we subjected each quantitative objective measure to a mixed model ANOVA analysis. Pairwise post-hoc tests between levels of the between-subjects variables (i.e. conditions) was conducted using Tukey's HSD analysis, whereas between levels of the within-subjects variables (i.e. target distance) was conducted using the Bonferroni adjusted alpha method. Greenhouse-Geisser correction and adjustment to degrees of freedom were applied when Mauchly's test of sphericity was violated. The subjective quantitative data gathered were subjected to a non-parametric statistical analysis. The non-parametric analysis method consisted of first subjecting the non-parametric scores to a Kruskal-Wallis H test, followed by post-hoc pairwise comparisons using the Mann-Whitney U test.

5 Results

5.1 Quantitative Objective Results

After verifying the assumptions of the parametric analysis, we subjected the objective metrics to a 4(conditions) x 2(target distance) mixed model ANOVA analysis with condition as a between-subjects variable and target distance as a within-subjects variable. Post-hoc pairwise comparisons between conditions was conducted using Tukey's HSD analysis, and between target distances was conducted using Bonferroni adjusted alpha method.

Mixed Model ANOVA Results: With regards to *total movement time* we found a significant main effect of distance only, F(1,57) = 33.856, p<0.001, part. η^2 = 0.373. With regards to the *number of attempts*, we found a significant main effect of condition, F(3,57) = 10.525, p<0.001, part. η^2 = 0.356, and main effect of distance, F(1,57) = 16.655, p<0.001, part. η^2 = 0.226. With regards to the *path length*, we found a significant main effect of condition, F(3,57)=11.215, p<0.001, part. η^2 = 0.371, main effect of distance, F(1,57)=12.884, p=0.001, part. η^2 = 0.184, and a distance-by-condition interaction effect, F(3,57)=10.513, p<0.001, part. η^2 = 0.356. With regards to the *total rotation*, we found a significant main effect of condition, F(3,57)=4.933, p=0.004, part. η^2 = 0.206, main effect of distance, F(1,57)=10.069, p=0.002, part. η^2 = 0.15, and a distance-by-condition interaction, F(3,57)=6.527, p=0.001, part. η^2 = 0.256. With regards to the *positional displacement (accuracy) with respect to y-axis*, we found a main effect of distance, F(1,57)=6.346, p=0.015, part. η^2 = 0.1. With regards to the *positional displacement (accuracy) with respect to z-axis*, we found a significant main effect of condition, F(3,57)=36.586, p<0.001, part. η^2 = 0.658, main effect of distance, F(1,57)=48.165, p<0.001, part. η^2 = 0.458, and a distance-by-condition interaction, F(3,57)=13.591, p<0.001, part. η^2 = 0.417. With regards to the *positional displacement overall (accuracy)*, we found a significant main effect of condition, F(3,57)=25.573, p<0.001, part. η^2 = 0.574, main effect of distance, F(1,57)=36.179, p<0.001, part. η^2 = 0.388, and a distance-by-condition interaction, F(3,57)=9.619, p<0.001, part. η^2 = 0.336. With regards to the *rotational shift with respect to up vector (rotational accuracy)*, we found a significant main effect of condition, F(3,57)=12.1, p<0.001, part. η^2 = 0.389, and main effect of distance, F(1,57)=8.975, p=0.004, part. η^2 = 0.136. With regards to the *rotational shift with respect to right vector (rotational accuracy)*, we found a significant main effect of condition, F(3,57)=12.692, p<0.001, part. η^2 = 0.4, main effect of distance, F(1,57)=10.357, p=0.002, part. η^2 = 0.154, and a distance-by-condition interaction, F(3,57)=3.127, p=0.033, part. η^2 = 0.141. With regards to the *rotational shift*, we found a significant main effect of condition, F(3,57)=12.393, p<0.001, part. η^2 = 0.395, main effect of distance, F(1,57)=9.66, p=0.003, part. η^2 = 0.145, and a distance-by-condition interaction, F(3,57)=2.931, p=0.041, part. η^2 = 0.134. With regards to the *time ratio of watching at the near-field objects*, we found a significant main effect between conditions, F(3,57)=385.807, p<0.001, part. η^2 = 0.953.

Metrics	Cond1 vs. Cond2	Cond1-Mean	Cond1-SD	Cond2-Mean	Cond2-SD	p-value
Number of Attempts	SN vs. DB	16.4	6.5	39.1	16.4	<0.001
	SN vs. B	16.4	6.5	37.5	18.5	<0.001
	SH vs. DB	22.7	9.9	39.1	16.4	0.007
	SH vs. B	22.7	9.9	37.5	18.5	0.014
Rotational Error Up(deg)	SN vs. DB	10	3.3	5.1	3.4	0.001
	SH vs. DB	12	3.0	5.1	3.4	<0.001
	B vs. DB	8.4	3.6	5.1	3.4	0.034
	SH vs. B	12	3.0	8.4	3.6	0.013
TimeRWN	SN vs. DB	0.52	0.1407	0.92	0.0563	<0.001
	SN vs. SH	0.52	0.1407	0.0041	0.0072	<0.001
	SN vs. B	0.52	0.1407	0.13	0.0751	<0.001
	SH vs. DB	0.0041	0.0072	0.92	0.0563	<0.001
	B vs. DB	0.13	0.0751	0.92	0.0563	<0.001
	SH vs. B	0.0041	0.0072	0.13	0.0751	<0.001

Table 1: Results of Tukey's HSD post-hoc pairwise comparison significant results by condition. TimeRWN stands for Time Ratio of Watching at Near-field replicas; B for BMSR; DB for Direct BMSR; SH for Scaled HOMER; SN for Scaled HOMER+NFSRV; and SD for standard deviation.

Post-hoc pairwise comparisons using Tukey's HSD analysis between conditions are shown in Table 1. Post-hoc pairwise comparisons using the Bonferroni method between distances are shown in Table 2. To examine the interaction effect further, we conducted block analysis on comparing conditions within a distance using Tukey's HSD analysis and between distances within a condition using the Bonferroni method. The results of the post-hoc pairwise comparisons between conditions within each distance (blue and red arrows) and between distances within a condition (yellow arrows) are all illustrated in Fig 6.

Metrics	Near-Mean	Near-SD	Far-Mean	Far-SD	p-value
Movement Time(sec)	112.8	18.1	99.5	15.8	<0.001
Number of Attempts	30.6	6.9	27.2	6.3	<0.001
Positional Error Y(m)	0.043	0.0197	0.05	0.0526	0.015
Rotational Error Up(deg)	8.1	1.7	9.6	2.0	0.004

Table 2: Results of post-hoc pairwise comparisons of target distance using Bonferroni method. SD for standard deviation.

5.2 Quantitative Subjective Results

Non-parametric analysis of the post-experiment quantitative subjective variables revealed significant differences between conditions in the IPQ-Presence questionnaire. In evaluating the scores collected from the IPQ-Presence questionnaire, we found that the condition significantly affected spatial presence (SP) with H(3) = 8.682, p = 0.034 and experienced realism (REAL) with H(3) = 9.202, p = 0.027. Post-hoc Mann-Whitney U test revealed that spatial presence and perceived realism was significantly higher in Direct BMSR as compared to Scaled HOMER, which is illustrated in Figure 7 (a) and (b). No other significant effects were found. There were no significant differences found in NASA-TLX, SSQ and our own system performance questionnaires.

5.2.1 System Usability Scale

In evaluating the scores collected from the system usability scale questionnaire, we found that the condition significantly affected "Easy to use" with H(3) = 8.484, p = 0.037. Using the Mann-Whitney U Test, we found pairwise effects between conditions, which are shown in Figure 8.

There were no significant differences found in NASA-TLX, SSQ and our own system performance questionnaires.

5.3 Qualitative Results

To analyze the qualitative data, we adopted the open coding method mentioned in [34]. We classified the responses by marking keywords. These keywords were defined in two ways: (1) Terms that we defined, for example, 1D Translation, 2D Translation, 3D Translation, etc. (2) Terms found in the responses, for example, "Efficient", "Convenient", "Long-range movement", etc. After marking the keywords, we considered them as trends if there were at least 50% effective responses with the same keyword in that question. Effective responses represented those responses after some comments that were irrelevant to the question had been removed.

In Fig. 9, we show the trends for preference questions. For BMSR, the trend for "What did you like most with regards to rotating an object" is "1D Rotation" (57.1%). The reasons were "It's intuitive to use 1D rotation", "It is consistent with real-world experience", "DOF separation into three different axes is helpful for fine motor control", and "It is fixed at a specific position to prevent unwanted transformations". Regarding "What did you like least with regards to rotating an object", the trend is also "1D rotation" (90%). The reasons are "The handle bar would disappear accidentally", "Sometimes I would forget how to use it", and "I could not apply any axis of rotation I wanted". The trend for "What did you like most with regards to translating and rotating an object simultaneously" is "Just like in the real world" (50%). The reasons are "It feels consistent with real-world experience", "It is flexible", and "I don't need to think in advance". The trend for "What did you like most with regards to moving objects relative to near-field replicas" is "Object observation" (66.7%). The reasons were "I could see it clearly", "It is easier to judge the orientation", and "I feel more control over the selected object". The trend for "What did you like least with regards to DOF separation transformations" was "1D rotation" (54.5%). The reasons were "It does not meet the requirement", "I need to try lots of times to find the solution", and "It is not free for any axis of rotation".

For Direct BMSR, the trend for "What did you like most with regards to rotating an object" was "1D Rotation" (75%). The reasons were "It was intuitive to use the axis of rotation", "Separating from translation helps to do fine motor control", "Dividing into three different axes is helpful for fine motor control", "I could finish tasks accurately." Regarding "What did you like least with regards to moving objects relative to near-field replicas", the trend was "Scaling Mechanism" (53.8%). The reasons were "It was troublesome to cancel and re-select to resize replicas", "Sometimes the replicas would be too large to manipulate", "We cannot choose the size ourselves, which is not convenient." Regarding "What did you like most with regards to selecting near-field replicas as target objects", the trend was "Convenience" (64.3%). The reasons were "It helps to maintain the rhythm of manipulation", "It was very efficient to switch between objects." Regarding "What did you like least with regards to selecting near-field replicas as target objects", the trend was "Limited context range" (50%). The reasons were "Sometimes replicas would disappear accidentally when they are no longer a context to the target replica", "When replicas disappear for being too far away from the target replica, I would lose the sense of direction." Regarding "What did you like most with regards to the platform", the trend is "Multiple perspectives" (57.1%). The reasons were "It was convenient to watch from the back of objects", "It helps us to adjust perspective quickly." Regarding "What did you like least with regards to the platform", the trend was "Difficult to use" (81.8%). The reasons were "The plane is sometimes too low to grab", "When replicas are large, they block the vision of the plane and ring", "It was difficult to use them."

For Scaled HOMER, the trend for "What did you like most with regards to rotating an object" was "Fine motor control" (60%). The reasons are "It reflects the angle of my wrist faithfully". The trend for "What did you like least with regards to rotating an object" was "Difficult to rotate" (78.6%). The reasons were "It was not easy to adjust small rotation degrees", "Not flexible", "It takes me several steps to reach the goal", and "I feel interruption during the rotation".

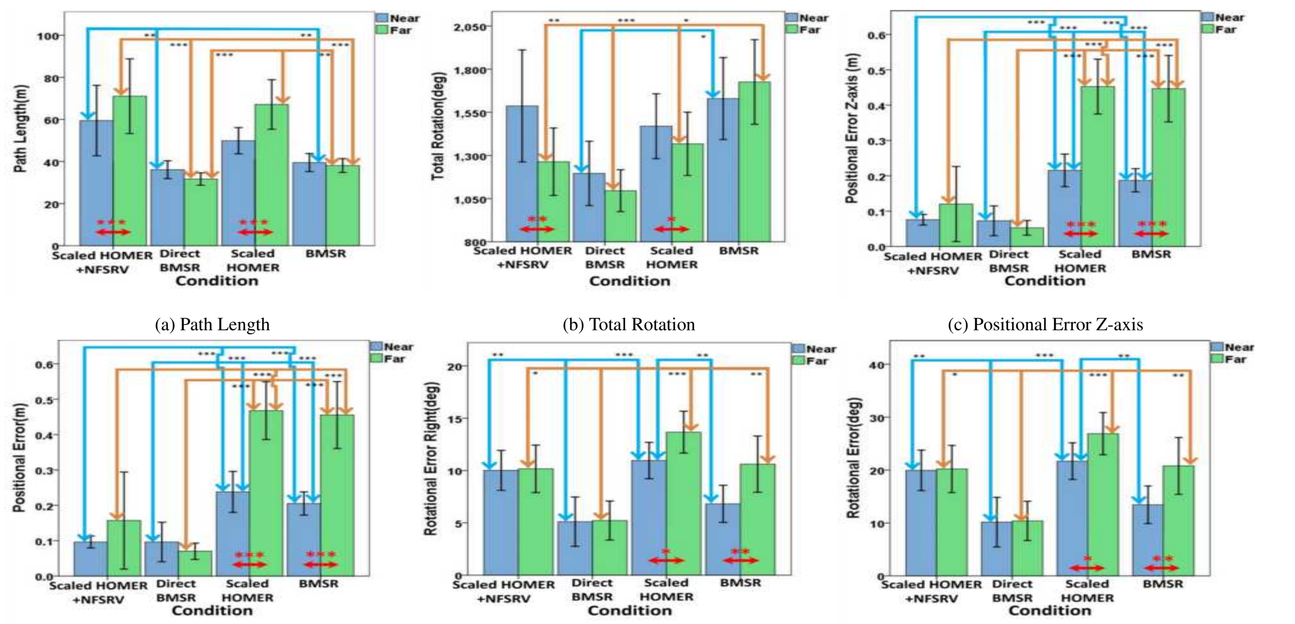
For Scaled HOMER+NFSRV, the trend for "What did you like least with regards to rotating an object" was "Difficult to rotate" (93.0%), with reasons being "Hard to rotate horizontally", "Cannot make slight rotation adjustment", and "Inefficient to rotate along vertical axis". The trend for "What did you like least with regards to translating and rotating objects simultaneously" was "Mixed transformation" (62.0%), with reasons being "Separating translation from rotation is needed when you need to transform as accurately as possible" and "Cannot align the edges well". The trend for "What did you like most with regards to viewing the replicas of the target object and its context in near field" was "Object observation" (68.0%), with reasons being "Viewing replicas helps me to avoid collisions" and "I can check orientation frequently". The trend for "What did you like least with regards to viewing the replicas of the target object and its context in near field" was "Unstable replicas' positions" (69.3%), with reasons being "Not familiar with the initial position of replicas" and "Sometimes the positions of replicas are weird and not helpful for manipulation".

6 Discussion

We operationalized the study's research question, *To what extent do users' performance and preferences differ between the study conditions some of which feature near-field replica viewing and interaction for distant object manipulation?*, via a series of hypotheses. The first hypothesis was that *Direct BMSR is expected to have greater accuracy and economy of movement than BMSR*. This hypothesis was supported by the results, as shown in Fig. 6 and Table 1. Without the ability to move the target replica relative to its context replicas in the near field, BMSR is a distal technique that might suffer from scale-up error, less visual acuity, and inaccurate depth and relative size perception, resulting in a higher error rate than Direct BMSR. However, while both methods allowed users to translate and rotate objects with the same interface, users of BMSR tended to finish trials even when the target objects were placed at incorrect positions, particularly with respect to the z-axis. This suggests that users had difficulty seeing clearly and judging the distance between objects correctly when they were manipulating objects at a distance. Fig. 6(c)(d) also shows that BMSR's accuracy was affected by the depth of the target objects, while Direct BMSR was not. In terms of economy of movement, BMSR required more total rotation than Direct BMSR, indicating that users of BMSR tended to make unnecessary rotations as compared to Direct BMSR.

The second hypothesis was that *Scaled HOMER+NFSRV is expected to have greater accuracy and economy of movement than Scaled HOMER.* This hypothesis was partially supported by the results shown in Fig. 6 and Table 1. There was no significant difference in the rotation error between the two conditions, but Scaled HOMER+NFSRV was significantly more accurate than the Scaled HOMER in terms of position error. This may be due to the fact that the 6DOF translation and rotation offered by both techniques were not well-suited for accurate rotations, which resulted in high rotation error in both conditions. Moreover, to rotate objects along the y-axis, users typically had to try multiple times, which was not ergonomically pleasing as shown by the qualitative results in Section 5.3. In terms of economy of movement, there were no significant differences between the two techniques.

The third hypothesis was that *BMSR is expected to have greater accuracy and economy of movement than Scaled HOMER.* This hypothesis was partially supported by the results as shown in Fig. 6 and Table 1. The position errors of both techniques were similar, but the



(a) Path Length

(b) Total Rotation

(c) Positional Error Z-axis

(d) Positional Error

(e) Rotational Error Right Vector

(f) Rotational Error

Figure 6: Bar graphs of distance-by-condition interaction effects with pairwise post-hoc comparisons embedded. Yellow arrows show the significant post-hoc pairwise differences in viewing distances (near field replica vs. far field object viewing) within an interaction method condition using the Bonferroni method, red arrows show the significant post-hoc pairwise differences between conditions in far field distance viewing of the target object using Tukey's HSD method, and blue arrows show the significant post-hoc pairwise differences between conditions in near-field replica viewing using Tukey's HSD method. The strength of the significant effects are shown using * for $p < 0.05$, ** for $p < 0.01$, and *** for $p < 0.001$. Error bars: 95% confidence interval.

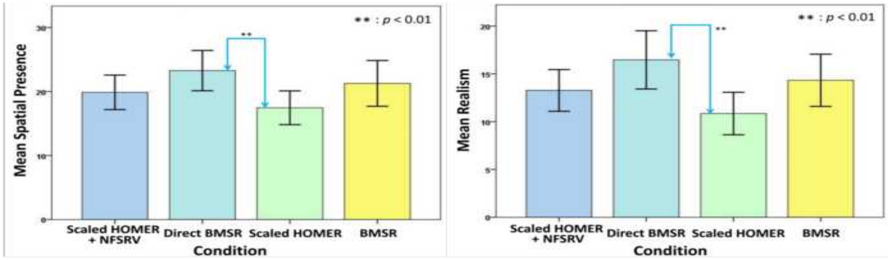
rotation error of BMSR was lower than that of Scaled HOMER for near-distance targets. This result was surprising because it implied that the DOF separation transformations of BMSR did not affect the performance, as much as near-field replica viewing of Scaled HOMER + NFSRV did. With regard to economy of movement, Scaled HOMER had a longer path length, while BMSR had a larger total rotation. These two techniques did not outperform each other.

The fourth hypothesis was that *Direct BMSR is expected to have greater accuracy and economy of movement than Scaled HOMER+NFSRV.* This hypothesis was partially supported by the results shown in Fig. 6 and Table 1. For positioning a target object, both Direct BMSR and Scaled HOMER+NFSRV had performed well in all the tasks. On the contrary, to rotate the target objects to a desired orientation, Direct BMSR outperformed Scaled HOMER+NFSRV for both near- and far- distance trials. For economy of movement, Direct BMSR had a significantly shorter path length and a smaller total rotation for far-distance trials than Scaled HOMER+NFSRV, implying that users made many more unnecessary movements using Scaled HOMER+NFSRV than using Direct BMSR.

The fifth hypothesis was that *Scaled HOMER and Scaled HOMER+NFSRV are expected to have higher efficiency than BMSR and Direct BMSR*. No significant differences in movement time were found, so this hypothesis was not supported. However, we observed that the participants who used BMSR and Direct BMSR often spent much more time figuring out the operations to be applied than the users of Scaled HOMER and Scaled HOMER+NFSRV, implying that task completion time is more appropriate than the movement time for presenting condition's efficiency. Furthermore, in Table 1, we saw that Scaled HOMER and Scaled HOMER+NFSRV had fewer attempts than BMSR and Direct BMSR, implying that with Direct BMSR and BMSR, a complex manipulation task tended to be decomposed into several DOF separation operations.

With regards to the subjective results, we made the following critical observations. As shown in Figure 7, we found that Direct BMSR provided better spatial presence and experienced realism than Scaled HOMER. While in Scaled HOMER, users directly manipulate distant objects distally with less visual acuity due to distant object viewing; in Direct BMSR, users directly manipulated the near-field replicas with potentially higher visual acuity, depth perception and possibly higher attention, which may have enhanced spatial presence and experienced realism. From Figure 8, we observed that Scaled HOMER received lower scores on the System Usability Scale than the other three conditions. With Scaled HOMER, users may have felt frustrated and uncertain potentially due to the inability in the simulated scenarios to see objects clearly for comfortable manipulation at a distance. This result is not surprising, as other relevant research have found similar findings in user studies evaluating distant object interaction techniques [21, 23, 26].

With regard to the qualitative results, we make the following critical observations. First, participants who used Direct BMSR recognized that near-field manipulation helped them perform more accurately and that the feature of selecting context replicas as targets led to convenient manipulation. Moreover, the platform could potentially allow users to view replicas from different perspectives. Second, participants who used Scaled HOMER+NFSRV indicated that near-field viewing helped them perform more accurately. Third, the 6DOF simultaneous translation and rotation of Scaled HOMER and Scaled HOMER+NFSRV was easy and natural to use, but challenging for precise or slight rotation, while the 1D rotation of BMSR and Direct BMSR was well suited for precise or slight rotation but may have required users to determine which axis to be used beforehand. Lastly, the automatic derivation of context size in Direct BMSR and Scaled HOMER+NFSRV, and the positioning of context replicas while the target is moved in Scaled HOMER+NFSRV were challenging to use.



(a) Spatial Presence

(b) Experienced Realism

Figure 7: Boxplots for IPQ-Presence questionnaire. Significant post-hoc pairwise differences are shown using blue arrows. Error bars: 95% confidence interval.

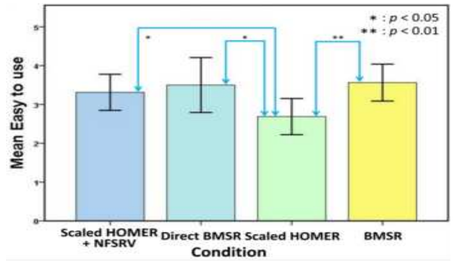
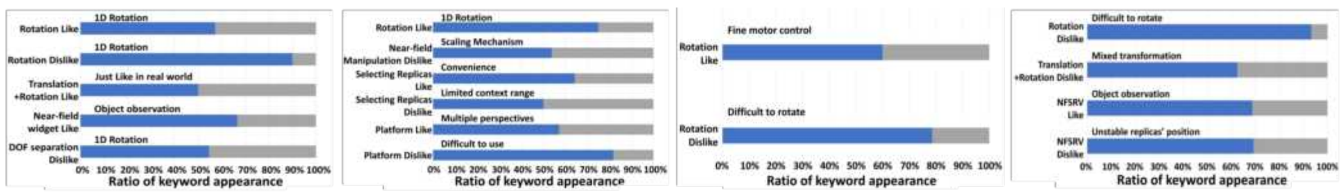


Figure 8: Boxplot for System Usability Scale. Significant post-hoc pairwise differences are shown using blue arrows. Error bars: 95% confidence interval.



(a) Trends for BMSR

(b) Trends for Direct BMSR

(c) Trends for Scaled HOMER

(d) Trends for Scaled HOMER+NFSRV

Figure 9: Qualitative trends in feedback to questions.

Limitations Several problems could be considered limitations of the proposed techniques, and possibly as potential future work. The radius of the sphere representing the context size in both Direct BMSR and Scaled HOMER+NFSRV was automatically determined by taking six times the maximum edge size of the target object's bounding box. However, if the target object is small, the context size will also be small, which may lead to a lack of context objects in the near field to use as a reference for manipulating the small target object. Conversely, if the target object is of considerable size, the size of the context could also increase, potentially resulting in a smaller replica for a context object that is small. However, interacting with this replica may not pose any issues given the proximity of viewing and manipulation. We strive to improve the method for determining the size of context objects in future versions of our technique.

The present implementation does not take into account the terrain or walls of the environment as contextual objects. Therefore, when the target object is moved closer to this type of object, no reference will be shown. In future versions of the system, we can consider labeling these kinds of object and perhaps only showing a part of the object that is within the specified context range in the user's near-field space. The inclusion of terrains or walls as contextual objects will fit well into the current implementation. If the target object is located behind a wall, the Direct BMSR method can be used effectively by following the procedures outlined in Section 3.1. However, the scaled HOMER+NFSRV approach may struggle to select and manipulate the occluded target.

7 Conclusion and Future Work

In this paper, we proposed Direct BMSR (Direct Bimanual Metaphor with Scaled Replica) and Scaled Homer+NFSRV (Scaled HOMER+Near Field Scaled Replica Viewing) and investigated objective performance and subjective impression by conducting a between-subjects user study between four interaction techniques, namely BMSR, Direct BMSR, Scaled HOMER and Scaled HOMER+NFSRV. Our findings revealed that Direct BMSR and Scaled HOMER+NFSRV were significantly more accurate than BMSR and Scaled HOMER, respectively, in terms of both position and rotation errors. From the feedback of the participants, we found that it was troublesome to do rotation with Scaled HOMER and Scaled HOMER+NFSRV, but it was easy to use with the 1D rotation of Direct BMSR and BMSR.

The results of the study suggested that near-field scaled replica manipulation and viewing are helpful in increasing the accuracy, economy of motion, and effectiveness of manipulating distant objects in VR. Applications where the user interacts with the objects in a region of interest from a distance can benefit from the proposed methods, for better vision and depth perception, more accurate manipulation, or better interaction, since the objects in the region of interest can be shown in the near field as replicas and interacted with. Typical examples include manipulation of distant objects, selection in occluded environments, view sharing in remote collaboration, and 3D maps or navigation aids showing part of the virtual worlds.

The idea of manipulating distant objects through distal or near-field scaled replica manipulation was to avoid having to move closer to the target object. However, in large virtual environments, users may have to traverse the scene to locate the objects they wish to manipulate. A more comprehensive interface for object manipulation would include global search and navigation tools, such as steering or teleportation, which would enable the user to move to the right spot in the environment and then manipulate the targets using distal or near-field scaled replicas manipulation techniques from a distance. The Direct BMSR transformation platform can facilitate long-distance translation; however, it can be troublesome to use. Consequently, a combination of Direct BMSR and Scaled HOMER could be a potential solution to resolve the problem and could be suitable for applications that necessitate different levels of precision.

Acknowledgments

This work was partly supported by the National Science and Technology Council, R.O.C., under Grant No. NSTC 112-2221-E-A49-111 and the US National Science Foundation (CISE IIS HCC) under Grant No. 2007435.

References

[1] A. Bangor, P. T. Kortum, and J. T. Miller. An empirical evaluation of the system usability scale. *Intl. Journal of Human–Computer Interaction*, 24(6):574–594, 2008.

[2] A. Bhargava, J. W. Bertrand, A. K. Gramopadhye, K. C. Madathil, and S. V. Babu. Evaluating multiple levels of an interaction fidelity continuum on performance and learning in near-field training simulations. *IEEE transactions on visualization and computer graphics*, 24(4):1418–1427, 2018.

[3] B. Bossavit, A. Marzo, O. Ardaiz, L. D. De Cerio, and A. Pina. Design choices and their implications for 3D mid-air manipulation techniques. *Presence: Teleoperators and Virtual Environments*, 23(4):377–392, 2014.

[4] D. Bowman, E. Kruijff, J. J. LaViola Jr, and I. P. Poupyrev. *3D User Interfaces: Theory and Practice*. Addison-Wesley, 2004.

[5] D. A. Bowman and L. F. Hodges. An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In *Proceedings of the Symposium on Interactive 3D Graphics*, pp. 35–38, 1997.

[6] D. A. Bowman, R. P. B. McMahan, and E. D. Ragan. Questioning naturalism in 3D user interfaces. *Communication of the ACM*, 55(9):78–88, 2012.

[7] D. Brickler, R. J. Teather, A. T. Duchowski, and S. V. Babu. A fitts' law evaluation of visuo-haptic fidelity and sensory mismatch on user performance in a near-field disc transfer task in virtual reality. *ACM Transactions on Applied Perception (TAP)*, 17(4):1–20, 2020.

[8] D. Brickler, M. Volonte, J. W. Bertrand, A. T. Duchowski, and S. V. Babu. Effects of stereoscopic viewing and haptic feedback, sensory-motor congruence and calibration on near-field fine motor perception-action coordination in virtual reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 28–37. IEEE, 2019.

[9] I. Cho and Z. Wartell. Evaluation of a bimanual simultaneous 7DOF interaction technique in virtual environments. In *Proceedings of 2015 IEEE Symposium on 3D user interfaces*, pp. 133–136, 2015.

[10] A. Coh́e, F. Décle, and M. Machet. tbox: A 3d transformation widget designed for touch-screens. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 3005–3008, 2011.

[11] B. R. De Araújo, G. Casiez, J. A. Jorge, and M. Hachet. Mockup builder: 3D modeling on and above the surface. *Computers & Graphics*, 37(3):165–178, 2013.

[12] F. El Jamiy and R. Marsh. Survey on depth perception in head mounted displays: distance estimation in virtual reality, augmented reality, and mixed reality. *IET Image Processing*, 13(5):707–712, 2019.

[13] S. Frees and G. D. Kessler. Precise and rapid interaction through scaled manipulation in immersive virtual environments. In *IEEE Conference on Virtual Reality*, 2005.

[14] S. Frees, G. D. Kessler, and E. Kay. PRISM interaction for enhancing control in immersive virtual environments. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 14(1), 2007.

[15] P. C. Gloumeau, W. Stuerzlinger, and J. Han. PinNPivot: Object manipulation using pins in immersive virtual environments. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2020. doi: 10.1109/TVCG.2020.2987834

[16] Y. Guiard. Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model. *Journal of motor behavior*, 19(4):486–517, 1987.

[17] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, vol. 52, pp. 139–183. Elsevier, 1988.

[18] S. Houde. Interactive design of an interface for easy 3D direct manipulation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 135–142, 1992.

[19] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3):203–220, 1993.

[20] M. S. Landy, M. S. Banks, and D. C. Knill. Ideal-observer models of cue integration. *Sensory cue integration*, pp. 5–29, 2011.

[21] C.-Y. Lee, W.-A. Hsieh, D. Brickler, S. V. Babu, and J.-H. Chuang. Design and empirical evaluation of a novel near-field interaction metaphor on distant object manipulation in VR. In *ACM Symposium on Spatial User Interaction, 2021 (SUI 2021)*, 2021.

[22] D. P. Mapes and J. M. Moshell. A two-handed interface for object manipulation in virtual reality. *Presence: Teleoperators & Virtual Environments*, 4(4):403–416, 1995.

[23] D. Mendes, F. M. Caputo, A. Giachetti, A. Ferreira, and J. Jorge. A survey on 3D virtual object manipulation: From the desktop to immersive virtual environments. *Computer Graphics Forum*, 38(1):21–45, 2019.

[24] D. Mendes, F. Fonseca, B. Araujo, A. Ferreira, and J. Jorge. Mid-air interactions above stereoscopic interactive tables. In *IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 3–10. IEEE, 2014.

[25] D. Mendes, F. Relvas, A. Ferreira, and J. Jorge. The benefits of DOF separation in mid-air 3D object manipulation. In *ACM Conference on Virtual Reality Software and Technology*, pp. 261–268, 2016.

[26] D. Mendes, M. Sousa, R. Lorena, A. Ferreira, and J. Jorge. Using custom transformation axes for mid-air manipulation of 3D virtual objects. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, pp. 1–8, 2017.

[27] M. Mine. Isaac: A virtual environment tool for the interactive construction of virtual worlds. *UNC Chapel Hill Computer Science Technical Report TR95-020*, 1995.

[28] M. Mine, F. Brooks, and C. Sequin. Moving objects in space: Exploiting proprioception in virtual environment interaction. In *Proceedings of SIGGRAPH"97*, pp. 19–26, 1997.

[29] J. S. Pierce, B. C. Stearns, and R. Pausch. Voodoo dolls: seamless interaction at multiple scales in virtual environments. In *Proceedings of the 1999 Symposium on Interactive 3D graphics*, pp. 141–145, 1999.

[30] I. Poupyrev, M. Billinghurst, S. Weghorst, and T. Ichikawa. 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.

[31] T. W. Schubert. The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness. *Z. für Medienpsychologie*, 15(2):69–71, 2003.

[32] P. Song, W. B. Goh, W. Hutama, C.-W. Fu, and X. Liu. A handle bar metaphor for virtual object manipulation with mid-air interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, p. 1297–1306, 2012.

[33] R. Stoakley, M. J. Conway, and R. Pausch. 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] A. Strauss and J. Corbin. *Basics of qualitative research techniques*. Thousand oaks, CA: Sage publications, 1998.

[35] M. Veit, A. Capobianco, and D. Bechmann. Influence of degrees of freedom's manipulation on performances during orientation tasks in virtual reality environments. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*, VRST '09, p. 51–58. Association for Computing Machinery, New York, NY, USA, 2009. doi: 10.1145/1643928.1643942

[36] C. Wilkes and D. A. Bowman. Advantages of velocity-based scaling for distant 3D manipulation. In *Proceedings of the ACM symposium on Virtual Reality Software and Technology*, pp. 23–29, 2008.