

Object Impersonation: Towards Effective Interaction in Tablet- and HMD-Based Hybrid Virtual Environments

Jia Wang*

HIVE Lab

Worcester Polytechnic Institute

Robert W. Lindeman†

HIVE Lab

Worcester Polytechnic Institute

ABSTRACT

In virtual reality, hybrid virtual environment (HVE) systems provide the immersed user with multiple interactive representations of the virtual world, and can be effectively used for 3D interaction tasks with highly diverse requirements. We present a new HVE metaphor called *Object Impersonation* that allows the user to not only manipulate a virtual object from outside, but also become the object, and maneuver from inside. This approach blurs the line between travel and object manipulation, leading to efficient cross-task interaction in various task scenarios. Using a tablet- and HMD-based HVE system, two different designs of Object Impersonation were implemented, and compared to a traditional, non-hybrid 3D interface for three different object manipulation tasks. Results indicate improved task performance and enhanced user experience with the added orientation control from the object's point of view. However, they also revealed higher cognitive overhead to attend to both interaction contexts, especially without sufficient reference cues in the virtual environment.

Keywords: Hybrid virtual environments, cross-task interaction, 3D user interface, tablet interface, virtual reality.

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—artificial, augmented, and virtual realities; H.5.2 [Information Interfaces and Presentation]: User Interfaces—evaluation/methodology, input devices and strategies, interaction styles.

1 INTRODUCTION

The popularity of immersive Virtual Reality (VR) technology has been booming recently thanks to a new generation of low-cost Head-Mounted Displays (HMD). In addition to the high fidelity sensory feedback, various realistic 3D User Interfaces (3DUIs) have also been developed, allowing an immersed user to naturally walk [22] or fly [29] in a virtual world, and grab and manipulate virtual objects using his/her hands [21]. However, in spite of the realistic feeling, researchers also realize that interaction in VR can be just as confusing, limiting, and ambiguous as in the real world [25], especially for tasks with highly diverse requirements [31]. For example, it is difficult to select and manipulate virtual objects of different sizes, from multiple angles, and at different distances to the user's virtual avatar, without spending significant time and effort on navigation [6].

One way to overcome such limitations is through the use of Hybrid Virtual Environment (HVE) systems. In an HVE system,

the same virtual world is represented in multiple heterogeneous display contexts, each of which can be interacted with using different virtual and/or physical interface elements [31]. Examples of HVE metaphors include the World-In-Miniature (WIM) that allows a user to interact with the surrounding virtual environment (VE) from both inside and outside his/her virtual body [25], the Voodoo Doll that enables direct manipulation of both nearby and faraway objects [18], the virtual portal that connects local with remote spaces [24], and the see-through lens that visualizes both the internal and external features of virtual objects [27]. Despite this variety, all of these metaphors assume that the user possesses a virtual body, and acts on other virtual objects from an exocentric (outside) point of view.

We present a new HVE metaphor which enables an immersed user to not only grab and manipulate a virtual object from outside, but also become the object, and maneuver from inside of it. For example, by impersonating a virtual spotlight, one can efficiently change its location by travelling around the space, and precisely illuminate a target area by turning and looking at it. In other words, object impersonation has the potential to turn complex object manipulation tasks into intuitive travel tasks. This has been referred to as cross-task interaction in Bowman's doctoral dissertation [4], but has not been formally implemented or studied, particularly in the context of HVEs. In this paper, we present six use cases of object impersonation, two different interface implementations in a tablet- and HMD-based HVE system, and results of a user study that evaluated efficiency and user experience in three object-target alignment tasks.

2 RELATED WORK

2.1 3D User Interaction

One oft-mentioned benefit of immersive VR is its affordance of body-centered natural 3D interaction [6]. The basic 3D interaction tasks in VR have been categorized into travel, way-finding, selection, manipulation, system control, and symbolic input [6]. Travel refers to the process of changing one's position and orientation in the 3D space in order to gain different perspectives of the surrounding VE. The task can also be reduced to the specification of a 3D vector and a speed value, which can be fulfilled by pressing a button while looking at, pointing at, or facing the intended destination [6]. Real walking is also possible, but can be difficult to implement due to limited tracking space in the real world [22].

Object selection and manipulation are usually bundled to complete the process of picking up an object and editing its spatial properties including position, orientation, and scale [6]. The two most common interaction techniques for object selection and manipulation are based on ray-casting, using a wand, and a virtual hand, using a data glove [21], while hybrid approaches such as the HOMER interaction technique also exist and combine the benefits of both [5]. One main challenge for object manipulation in 3D space is our inability to reason about 3D rotations, which is usually alleviated by experimenting from different angles [32].

*e-mail: wangjia@wpi.edu

†e-mail: gogo@wpi.edu

Since travel and manipulation both require changing an object's position and orientation (the user's avatar being the object in travel), Bowman speculated about unifying the two tasks, and creating an experience of cross-task interaction in immersive VR [4]. For example, the user can choose to become an object in the VE, and translate and rotate the object by simply moving and looking around. This approach breaks the basic assumption of VR that every user possesses a unique virtual self, but has great potential in certain application scenarios. However, to the authors' knowledge, it has never been formally implemented and studied, especially in the context of HVEs.

2.2 Hybrid Virtual Environment

The unique advantage of discussing object impersonation in the context of HVEs is that one can still preserve the benefits of egocentric, body-centered interaction. The early work of HVEs proposed several different metaphors that can define the relationship between the multiple interaction contexts (ICs). These include the World-In-Miniature [25], the Voodoo Dolls [18], the SEAM [24], and the see-through lens [27]. Compared to traditional, single context VR setups, these systems can offer the user more power and flexibility to handle tasks with diverse requirements, such as manipulating objects in different scales, distances, reference frames, and dimensions [31].

Based on these metaphors, various HVE implementations have been proposed. For example, the SCAPE system puts a see-through workbench display in the center of a room with projection walls [7]. The projection walls form a CAVE environment for immersive interaction, while the workbench shows a WIM view of the same VE that can be navigated using a magic lens interface. The HybridDesk system features a similar setup, in which a traditional desktop computer is embedded in the space of a desktop CAVE display [8]. When a remote virtual object is selected in the CAVE context, a Voodoo Doll version of it appears on the computer screen, and can be manipulated and annotated using a wand, a mouse, or a keyboard device. A recent HVE development took advantage of the advanced interactivity and computing power of modern tablet devices [31]. In the HVE level editor, a WIM representation of the VE is rendered on a tablet, which can be viewed from the peripheral vision of a non-occlusive HMD, and interacted with using 2D GUI and multi-touch gestures. Inspired by research in coordinated multi-view visualization systems [28], the basic 3D interaction tasks were synchronized between the tablet and immersive ICs, in order to reduce the cognitive overhead involved in the context-switching process. In this paper, we apply the object impersonation metaphor to the tablet- and HMD-based HVE setup, and study its effectiveness in three different object manipulation tasks.

2.3 Tablet-Based Interfaces

Interactive tablet devices have had successful application in HVEs, thanks to the affordance of natural bimanual interaction [11] and passive haptic feedback [14]. Early pen-and-tablet prototypes displayed an interactive 2D map on a tracked touchpad to aid way-finding and travel in the immersive context [1]. The personal interaction panel [26] and the virtual notepad [20] demonstrated object selection and manipulation, system control, and symbolic input tasks offloaded to the 2D surface, through displayed 3D widgets and 2D GUI that could be controlled using a stylus. See-through interaction is also possible using a transparent prop with a workbench display [23]. The recent research trend of tablet-based interface uses mobile tablet and phone devices as an integrated solution of computing, display, and touch-based interaction. For example, Bornik *et al.* demonstrated an HVE system which combined a projection screen and a tablet PC, and used spatial stylus input to seamlessly control contents in

both contexts [3]. Finally, tablet and phone devices can also be bundled with spatial tracking sensors, and used as object manipulation tools [15], or see-through lens interfaces [16].

3 METHODOLOGY

3.1 Object Impersonation

We define *object impersonation* as an interaction technique that enables an immersed user to select an object in the virtual world as his/her virtual self, and view, move, and interact with other objects from its point of view. Generally speaking, impersonating a different virtual object can cause various changes in view point location, orientation, field of view, body scale, and reference frame, and mappings between the user's body motion and his/her avatar actions. As our first exploration of this paradigm, the discussion in this paper is limited solely to view point position and orientation changes, as well as reference frame changes of the virtual object.

Object impersonation can be implemented as a transitional user interface by allowing the user to jump in and out of his/her avatar [2], or used as a metaphor to define the relationship between ICs in an HVE system [31]. The discussion in this paper is focused on the latter scenario, where a user is given two ICs, one using the traditional, avatar-based approach, while the other is based on the perspective and reference frame of the selected object. Being the object, the user is still able to perform the same 3DUI tasks (i.e., travel, way-finding, selection, manipulation, etc.), thereby supporting effectiveness in the following application scenarios:

- **Remote space inspection (way-finding):** Using object impersonation, an enhanced version of Worldlets [10] can be implemented, which allows a user to navigate and inspect remote spaces without having to travel there him/herself. By jumping between objects at different geo-locations, the views of each object's surrounding environment can be connected to accumulate survey knowledge [9] of a large VE relatively quickly.
- **Avatar transportation (travel):** From the object's point of view, the user can also drag and drop his/her virtual avatar to locations in nearby space. This enables quick and accurate transportation, and can be helpful for collaboration, or tasks with distributed goals (e.g., annotation of landmarks). However, certain awareness cues may be necessary to highlight the spatial relationship between the multiple views [31], as seeing one's previous avatar in his/her current view may cause disorientation.
- **Occlusion-free object selection (selection):** Selecting objects in cluttered virtual space can be difficult due to the large amount of occlusion in the scene. Applying object impersonation in two different ways can alleviate this challenge. First, the user can select and impersonate an object to the side of the occluded space, offering an orthographic view to complement the current perspective [19]. Second, the user can even become the occluding object itself, and use its perspective as a see-through lens [16] to select the objects behind it. Using these approaches, the amount of travel needed to gain different viewing angles of the VE can be effectively reduced to one click of a button.
- **Multi-perspective object manipulation (manipulation):** Like the previous use case, object impersonation can also be utilized to enable object manipulation from two orthographic perspectives. Similar approaches have been shown to be effective in collaborative virtual environments for a variety of cooperative object-manipulation tasks [19]. In addition, objects at high elevation can be impersonated to gain a God view of the VE, offering the user a WIM-like interface [25] to ease large-scale manipulation tasks.

- **Object-target alignment (manipulation):** The previous use cases all focused on what the user can do to *other objects* from the impersonated object's perspective. The user can also affect the impersonated object *itself*, by simply looking, turning, and moving around from its frame of reference. This approach can be used to simplify object manipulation tasks in which the goals are related to the object's view. For example, the user can impersonate a spotlight, and simply look at the target to accurately illuminate its surrounding area. This crosses the 3DUI tasks of travel and manipulation, implying an interesting What-I-Sec-Is-What-I-Do (WISIWID) metaphor.
- **Path editing (manipulation):** In addition to looking around, the user can also travel around the VE using the object as his/her virtual self. Opposite to the path-drawing technique used for navigation [13], a Where-I-Go-Is-What-I-Do (WIGIWID) metaphor can be implemented, letting the user impersonate a brush to draw a 3D spline, or the front of a train to lay out a roller coaster in the VE. Compared to a traditional interface such as a 3D stylus, this object egocentric approach can make it easier to draw a spline across multiple anchor points, especially through cluttered or enclosed spaces.

It should be mentioned that despite the advantages listed above, object impersonation also has its limitations, so one should not rely on it completely for all 3DUI tasks. For example, directing the orientation of a spotlight may be easier from its own point of view, but setting its position can be difficult without seeing it from a third-person view. Similarly, a third-person view is necessary to keep track of the overall layout of the spline, even though passing through the anchor points can be easily done by being the brush itself. Fortunately, the advantages of object impersonation and traditional avatar-based approaches appear to complement each other's drawbacks in many aspects. Therefore, we propose a hybrid solution based on an HVE system, and expect it to combine the strengths of both techniques, to offer effective cross-task 3D interaction in immersive VR. This paper specifically studies this methodology using the object-target alignment task as a test bed.

3.2 System Development

To investigate the effectiveness of object impersonation, an HMD- and tablet-based HVE system was developed. As shown in Figure 1, the HVE system consists of two ICs. The immersive IC uses an eMagin z800 HMD as the display device, and a Wii Remote controller-based wand interface as the input device. Both the HMD and the wand have a six degrees-of-freedom (DOF) LED tracking constellation attached, which can be tracked by sixteen PhaseSpace motion capture cameras mounted on a frame surrounding the user. Seated on a swivel chair, the user can freely turn his/her head to look around the VE, point the wand at virtual objects, and press buttons on the Wii Remote controller to select and manipulate them in 3D space (Figure 1b). The tablet IC is implemented using a Google Nexus-7 Android tablet, which is placed on an armrest table on the non-dominant hand side of the user. Since the HMD is non-occlusive, the user can view the VE rendered on the tablet screen from his/her peripheral vision, and perform 3D interaction using multi-touch gestures (Figure 1c).

On the software side, the HVE system was developed using the Unity game engine as a multi-player game running on two different platforms. The input data from the motion tracking system and the Wii Remote controller are streamed to a desktop PC through VRPN and the Unity Indie VRPN Adapter (UIVA) [31]. Both the desktop PC and the tablet simulate the same virtual world locally at a steady 30 frames per second. By sending UDP data streams and RPC calls over a local WiFi network, the effects of interaction in one IC can be propagated to the other IC in real time, giving the user a convincing feeling that they are looking

and interacting with the same VE, only from two different perspectives.

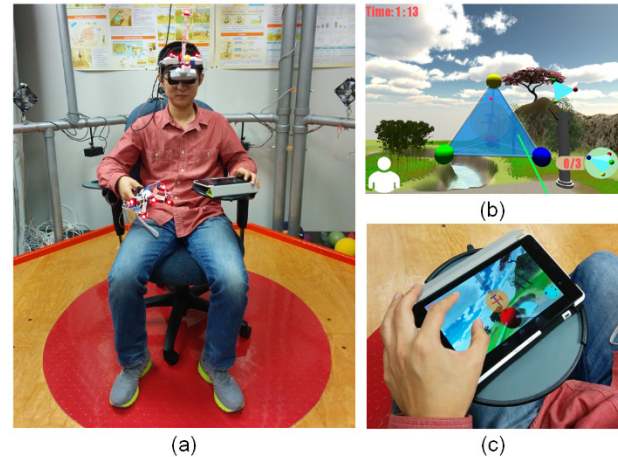


Figure 1: The hardware setup of the HVE system; (b) the avatar's view on the HMD; (c) the object's view on the tablet.

3.3 Interaction Tasks

As discussed in Section 3.1, the object-target alignment task was selected as the test bed to evaluate object impersonation for cross-task 3D interaction in HVEs. To gain an in-depth understanding of all task scenarios, three different object-target alignment tasks were implemented. As shown in Figure 2a, the spotlight task asked the user to translate and rotate a spotlight, in order to have it placed in the position of a street lamp, and oriented to illuminate a text plate. One efficient hybrid strategy, as speculated by the authors, is to first drag the spotlight to its destination using an avatar-based third-person view interface, and then to impersonate the spotlight, and illuminate the text plate by simply looking at it.

The spotlight task presents a special case of object manipulation in VR. More generally, the impersonated object may neither feature a shape similar to the viewing frustum, nor afford a visual indicator (the light) to naturally connect the goal of the task to the style of the first-person view. Therefore, a second task is illustrated in Figure 2b, which asks the user to translate and rotate a house in 6-DOF, in order to have it stand on the ground, and face another house door to door. Without the visual cues, the advantages of object impersonation in this task may not be as significant as in the spotlight task. However, the user may still find it helpful to level the house on the ground, or determine its alignment with the other house.

To facilitate controlled comparison with traditional 3D interfaces, a further generalized object-docking task was developed, following the classic object manipulation task proposed by Zhai [32]. As shown in Figure 2c, this task requires the user to manipulate a tetrahedron in 6-DOF, and match it with another tetrahedron with arbitrary position and orientation. To avoid ambiguity of the orientation matching, a uniquely colored sphere is attached to each vertex of the tetrahedron. Using object impersonation, the user can become the tetrahedron, and change its position and orientation by moving and looking around the VE, respectively. To make the task goal visible in the object's view, a crosshair was added to both tetrahedra, which can be matched to align their orientations. This approach reduces the overhead of mentally rotating the tetrahedron by separating the interrelated 3-DOF object rotation control to the combination of a 2-DOF looking action (i.e., crosshair translation) and a 1-DOF rolling action (i.e., crosshair rotation), and is expected to enhance user performance and experience in comparison to traditional, non-hybrid spatial input interfaces.

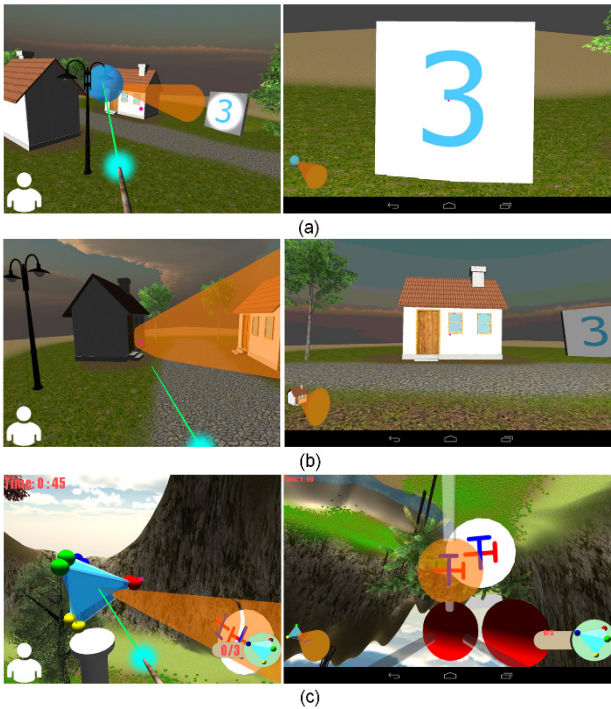


Figure 2: Alignment tasks from the avatar's view on the HMD (left), and the object's view on the tablet (right): (a) the spotlight task, (b) the house task, and (c) the tetrahedron task.

3.4 Interface Design

Within the aforementioned HVE system, two object impersonation modes, VIEW and DRIVE, were implemented to support the object-target alignment tasks. The main differences between these two modes, from a user's standpoint, are the depth of immersion in, and the degree of control over, the impersonated object. As shown in Figure 2c, the Object View Impersonation (VIEW) mode displays the view of the impersonated object on the screen of the tablet, leaving the HMD to the traditional avatar-based immersive model. Object translation in the immersive IC is realized using a combination of virtual hand and ray-casting techniques. Ray-casting-based translation is triggered when the user points the wand at the tetrahedron and presses down the "B" button on the Wii Remote controller. The tetrahedron will follow the movement of the wand at the original hit point, while two fishing rod buttons can be used to move it further or closer, along the direction of the ray [5]. Virtual-hand-based translation starts when the user points at the tetrahedron and presses down the "Home" button. The tetrahedron will then follow the position change of the user's hand, allowing more accurate position control over a much smaller range. This hybrid control approach combines position control with rate control [30], allowing the user to match the targets both quickly and precisely. In the tablet IC, the user can see the tetrahedron's first-person view, and look around using a single-finger swipe gesture, or roll the view using a two-finger rotation gesture. Since the tetrahedron's viewing frustum is fixed to its body, changing its first-person view will also affect the object's orientation. Therefore, by moving the tetrahedron onto the target, and matching the two crosshairs, the orientation of the objects can be roughly matched. The result can then be perfected by micro-adjusting the tetrahedron's position using the wand interface, until a "Right There!" text is shown on the screen to indicate the completion of the task.

Figure 3 illustrates the object drive impersonation (DRIVE) mode. The tablet screen is used to display a third-person view looking towards one vertex of the tetrahedron from behind,

allowing the user to use the HMD to gain deeper immersion in the tetrahedron itself. The experience in the immersive IC is similar to driving a spacecraft from the inside, with the tetrahedron being the spacecraft, and following pointing-directed locomotion of the user. Rotations around the up- and right-axes (i.e., yaw and pitch) are realized by turning the head. To avoid straining the neck, rolling is implemented by pressing down two buttons on the wand, with one being clockwise and the other counter-clockwise. It should also be mentioned that pressing these two buttons will only rotate the tetrahedron; the immersive view is always kept upright, in order to prevent disorientation and motion sickness induced by looking at the VE upside-down. To match the tetrahedron with the target in DRIVE mode, a three-step procedure is suggested. The first step is to drive the tetrahedron to the center of the target object, which sets up a base point to align the crosshairs. The user can then hold down the "B" button, and turn his/her head to find and match the reference crosshair, which will match the orientation of the tetrahedrons as well. Finally, the user switches to the tablet, and uses one-finger swipe and two-finger pinch gestures to precisely match the positions of the tetrahedrons. It should be mentioned that the last two steps may need to be repeated, depending on the precision of the initial position match in the first step.



Figure 3: The tetrahedron docking task in the object drive impersonation (DRIVE) mode, from (a) the avatar's view on the tablet and (b) the tetrahedron's view on the HMD.

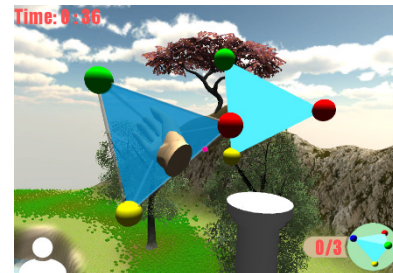


Figure 4: The virtual hand technique used to rotate the tetrahedron from the avatar's view on the HMD.

To evaluate object impersonation, a traditional, non-hybrid 6-DOF manipulation interface was implemented as a control condition, using the wand device only (WAND). The interaction technique adopted is similar to HOMER [4]. Translation control is the same as in VIEW mode, with the "B" button dedicated to enabling ray-casting-based object dragging, and the "Home" button used to trigger virtual hand-based accurate position adjustment. Instead of turning the object's first-person view, object rotation in WAND mode is done by holding down the "A" button, and directly rotating the wand device, as shown in Figure 4. Clutching is supported by releasing and repressing the "A" button. Furthermore, since there is only one object to control in the tetrahedron task, the user can start rotating it as soon as the button is pressed down, without having to point the wand at the object first. These two settings compensate for the physical constraint of the wrist, giving the user more freedom and flexibility to operate the wand interface effectively [12].

4 USER STUDY

4.1 Hypotheses

Object impersonation offers the user a cross-task approach to perform 3D manipulation tasks based on the target object itself. As proposed in Section 3.1, this metaphor can benefit many task scenarios where traditional 3DUIs fall short, such as the object-target alignment task selected as the study test bed in this paper. However, the authors also believe that despite its advantages, object impersonation also has its limitations, and should not be used to *replace* traditional, avatar-based 3DUI techniques, but rather to *supplement* them. The HVE system thus offers a hybrid solution to combining the benefits of both approaches, allowing the user to select and drag the object from the outside, as well as maneuvering its orientation from the inside. We feel that the WAND interface does offer a more realistic simulation of object rotation, and integrates all interaction in one single IC. Based on these analyses, we make the following hypotheses:

- **H1:** Users will spend less time completing the tetrahedron docking task in the VIEW and DRIVE modes.
- **H2:** Users will feel the WAND interface to be more intuitive and natural to understand and learn.
- **H3:** Users will find the VIEW and DRIVE modes to be more efficient and precise, and easier and less tiring to use.
- **H4:** The mental rotation skill required to manipulate the object in 6-DOF will be lower in the VIEW and DRIVE modes.
- **H5:** Higher cognitive overhead will be induced on the user when multiple ICs are involved in the VIEW and DRIVE modes.

4.2 Procedure

To validate these hypotheses, a within-subjects user study was designed and conducted. The study was approved by the institutional review board (IRB), and 26 university students were recruited with no remuneration. Each session began with the subject reading and signing a consent form, followed by a demographic questionnaire that asked about gender, age, and handedness, as well as experiences with video games, 3D modeling software (e.g., Maya, SketchUp), immersive VR, multi-touch devices, and multi-screen devices (e.g., the Nintendo WiiU). The subject was then asked to complete Peters' redrawing of Vandenberg & Kuse Mental Rotation Test (MRT), which presented 24 questions with a time limit of 10 minutes [17]. After the MRT, the experimenter gave the subject a brief introduction to the hardware used in the study, including the HMD, the wand, and the tablet. The experimenter also explained the details of the three object-target alignment tasks, especially the tetrahedron docking task, which served as the primary task to compare the efficiency of the WAND, VIEW, and DRIVE interfaces.

After the introduction, the subject put on the equipment, and completed the tetrahedron docking task using each of the three interfaces, following a counterbalanced order based on a Latin square. Each of the three conditions included a training session and an experiment session, in which the same VE was used. As shown in Figure 5, the VE included three tetrahedral targets in different positions and orientations. The subject was asked to practice the specific interface in each training session, by matching the three targets one after another. In the experiment sessions, the subject was asked to match up to three rounds (nine trials) of the same targets, within a time limit of 10 minutes. At the beginning of a session, the subject's avatar was spawned in the center of the VE, together with a semi-transparent tetrahedron object floating right in front of him/her. The subject could then use this tetrahedron to match the targets one by one, as quickly as possible. The distances between each pair of the colored spheres were calculated, and were compared to a threshold variable d to

determine whether the tetrahedrons had been matched. When the threshold was reached, a "Right There!" text would show up on both screens to indicate a match. The subject could then let go of the control, and wait for the current target to disappear, and the next target to appear, in three seconds. This process was repeated three times in training (one round, with $d = 0.8\text{m}$ and tetrahedron's edge length = 5m), and up to nine times (three rounds, with $d = 0.8\text{m}$, 0.4m , and 0.2m), or 10 minutes in the experiment sessions. The experiment sessions had increased precision requirements with each round, in order to inspect the effects of the interfaces on task precision. During the experiment, a timer was displayed in the top-left corner, and a target counter was shown in the bottom right, on both screens. The crosshair plates accompanying each target as shown in Figure 5a were only made visible in the VIEW and the DRIVE conditions. They indicated the targets' first-person views, and were used to aid rotation alignment from the impersonated object's perspective. After completing all three conditions, the subject was asked to fill in a questionnaire to compare the WAND, VIEW, and DRIVE interfaces, and to rate them on a one to six scale regarding six different questions (see Figure 7, discussed later). The subject was also asked to indicate his/her general preference for the three interfaces, and provide comments on what they liked and disliked about each of them. Lastly, to expand the investigation to real world applications, the house and spotlight tasks were also included in the study. However, instead of being formally evaluated, they were only tested in a short session after the tetrahedron experiment. The subjects casually selected the houses and spotlights in a VE, and tried each aforementioned interface to align them with their targets. During the process, the experimenter kept an active conversation with the subject, so that he/she could give anecdotal comments on the go about the advantages and drawbacks of each interface for the two tasks.



Figure 5: The task VE of the tetrahedron docking task.

Of the 26 participants, 14 were males and 12 were females. All subjects were right-handed. Their ages ranged from 19 to 31 years (mean=23.9, SD=3.1). With 1 being "Never" and 6 being "Every day," their experiences with video games ranged from 1 to 6 (mean=3.2, SD=1.5), 3D modeling software ranged from 1 to 4 (mean=1.9, SD=1.0), VR ranged from 1 to 3 (mean=1.4, SD=0.6), multi-touch devices from 1 to 6 (mean=5.7, SD=1.0), multi-screen devices from 1 to 6 (mean=2.8, SD=1.7). Their responses to the MRT test were also graded. With 24 being the maximum score, their answers ranged from 7 to 24 (mean=14.9, SD=4.9).

5 RESULTS

5.1 Task Performance

For each experiment session, the system recorded how many targets were successfully completed in 10 minutes, as well as the exact time stamp when each target was matched. The numbers of completed targets of the three interface conditions were compared using a Friedman test, however the results were not significant.

Since many subjects were able to match all nine targets before the time expired, a more accurate indicator of task efficiency was needed. To do this, for each subject, we averaged the time he/she spent to match the targets, for all targets collected by the subject, as well as targets in the first round, or the second round alone (all subjects were able to complete all three targets in the first round, and at least one target in the second round). The seconds-per-target data produced from this process was analyzed using a one-way ANOVA, and the results are shown in Figure 6. Although no results are strictly significant (i.e., $p < 0.05$), statistical trends towards significance were evident in all of them (i.e., $p < 0.1$), suggesting further post-hoc investigation. Using Tukey HSD test, we found trends suggesting better efficiency in DRIVE mode than the WAND interface, for all targets in general ($p = 0.074$), and low-precision-requirement targets in the first round ($p = 0.085$). Additionally, a trend was also identified indicating better efficiency in VIEW mode than using the WAND interface, for the second-round targets that required medium-precision matching ($p = 0.061$). Finally, a Pearson correlation analysis was performed between these task performance measurements and the subject's prior experiences and mental rotation skills. However, no strong correlation was discovered (all correlation coefficient values are below 0.7).

5.2 Post Questionnaire

The six-point rating scores of the three conditions were analyzed using a Friedman test. As indicated in Figure 7, the differences among the three conditions were significant regarding efficiency ($p = 0.032$) and precision ($p = 0.040$), and just short of significance for ease-of-learning ($p = 0.054$) and fatigue ($p = 0.057$). Post-hoc analyses were performed using pairwise Wilcoxon signed-rank tests. The results suggested that the subjects considered VIEW mode to be more efficient, and less tiring to use than DRIVE mode ($p = 0.075$ and 0.009 , respectively) and the WAND interface ($p = 0.004$ and 0.007 , respectively). Additionally, VIEW mode was also considered to be more precise than the WAND interface ($p = 0.004$), and easier to learn than DRIVE mode ($p = 0.025$). Pearson correlation analyses between the rating scores and the subjects' prior experiences and MRT scores were also performed. However, no strong correlation coefficient was identified.

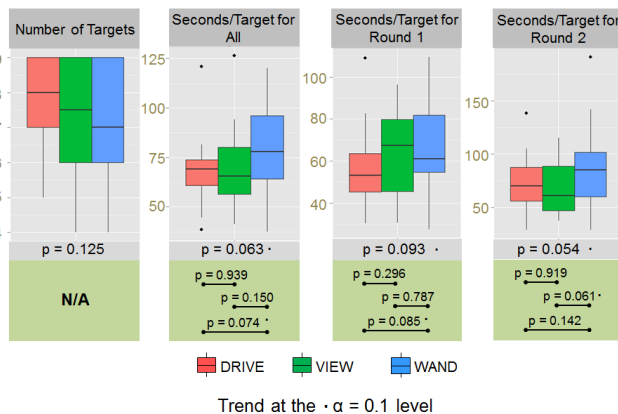


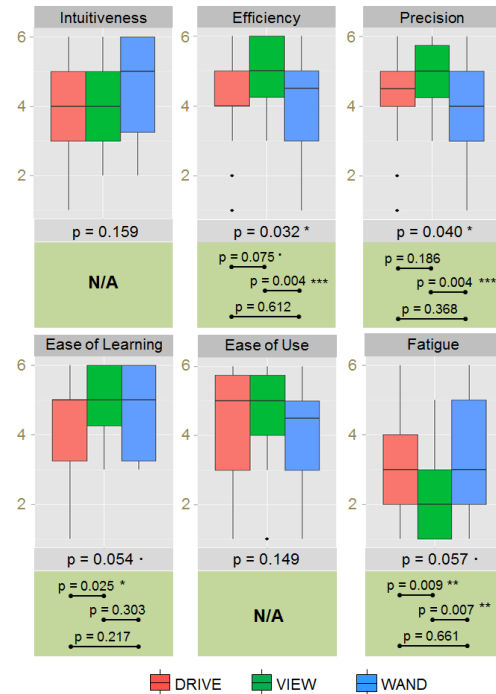
Figure 6: The analysis of the task performance indicators.

5.3 User Feedback

5.3.1 Tetrahedron Docking Task

After finishing all tetrahedron experiments, the subjects were asked to compare the three interfaces, and comment on what they liked and disliked about each of them. By summarizing the comments, we found that each interface had its positives and negatives for this task. For the WAND interface, 10 subjects

complimented its affordance of simple, natural, and realistic rotation control, praising that a 3D interface was used for a 3D task, and that the wand offered more tactile control than the tablet. Eight subjects also liked it because it was easy to learn and use, as it did not require switching between two displays, so that their immersion in the VE could remain unbroken. On the downside, 14 subjects felt the wand device was physically difficult and tiring to rotate. Two subjects suggested replacing the wand device with a ball-shaped prop [12]. In addition, nine subjects disliked the WAND interface because controlling 3-DOF rotation was more challenging than matching the crosshairs. Six of them mentioned that to precisely match the targets, they had to frequently switch the viewing angles by flying to different spots around the target object.



Trend at the $\cdot \alpha = 0.1$ level.
Significant at $\ast \alpha = 0.05$, $\ast \ast \alpha = 0.01$, $\ast \ast \ast \alpha = 0.005$

Figure 7: The analysis of the subjective rating scores.

For VIEW mode, 14 of 26 subjects complimented it for making the target matching process easier and faster. Specifically, six subjects found the combination of the avatar's view and the object's view helpful, as third-person control from the avatar's view allowed them to translate the object efficiently, while first-person control from the object's view allowed them to match the rotation intuitively and precisely. Four subjects preferred this mode because matching the 2D crosshairs was easier than figuring out the mental rotations to match the targets in 3D space. Five subjects liked to use the tablet device, because it was more stable and precise to touch on a 2D plane than holding and manipulating a wand. On the other hand, seven subjects disliked having another display, as it made the task more complicated, and took away the immersion and spatial orientation established in the HMD view. In addition, nine subjects pointed out that searching for the reference plates (i.e., the ones that accompanied each target tetrahedron) could sometimes become very difficult to do on the tablet, partly due to the first-person view [9]. Based on the experimenter's observations, some subjects attempted to alleviate this challenge during the experiments by looking at the HMD while touching the tablet. Nevertheless, many of them struggled,

because the the mapping of the swiping gesture was based on the object's view, and felt inverted from the avatar's view. Noticing this problem, one subject asked the experimenter if it was possible to detect his gaze change to the HMD, and base the touch control on its perspective instead – a solution similar to the interface sharing idea proposed in our previous work [31].

5.3.2 Spotlight and House Tasks

The anecdotal feedback session presented the user with a hybrid interface that combined all three aforementioned interface conditions. The user could point at an object and hold down different modal buttons to translate and rotate it using the wand device. The first-person view of the object was displayed on the tablet upon selection, and could then be rotated using multi-touch gestures. Furthermore, pressing the “+” and “-” buttons on the wand device made the user *jump in and out* of the selected object respectively, realizing DRIVE mode through a transitional user interface approach [2]. Due to user fatigue and other logistical reasons, only 15 of the 26 subjects participated in this session. Nonetheless, they all tried different modes for both the spotlight task and the house task, and provided oral feedback to the experimenter on the go.

By summarizing their comments, the authors found that a majority group of 12 subjects did not state a clear interface preference for either task. Instead, they liked the fact that they could switch between interfaces, and felt that having all three options was actually better than any of them alone. For example, seven subjects preferred using the wand device for positioning the house, but would rather use the object's first-person view to align the orientation. For object-impersonation-based rotation control, eight subjects considered the jump-in DRIVE mode to be more effective than VIEW mode, as target house searching was easier by looking around, and the house orientation could quickly follow the view by pressing the “B” button. Specific to the house task, five subjects pointed out that none of the two modes had made the object's first-person view appropriate for judging the leveling of the house; without any visual cue added to the VE, the user still had to refer to the avatar's third-person view to place the house on the ground, using either the WAND interface or the tablet in DRIVE mode. The three interfaces also had different and complementary advantages in the spotlight task. Nine subjects preferred using the wand device for translating the spotlight, and two subjects were willing to use it for orienting it as well, since it only involved 2-DOF rotation, in comparison to the 3-DOF tetrahedron rotation task. In addition, a majority group of 12 subjects preferred to control orientation from the spotlight's first-person view, as it was more direct, intuitive, and efficient. Four subjects even felt it *too easy to do* using the jump-in DRIVE mode, as they could simply look at the target, and press a button to accurately illuminate it.

5.4 Discussion

Using the object-target alignment tasks as the test bed, the user study results revealed various advantages and limitations of object impersonation in HVE systems. Although the results were not conclusive, the performance results, such as the average time spent on each round of targets, did show statistical trends that object impersonation could complement a traditional 3D wand interface to make performance of 6-DOF manipulation tasks more efficient (**H1**). Analyses of subjective measurements revealed advantages and shortcomings of each interface condition. According to the subjects' ratings, VIEW mode provided the most efficient, precise, and least-fatiguing interface of the three conditions (**H3**). The subjects' post-study comments suggested two explanations for these preferences. First, by requiring the user to align the crosshairs to match the rotation, the object

impersonation techniques transformed complex and hard-to-reason 3D rotation tasks [32] into simpler and more-intuitive 2D target-matching tasks. Second, the tablet device offered a physical surface to touch on, leading to an increase in operator effectiveness and precision, and a reduction in user fatigue confirming results of other studies [1] [14] [15], especially in comparison to spatial input devices (i.e., the WAND interface).

Partially refuting **H3**, neither the VIEW nor DRIVE mode was considered to be easier to use than the traditional WAND interface. A summary of user comments suggests they had difficulty searching for the reference plates from the object's first-person view, especially using the tablet in VIEW mode. By allowing the user to search with head and chair turning, this challenge was alleviated in DRIVE mode. However, DRIVE mode forced the user to completely immerse themselves in the object's body, without providing a maneuverable avatar camera on the tablet to adjust the position of the tetrahedron from all angles. On the other hand, refuting **H2**, the WAND interface was not rated to be more intuitive or easier to use than the two object impersonation modes, although DRIVE mode was commented as being more difficult to understand and learn. This suggests that object impersonation may be better accepted as an augmentation to, instead of a complete replacement for, existing interaction metaphors.

Our hypotheses in **H4** and **H5** were only evaluated anecdotally. Six subjects complimented VIEW and DRIVE modes for requiring less mental rotation, as the DOFs involved in the rotation alignment process were reduced from three to two (**H4**). Increased cognitive overhead of attending to two ICs was mentioned by seven subjects for VIEW mode, and six subjects for DRIVE mode (**H5**). According to them, dividing the task sequences to different ICs made it more complex to complete, and also broke the immersion established in the HMD. This issue was mainly caused by divided attention during context switching, and could be alleviated by peripheral displays [9], display blend-in, and interaction coordination mechanisms [31].

The anecdotal feedback collected during the spotlight and house task sessions suggest a need to further combine the three interface conditions to form a more-advanced hybrid interface on top of the current HVE system. In other words, *such a system should not only combine the immersive and tablet ICs (the avatar and object perspectives), but also the different interface approaches, to counter each of their disadvantages*. In addition, this session also provided interesting insights about the applicability of object impersonation in real world application tasks. The preference of object impersonation was most evident in the spotlight task. On one hand, the cone shape of the spotlight was similar to the frustum of the first-person view, offering good visual affordance for the object impersonation metaphor. On the other hand, the goal of the task (i.e., having the light illuminating the target) also had a strong similarity to the user's action of looking at a target. In contrast, the effectiveness of object impersonation fell short for leveling the houses on the ground, due to the lack of visual cues from inside the house itself, and users needed to refer to the traditional exocentric interaction paradigm for better efficiency. These findings again suggest that object impersonation should be used in a hybrid context to complement the limitations of traditional VR interfaces, instead of replacing them.

6 CONCLUSION

To conclude, this paper proposed a new interaction technique that can benefit various 3DUI task scenarios in immersive VR. By impersonating a virtual object, the user can perform 3D interaction from a different perspective, or even manipulate the impersonated object by looking and traveling around the VE. This

blurs the line between basic 3DUI tasks, and can be used in HVE systems to complement the limitations of traditional 3D interfaces. As listed in Section 3.1, object impersonation can be used to enhance 3D interaction in many task cases. As a start, the user study presented in this paper used three types of object-target alignment tasks as the test bed to investigate the task performance and user experience with two different object impersonation implementations, within a tablet-and-HMD-based HVE system. The results showed improved task performance and user experience using object impersonation together with traditional 3D UIs, but also suggested issues and limitations that make it less useful by itself.

The cross-task interaction paradigm presents new opportunities in 3D UI research. As our first attempt, the system and study presented in this paper clearly have their limitations and drawbacks. The impersonation studied here is still limited to view point and reference frame changes, and does not allow the user to use his/her full body motion to *act* as the impersonated virtual object. The divided attention between the tablet and HMD induces cognitive overhead in context transitions. The study results show promising performance and user experience improvements, but due to the compound effect of touch input and reduced task DOF, it is difficult to isolate and precisely appraise the real benefits of object impersonation. Lastly, to advocate cross-task interaction as a mainstream 3D UI design, many more convincing use cases, like the spotlight alignment task, need to be discovered and tested. Therefore, the authors plan to address these questions carefully in future work, by developing and testing different IC coordination mechanisms to reduce the mental overhead of context transitions [31], as well as by designing and studying various application scenarios.

REFERENCES

- [1] I. Angus and H. Sowizral, "Embedding the 2D interaction metaphor in a real 3D virtual environment," *Proc. IS&T/SPIE's Symposium on Electronic Imaging: Science & Technology*, pp. 282-293, 1995.
- [2] M. Billinghurst, H. Kato, and I. Poupyrev, "The MagicBook: a transitional AR interface," *Computers and Graphics*, vol. 25, no. 5, pp. 745-753, 2001.
- [3] A. Bornik, R. Beichel, E. Kruijff, and D. Schmalstieg, "A hybrid user interface for manipulation of volumetric medical data," *Proc. IEEE 3DUI*, pp. 29-36, 2006.
- [4] D. Bowman, "Interaction techniques for common tasks in immersive virtual environments," PhD Dissertation, Georgia Institute of Technology, 1999.
- [5] D. Bowman and L. Hodges, "An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments," *Proc. ACM i3D*, pp. 35-ff, 1997.
- [6] D. Bowman, E. Kruijff, J. J. LaViola, and I. Poupyrev, *3D User Interfaces: Theory and Practice*. Addison-Wesley Professional, 2004.
- [7] L. Brown and H. Hua, "Magic lenses for augmented virtual environments," *IEEE Computer Graphics and Applications*, vol. 26, no. 4, pp. 64-73, 2006.
- [8] F. Carvalho, D. Trevisan, and A. Raposo, "Toward the design of transitional interfaces: an exploratory study on a semi-immersive hybrid user interface," *Virtual Reality*, vol. 16, no. 4, pp. 271-288, 2011.
- [9] J. Chen, M. Narayan, and M. Perez-Quinones, "The use of hand-held devices for search tasks in virtual environments," *Proc. IEEE VR (Workshop on New Directions in 3DUI)*, pp. 15-18, 2005.
- [10] T. Elvins, D. Nadeau, R. Schul, and D. Kirsh, "Worldlets: 3D thumbnails for wayfinding in large virtual worlds," *Presence: Teleoperators and Virtual Environments*, vol. 10, no. 6, pp. 565-582, 2001.
- [11] K. Hinckley, R. Pausch, D. Proffitt, J. Patten, and N. Kassell, "Cooperative bimanual action," *Proc. ACM CHI*, pp. 27-34, 1997.
- [12] K. Hinckley, J. Tullio, R. Pausch, D. Proffitt, and N. Kassell, "Usability analysis of 3D rotation techniques," *Proc. ACM UIST*, pp. 1-10, 1997.
- [13] T. Igarashi, R. Kadobayashi, K. Mase, and H. Tanaka, "Path drawing for 3D walkthrough," *Proc. of ACM UIST*, pp. 173-174, 1998.
- [14] R. Lindeman, J. Sibert, and J. Hahn, "Towards usable VR: an empirical study of user interfaces for immersive virtual environments," *Proc. ACM CHI*, pp. 64-71, 1999.
- [15] A. Marzo, B. Bossavit, and M. Hachet, "Combining multi-touch input and device movement for 3D manipulations in mobile augmented reality environments," *Proc. ACM SUI*, pp. 13-16, 2014.
- [16] M. Miguel, T. Ogawa, K. Kiyokawa, and H. Takemura, "A PDA-based see-through interface within an immersive environment," *Proc. IEEE Artificial Reality and Telexistence*, pp. 113-118, 2007.
- [17] M. Peters, B. Laeng, K. Latham, M. Jackson, R. Zaiyouna, and C. Richardson, "A redrawn Vandenberg and Kuse mental rotations test-different versions and factors that affect performance," *Brain and Cognition*, vol. 28, no. 1, pp. 39-58, 1995.
- [18] J. Pierce, B. Steams, and R. Pausch, "Voodoo Dolls: seamless interaction at multiple scales in virtual environments," *Proc. ACM i3D*, pp. 141-145, 1999.
- [19] M. Pinho, D. Bowman, and C. Freitas, "Cooperative object manipulation in collaborative virtual environments," *Journal of the Brazilian Computer Society*, vol. 14, no. 2, pp. 54-67, 2008.
- [20] I. Poupyrev, N. Tomokazu, and S. Weghorst, "Virtual Notepad: handwriting in immersive VR," *Proc. IEEE VR*, pp. 126-132, 1998.
- [21] I. Poupyrev, S. Weghorst, M. Billinghurst, and T. Ichikawa, "Egocentric object manipulation in virtual environments: empirical evaluation of interaction techniques," *Computer Graphics Forum*, vol. 17, no. 3, pp. 41-52, 1998.
- [22] S. Razzaque, "Redirected walking," PhD Dissertation, University of North Carolina at Chapel Hill, 2005.
- [23] D. Schmalstieg, M. Encarnacao, and Z. Szalavari, "Using transparent props for interaction with the virtual table," *Proc. ACM i3D*, pp. 147-154, 1999.
- [24] D. Schmalstieg and G. Schaufler, "Sewing worlds together with SEAMS: a mechanism to construct complex virtual environments," *Presence: Teleoperators and Virtual Environments*, vol. 8, no. 4, pp. 449-461, 1999.
- [25] R. Stoakley, M. Conway, and R. Pausch, "Virtual reality on a WIM: interactive worlds in miniature," *Proc. ACM CHI*, pp. 265-272, 1995.
- [26] Z. Szalavari and M. Gervautz, "The personal interaction panel - a two-handed interface for augmented reality," *Computer Graphics Forum*, vol. 16, no. 3, pp. C335-C346, 1997.
- [27] J. Viega, M. Conway, G. Williams, and R. Pausch, "3D magic lenses," *Proc. ACM UIST*, pp. 51-58, 1996.
- [28] M. Wang Baldonado, A. Woodruff, and A. Kuchinsky, "Guidelines for using multiple views in information visualization," *Proc. ACM AVI*, pp. 110-119, 2000.
- [29] J. Wang and R. Lindeman, "Comparing isometric and elastic surfboard interfaces for leaning-based travel in 3D virtual environments," *Proc. IEEE 3DUI*, pp. 31-38, 2012.
- [30] J. Wang and R. Lindeman, "ForceExtension: extending isotonic position-controlled multi-touch gestures with rate-controlled force sensing for 3D manipulation," *Proc. IEEE 3DUI*, pp. 3-6, 2013.
- [31] J. Wang and R. Lindeman, "Coordinated 3D interaction in tablet-and HMD-based hybrid virtual environments," *Proc. ACM SUI*, pp. 70-79, 2014.
- [32] S. Zhai and P. Milgram, "Human performance evaluation of manipulation schemes in virtual environments," *Proc. of IEEE VR*, pp. 155-161, 1993.