To appear in Proceedings of IEEE Virtual Reality (VR '05). Bonn, Germany. March 10–12, 2005.

Cross-Dimensional Gestural Interaction Techniques for Hybrid Immersive Environments

Hrvoje Benko Edward W. Ishak Steven Feiner

Department of Computer Science, Columbia University
500 W. 120th St. 450 CS Building, New York, NY
{benko | ishak | feiner}@cs.columbia.edu

ABSTRACT
We present a set of cross-dimensional interaction techniques for a hybrid user interface that integrates existing 2D and 3D visualization and interaction devices. Our approach is built around one- and two-handed gestures that support the seamless transition of data between co-located 2D and 3D contexts. Our testbed environment combines a 2D multi-user, multi-touch, projection surface with 3D head-tracked, see-through, head-worn displays and 3D tracked gloves to form a multi-display augmented reality. We address some of the ways in which we can interact with private data in a collaborative, heterogeneous workspace. We also report on a pilot usability study to evaluate the effectiveness and ease of use of the cross-dimensional interactions.

CR Categories: H.5.1 **[Information Interfaces and Presentation]**: Multimedia Information Systems—Artificial, augmented, and virtual realities; H.5.2 **[Information Interfaces and Presentation]**: User Interfaces—Graphical user interfaces (GUI), Interaction styles

Keywords: hybrid user interfaces, augmented reality, tabletop interaction, gesture-based and touch-based interaction.

1. INTRODUCTION

Augmented Reality (AR) makes it possible to visualize computer-generated graphics overlaid onto the surrounding environment. Many AR systems focus on the use of personal displays (such as see-through head-worn displays), and their user interfaces often consist of purely virtual elements, such as 3D widgets or 3D interaction metaphors. However, since the user is not completely immersed in the virtual world, AR also allows for the simultaneous use of all existing user interfaces that surround the user. These surrounding user interfaces are often better suited for some tasks than the personal AR user interface, but the lack of simple ways to move data to and from them discourages the AR user from interacting with them.

In this paper, we present an experimental system that combines multiple see-though head-worn 3D displays with a 2D user interface projected on a shared touch-sensitive surface. We refer to a heterogeneous combination of displays and interaction technologies as a *hybrid* user interface [6], in this case forming a hybrid immersive environment. Our research focuses primarily on the border between 2D and 3D user interfaces.

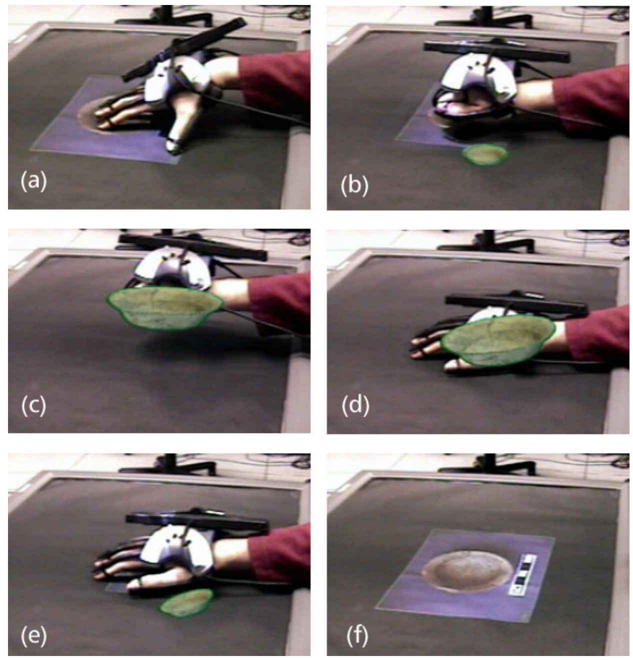

Figure 1: Frames from a sequence of cross-dimensional pull and push gestures. (a) Selecting the 2D object and beginning to form a grabbing gesture. (b) The 2D object disappears, while the 3D object appears from the table. (c) Holding the 3D object. (d) Push gesture begins by pressing on the table through the 3D object. (e) The 3D object disappears and the 2D object appears. (f) The 2D object, projected on the table. (Table surface is covered with black paper to provide a darker background for imaging the projected display through the live tracked video see-through display used to create all images in this paper.)

Given the large body of existing work on gestural interaction in 2D (e.g., [27]) and 3D (e.g., [15]), what are missing are hybrid gestural interactions that create a unified interaction space. To bridge this divide, we have designed *cross-dimensional* gestural interactions that require the use of both 2D and 3D modalities, as shown in Figure 1. These context-aware cross-dimensional gestures build on and integrate the ways in which we already interact gesturally in 2D and 3D environments.

There are three pieces of information required to complete a precise and successful transition across the dimensionality border: the *source object*, the *type of transformation*, and the *destination location*. For example, to transform an object from a 2D representation into a 3D representation (similar to Figure 1a–c), one needs to specify the 2D object that is undergoing a transformation, that the desired effect is a transformation between 2D and 3D, and finally, where the resulting 3D object will be located once it is transformed. We believe that our cross-dimensional gestures provide a way to satisfy all three requirements simultaneously and unambiguously, and thus provide a seamless transition across the 2D–3D divide.

In addition to supporting transitions between 2D and 3D environments, we extend our cross-dimensional gestural techniques with gestures that allow for the use of 2D interactions on 3D objects. We also explore the issue of data privacy in a collaborative hybrid environment, and provide additional gestures that facilitate private dimensionality transitions.

2. Related Work

The inspiration for our work has come from two different themes: hybrid user interfaces and gesture- and touch-based interactions.

2.1. Hybrid User Interfaces

Our first research on hybrid user interfaces embedded a small, stationary, flat-panel "detail" display within the field of view of a monocular, head-tracked, see-through, head-worn "context" display [6]. Low and colleagues reversed this approach, using a stereo, head-tracked, opaque, head-worn display to provide detail within the context of a larger monocular, projected display [14]. Neither of these systems addressed gestural interaction, although the former supported seamless mouse-based interaction across the interface between the displays.

EMMIE [3] demonstrated a hybrid user interface that integrated a variety of tracked handheld and stationary displays, head-worn displays, and physical input devices. See-through head-worn displays were used to overlay virtual information on the physical environment, including the displays within it. EMMIE's primary cross-dimensional user interaction technique accomplished *drag-and-drop* across displays using a 3D mouse. Privacy was supported by manipulating 3D "privacy lamps" that caused the virtual objects they illuminated to become invisible to others.

Regenbrecht and colleagues created MagicMeeting [19], a collaborative tangible AR environment that mixes 2D and 3D displays and employs tangible tools for seamless interaction between 2D and 3D data. In contrast to our work, all of their interaction techniques rely on the use of tangible tracked tools.

Both EMMIE and MagicMeeting have used a "universal tool" (a 3D mouse) to interact with components in a hybrid environment, much like using a common remote control for multiple components in a home entertainment system. An obvious benefit to this approach is that the user needs to master only one tool to control the environment. However, using that tool becomes increasingly complex as the number of possible interactions in the environment grows. In addition, simultaneously specifying all three of the requirements outlined in Section 1 for crossing the dimensionality border is difficult with a universal tool approach, and it often results in a two- or three-step process. In contrast, our cross-dimensional gestures allow users to maintain the familiar interaction techniques that are native to each of the user interfaces being combined, as well as make sure that data can be shared seamlessly between all interfaces.

The StudierStube collaborative AR environment [23] provides the Personal Interaction Panel (PIP), a two-handed interaction tool for manipulating 2D and 3D data. The PIP offers passive tactile feedback and allows for 2D and 3D interactions; however, it does not use any gestural interaction and the visualization remains completely within the 3D context. Rekimoto's *pick-and-drop* [20] and Rekimoto and Saitoh's *hyperdragging* [22] techniques were developed as part of an exploration of multi-computer direct-manipulation interfaces that examined the hybrid interactions between multiple portable computers, table and wall displays, and other physical objects. However, their interaction set addresses only pointer-based interaction with 2D data.

Brown and colleagues, in their SCAPE collaborative AR environment [2], use head-worn projective displays that project onto many retroreflective surfaces in a room, including the walls, a tabletop, and several hand-held widgets. Since all material is projected and visible only along a projector's line of sight, all data is potentially private. However, data cannot be overlaid on conventional displays in the room because of the need for a retroreflective surface to display it. Although Brown and colleagues employ hand gestures for navigation, all interactions with objects in the world are accomplished through their hand-held widgets.

2.2. Gesture- and Touch-Based Interactions

Much research has been performed on gesture-based and touch-based interaction in both 2D desktop systems and 3D virtual reality. Krueger's VIDEOPLACE [12] is an early example of multi-hand and multi-finger interaction that recognizes gestures through 2D video silhouettes. Recent work by Wu and Balakrishnan [27] presented multi-finger and whole-hand gestural interaction techniques for multi-touch, multi-user displays. While their system was entirely 2D and used a single display, it highlighted the touch- and gesture-sensing capabilities of the DiamondTouch table [5]. Rekimoto's SmartSkin technology [21] also addresses tabletop interaction and offers more powerful multi-touch capabilities than DiamondTouch; however, it cannot differentiate input from different users.

The i-LAND [24] and Connectables [25] projects both explore collaboration and interaction with displays built into furniture. Hinckley [8] uses synchronous gestures for collaborating and interacting with several independent mobile displays.

Natural 3D gestural interactions in virtual reality have been explored in numerous research projects (e.g., [16, 26]). Our 3D gestures share some similarities with the work of Cutler and colleagues [4], who developed a set of both single- and two-handed gestural interaction techniques for manipulating 3D objects on the Responsive Workbench. In work by Zeleznik and colleagues [28], as well as by Pierce and colleagues [18], 2D gestures are used to select and manipulate 3D objects. While all of these researchers focus on selection and manipulation of 3D objects, our gestures also deal with the transformation of objects between 2D and 3D environments.

In attempting to cross the 2D–3D divide, we have been inspired by the gestures that Ishii and colleagues have developed for manipulating tangible user interfaces [10], as our work similarly addresses "seamless workspaces" that stretch across different dimensionalities, displays, and users [9].

3. Implementation Overview

The cross-dimensional gestures presented in this paper have been developed as a part of *VITA (Visual Interaction Tool for Archaeology)* [1], an experimental collaborative mixed reality system for offsite visualization of an archaeological excavation site. As shown in Figure 2, VITA's users can explore conventional ar-

Figure 2: Two users, wearing tracked, see-through, head-worn displays and tracked 3D gloves, interact with 2D and 3D objects in VITA.

chaeological data, which is mostly 2D in nature (drawings, pictures, and notes), complemented with a variety of 3D acquired data (panoramic images, and 3D models of small finds and of the site itself at various excavation stages), in one seamless collaborative mixed reality environment.

The shared hybrid collaborative space that VITA supports is composed of the multi-user, multi-touch, projection surface, high resolution displays, and the entire surrounding 3D space tracked by a 6DOF tracking system, viewed through head-tracked, see-through, head-worn displays. Each user can use a tracked glove, speech commands, and the multi-touch sensitive surface to interact multimodally with the system and collaborate by jointly navigating, searching, and viewing data. Although, several archaeologists have used and evaluated the complete VITA system, for the purpose of user evaluation of cross-dimensional gestures (as described in Section 5), we isolated the functionality of the cross-dimensional gestures into a subset system described below.

3.1. Hardware Setup

Our hybrid environment consists of two types of modules: a 2D module and a 3D module. Each module powers one display unit; therefore, with two users sharing our interaction table, we require one 2D module and two 3D modules (one for each AR user). The 2D visualization module uses a Mitsubishi Electric Research Laboratory DiamondTouch table [5], a multi-user multi-touch surface (26"×20"), onto which we project imagery with a Proxima Ultralight x350 DLP projector (1024×768 resolution). We covered the table surface with black paper to increase the apparent contrast and improve the quality of images shot through the video see-through display used to create the figures in this paper.

Our 3D augmented reality modules consists of a modified EssentialReality P5 glove to sense right-hand gestures, a Sony LDI-D100B optical see-through head-worn display (800×600 resolution) to present the AR overlay, and two InterSense IS900 six–degree-of-freedom trackers to monitor head and hand position and orientation.

One PC workstation (dual Athlon MP 2.0, 1GB RAM, NVIDIA Quadro4 750 XGL) powers each module and all communication is implemented via message passing through a publish-and-subscribe message board system. We use the Adaptive Agent Architecture (AAA) [13] to facilitate easy connection, discovery, and communication management of our modules.

4. Cross-Dimensional Interaction Techniques

The cross-dimensional interaction techniques that we have developed can be categorized into two main classes: one-handed techniques and two-handed techniques. One-handed techniques are generally simpler and more intuitive, thus forming a set of basic interactions for a hybrid environment, while two-handed techniques comprise a more complex set that addresses higher-level concepts, such as privacy. Furthermore, having a relatively rich set of techniques that each require only a single hand opens up many interesting opportunities for using two one-handed techniques simultaneously.

In the remainder of this paper, we use "2D object" to refer to a pictorial representation of a real object presented on the table surface and "3D object" to refer to a 3D model of a real object presented to the user on the head-worn display.

4.1. Building Blocks: Existing 2D and 3D Gestures

We begin by describing the gestures that our system recognizes independently in 2D and 3D, as shown in Figure 3. Both 2D and 3D recognizers are heuristic-based and implemented as finite state automata. In our 2D environment, we support *one-finger touch*, *tap* (typically with one or more knuckles), *flat hand*, and *vertical hand*. The DiamondTouch information that we use for discriminating 2D gestures is the width and height of the axis-aligned bounding box of a user's contact points. In addition, we extend these gestures by considering motion, such as a *swoop* gesture (the transition between the idle flat hand and a closed fist). In 3D, we support conventional *point*, *grab*, and *idle* (all fingers unbent) gestures. We have built in part on our previous work in multimodal interaction for augmented and virtual immersive environments [11, 17], in which our 3D gesture-recognition technology was originally developed.

Our primary considerations for selecting these basic gestures were simplicity and relative lack of ambiguity, rather than completeness. While there are many other useful gestures possible in 2D and 3D, we feel that this set forms a good basis for creating interactions that span the 2D–3D divide in a wide range of potential applications.

Hand Gesture	2D Interpretation	3D Interpretation
	One finger touch	Point
	Flat hand	Idle
	Tap	Grab
	Vertical hand	Idle

Figure 3: A set of basic hand gestures and their interpretations in 2D and 3D environments.

While these gestures are recognized independently, both in 2D and in 3D, in a hybrid environment it is important to look at them in combination, rather than in isolation. This is especially true for hand gestures that have different meanings, depending on whether they are interpreted in 2D or in 3D. In our system, the 2D and 3D gestures are considered jointly and integrated together using an additional finite state automaton, thus facilitating more complex interactions. In the following sections, we define a set of hybrid interactions (each of which is prefixed by the modifier "Cross-Dimensional") by combining interpretations across the 2D–3D divide.

Selection is probably the most basic interaction in any environment. Because we believe that selection is inherently environment-dependent, we have chosen to implement it differently in 2D and 3D. 2D selection is implemented by a tap gesture on top of a 2D object, while 3D selection is implemented by a grab gesture near a 3D object. These gestures are extended by factoring in hand motion, yielding *touch and drag* and *grab and move*, both of which produce movement of the selected object. While these techniques are not "hybrid," selection is a critical component of the interaction set from which we form our hybrid gestures.

4.2. Single-handed Interactions

4.2.1. Cross-Dimensional Pull

To make an object transition from a 2D representation to a 3D representation, the user employs a *cross-dimensional pull* gesture. Wanting this gesture to mimic a real-world "grab" from a flat surface, we asked several users to perform exaggerated grabs of actual physical objects and observed each user's hand's location, movement, and finger contractions and expansions as they grabbed items, such as a CD, tennis ball, and a pair of scissors, from a table. Based on these observations, we designed the gesture so that the user places the hand flat on the table on top of the object's image representation, then moves the fingers towards the center of the object, and finally forms a fist (Figure 1a–c). As this occurs, the 2D image on the table scales down and disappears, while a 3D model appears and scales up, attached to the user's grabbing hand, representing the transformation from 2D to 3D. This gesture exemplifies a theme common to the entire set: preservation of the interaction context. If the intention is to transfer something from 2D to 3D, the user starts the gesture on the 2D surface, and finishes it in the 3D world.

4.2.2. Cross-Dimensional Push

As a complement to the cross-dimensional pull, an object can transition from 3D to 2D through use of the *cross-dimensional push* gesture. Once the 3D object is placed above the table, a user can position a flat hand above the object and "push" downwards until her hand touches the table (Figure 1d–e), mimicking an attempt to push a real physical object into the table. As this occurs, the 3D model being pushed scales down and disappears, while the 2D image appears on the table and scales up. (Alternatives to our uniform-scaling visualization include one-dimensional scaling along the axis perpendicular to the table to "flatten" a 3D object or extrude a 2D one, or clipping the object against a virtual plane coplanar with the table. Visual differences among many of the 2D–3D pairs have led us to select our current approach.)

4.2.3. Cross-Dimensional Connect/Disconnect

There are instances when a user needs to see both the 2D and 3D object representations at the same time. To establish this relationship between the two environments, the user performs a *cross-dimensional connect* gesture by first selecting (grabbing) the 3D object, pulling it towards the table, and then briefly tapping the table anywhere while continuing to hold the object (Figure 4a). This results in a 2D object appearing on the table surface while the 3D object remains visible and attached to the hand. Once the two representations are connected, a leader line appears between them, highlighting the connection (Figure 4b). This leader line visualization is similar to EMMIE's use of leader lines that extend from a search control panel on a hand-held display to 3D virtual objects that meet the search criteria [3]. We designed this gesture to resemble pulling a virtual string between one object and the other. This allows 3D objects placed anywhere in 3D space (even behind the user) to be located by the user by first locating its 2D representation and then following its leader line to its 3D location.

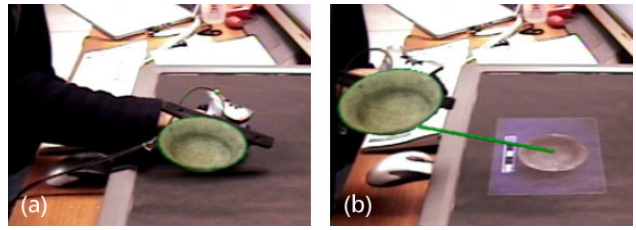

Figure 4: Cross-dimensional connect gesture. (a) Tapping the table while holding a 3D object establishes connection with the table. (b) Leader line connects two representations of the same object.

We also experimented with the idea of having a connected 2D object appear as a "virtual shadow" on the table following the 3D object in the world, similar to the ideas presented by Herndon and colleagues [7]. If the user moves the 3D object away from the table, the "shadow" maintains the position on the table that is closest to the 3D object's position in the world. While the "shadow" representation felt like a natural way to showcase the connection between the two objects, the leader lines alone allow for more freedom, since the position of each object is independent of the other.

The same tapping gesture used to connect the two representations can also be used to break the connection between them to perform a *cross-dimensional disconnect.* All that is needed is to tap the 2D representation on the table in the same manner used to perform a connect gesture. The 2D object disappears, leaving the 3D object in the user's hand. This cross-dimensional interaction thus serves as a simple, but powerful, toggle action.

4.2.4. Cross-Dimensional Pin/Unpin

Another way that the user can connect the 3D and the 2D environments is by "pinning" a 3D object to the table. In contrast to the cross-dimensional push, which transforms a 3D object into a 2D object, a *cross-dimensional pin* attaches the 3D object to the 2D surface. This is useful when wanting to perform a direct manipulation or transformation of the 3D object using a tactile 2D

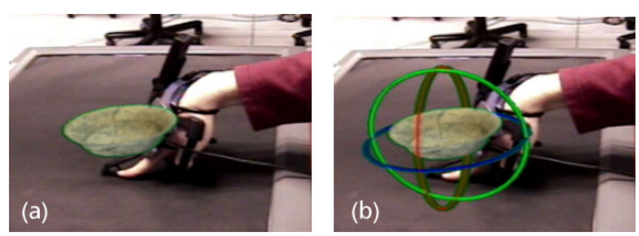

Figure 5: (a) Cross-dimensional pin gesture associates the object with the table. (b) Rings appear once the object is pinned to provide visual feedback to the user.

table surface. Pinning an object is accomplished by placing the 3D object close to the table, making a pointing gesture with the hand, and then pushing through the object until touching the table (Figure 5a). Three colored rings appear around the object to signal that the object is pinned to the table (Figure 5b). To cross-dimensionally unpin the 3D object, the user needs to select it (i.e., grab it) and move it away from the table surface. As soon as the object is grabbed in 3D, it is no longer pinned, which can be seen by the disappearance of the three colored rings around it.

Although the differences between the pin and the connect techniques might appear subtle, we feel that they are distinctly and usefully different. The connect technique allows for simultaneous viewing and manipulation of both 2D and 3D representations. By pinning the object to the table, there is no 2D representation shown below the 3D object and the focus is on allowing the use of the rich set of existing 2D touch-based interaction techniques for directly manipulating the 3D model, as described below. For example, when a 3D object is pinned to the table, the user can drag it around in exactly the same way that the 2D object can be moved. Simple tap and drag will translate the pinned object anywhere on the table.

4.2.5. Context-Sensitive Menus and Parameter Adjustment

We extend the gestures developed by Wu and Balakrishnan [27] for context-sensitive menus to manipulate 3D objects. This makes it possible to invoke a context-sensitive menu for a pinned 3D object in exactly the same way as for the regular 2D object: by tapping and holding on the table at the location of the object. A context-sensitive menu will then appear to the left of the hand. (The placement of the menu is currently determined by the fact that all our tracked gloves are right handed, and therefore minimal occlusion occurs when the menu is shown on the left side. Alternatively, the system could position the menu based on which hand is currently touching the table.)

Oftentimes, a user may wish to adjust the parameter values of a pinned or connected 3D object; for example, to perform transformations on the object, such as a rotation, scale, or hue-change. However, these can be cumbersome to accomplish in 3D space using a tracked, instrumented glove without haptic feedback. Instead, such parameter adjustment widgets can be selected from the context-sensitive menu, and by taking advantage of the multi-touch table, a user can apply 2D touch-based interactions with passive haptic feedback to modify the 3D pinned object. For example, Figure 6 shows a user rotating a pinned vessel object.

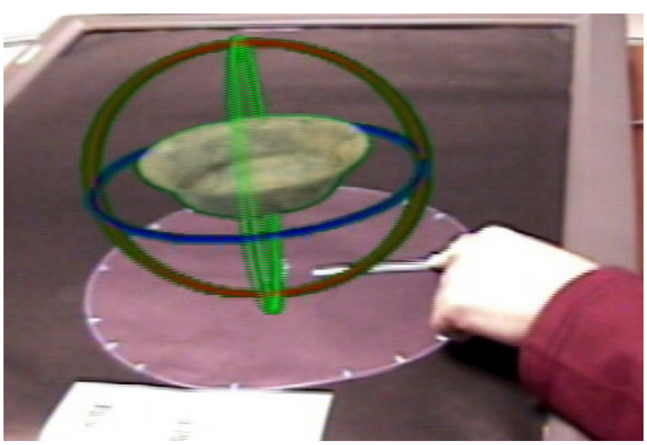

Figure 6: The context-sensitive menu is invoked on the 3D object and the 2D rotation tool is applied.

4.3. Two-Handed Interactions

4.3.1. Interaction Combo

To create more powerful gestural interactions, we allow the user to use both hands simultaneously to adjust two different independent transformation parameters of an object. For example, a user may wish to perform transformations on the object, such as changing its scale or color, while holding the object in her hand and inspecting it from all sides. As before, to avoid having to adjust these parameters in 3D, we use the multi-touch table to allow users to perform these adjustments on a 2D surface with their non-tracked hand, while their tracked hand can simultaneously manipulate the 3D object representation directly. For example, with remote parameter adjustment, the user can grab and move a 3D object in 3D space to change its location, while simultaneously using the other hand on the multi-touch surface to adjust its scale, as shown in Figure 7.

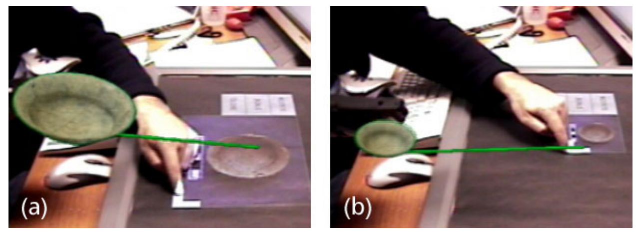

Figure 7: Interaction combo: Combining scaling with 3D manipulation. (a) The 3D object is held by the right hand, while the left hand adjusts the scale. (b) The right hand rotates the object to get a better view, while the object is simultaneously being scaled.

4.3.2. Cross-Dimensional Privacy Pull

In a hybrid collaborative environment, the users share a common workspace, both on the 2D surface and in the surrounding 3D volume. This allows each user to clearly observe any other user's actions and object information. However, in certain situations, users may wish to view sensitive data privately and not allow other users to visualize this information. Although privacy gestures on a 2D surface have been explored before [27], these gestures do not facilitate absolute privacy if the hand is not correctly positioned or angled on the multi-touch table relative to other users' viewpoints.

To address this, we introduce a two-handed hybrid privacy gesture, *cross-dimensional privacy pull*, which allows a user to have complete and absolute privacy when visualizing 3D data. Placing a vertical hand on the multi-touch surface to "shield" [27] a cross-dimensional pull performed with the other hand, allows the user to privately visualize the object information in 3D through his head-worn display (Figure 8a–c). After executing the privacy pull, the 3D object is only visible in the head-worn display of the user who performed the interaction, thus ensuring that the information is private.

4.3.3. Cross-Dimensional Privacy Push

The complementary *cross-dimensional privacy push* allows a user to view 2D object data privately. As mentioned earlier, attempting to visualize private information projected on the multi-touch surface using a single or two-handed gesture may not completely obstruct the information from other users. However, when a user places a vertical hand on the surface, while performing a single-handed cross-dimensional push with the other hand, this allows a private view of 2D information in the head-worn display, rendered geometrically as if it were projected onto the 2D multi-touch surface (Figure 8d–f). This private 2D object behaves just like a normal 2D object and can be manipulated using the standard 2D interaction on the surface, but is completely protected from the other user's view. Although we recognize that current head-worn displays are not the most suitable means to view high-resolution 2D data, this allows for complete user privacy, unlike projection onto the multi-user 2D surface.

5. Usability Testing

Twelve individuals volunteered to participate in a pilot study of cross-dimensional gestures in our hybrid environment. The participants were mostly male (11 males, 1 female), between the ages of 17 and 32, recruited by email solicitation, and all were frequent computer users. Only two participants had some previous virtual reality or augmented reality experience. Each participant spent 45 minutes in our hybrid environment, learning about and performing cross-dimensional gestures. Participants were given a post-test questionnaire in which they were asked to rate each gesture's ease of use and intuitiveness (as shown in Figures 9 and 10), and rate how stimulating and satisfying their overall experience was, each on a five-point Likert scale, and provide free-form comments on each gesture and their overall experience. For each user, we first demonstrated how to perform each of the gestures exactly once, and then afterwards asked them to try and perform the gestures. If at any point the user had difficulty performing a particular gesture, we verbally coached them to completion. All twelve users successfully completed every gesture once in the allotted time, with some gestures performed successfully on the first attempt. Once the entire set of gestures was completed, users were encouraged to explore the system during the remainder of their time slot.

For one-handed gestures, participants were moderately successful at performing the interactions with little difficulty. Most participants found both the push and pull to be fairly intuitive (mean push score 4.58, mean pull score 4.00) and easy to use (mean push score 4.08, mean pull score 3.62), especially after a few attempts. For such gestures, we received comments such as "very easy" and "the gesture itself is very intuitive." The pin gesture received mixed reactions, with most of the negative comments addressing hardware limitations of the instrumented glove. However, there were also many positive reactions. One participant stated that performing a pin was "pretty easy" and required a "small amount of learning."

We found similar results for connect, where hardware limitations appeared to frustrate some participants: part of the glove sometimes prevented the user's hand from directly contacting the table while tapping with a closed fist. Further frustrations regarding the intuitiveness of the connect gesture (mean score 4.00), caused one user (who decided it was easier to use two hands to connect) to write that "there was a lack of an intuitive connection between what you do and the result you get." This participant's reaction opens up a future possibility of allowing multiple ways to perform a single gesture, depending on the user's personal preference. Despite this, most participants found the connect gesture to be fairly intuitive, commenting "it was very straight forward" and "it was very easy to do with no mistakes."

The two-handed privacy gestures proved to be the most difficult gestures to perform for many participants. For both privacy pull (ease of use mean score, 3.06) and privacy push (ease of use mean score, 3.35, several users asked for more tolerance when shielding with the left hand, stating that it was difficult to maintain a completely straight and orthogonal (to the side of the table) vertical hand. Some stated that "using two hands is annoying, it would be nice to be able to do it with one hand" and there was a "learning

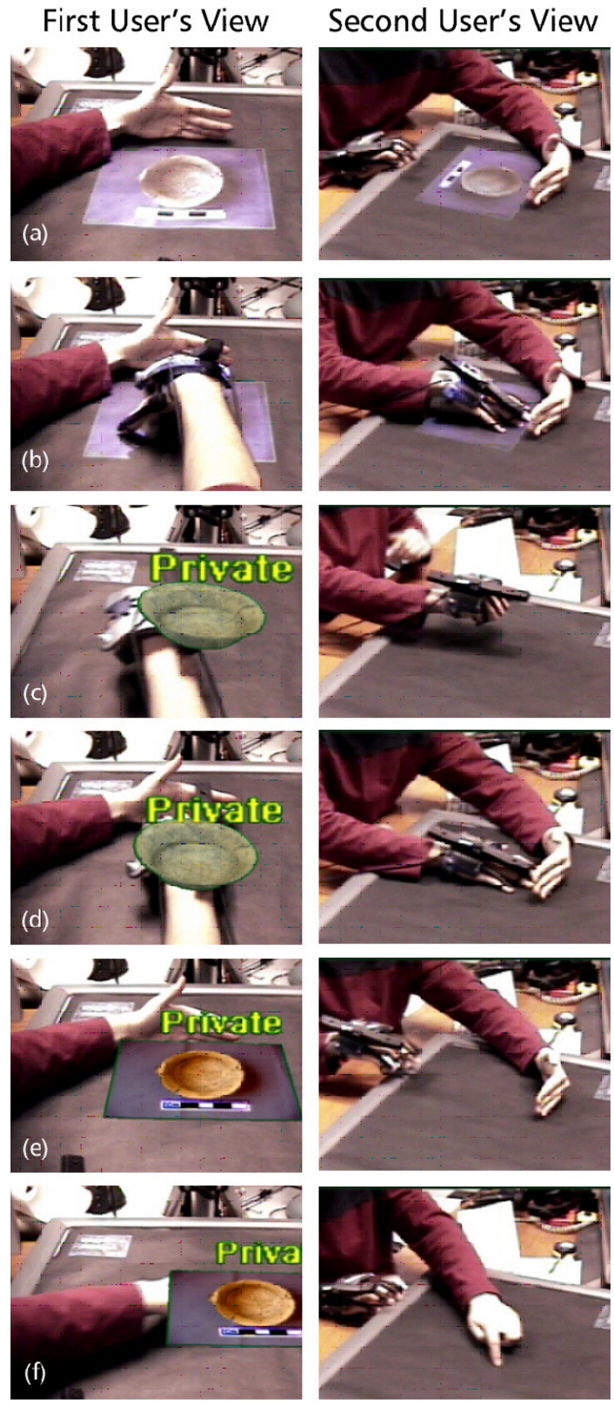

Figure 8: Side-by-side comparison of two users' views when using two-handed privacy gestures. (a) User places a vertical hand on the table to indicate a private gesture. (b) Cross-dimensional pull gesture is performed with the vertical hand above it, yielding a cross-dimensional privacy pull. (c) 3D object is private and invisible to the other user. (d) Performing cross-dimensional push with a vertical hand above it, yielding a cross-dimensional privacy push. (e) 2D object appears as if on the table, but is invisible to the other users. (f) 2D private object behaves like all other 2D objects and is dragged around.

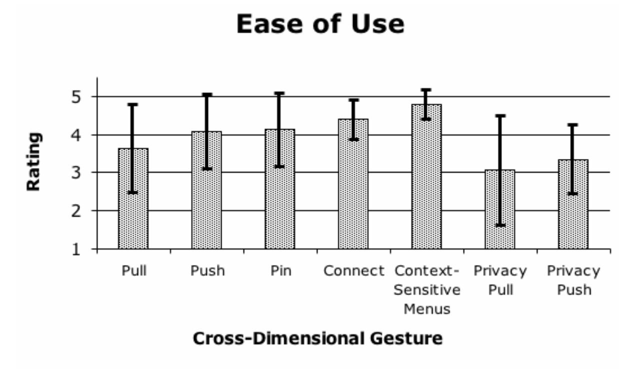

Figure 9: Mean and standard deviation scores for ease of use of cross-dimensional gestures. Participants ranked each gesture on a 5-point scale (1-difficult, 5-easy). Vertical error bars show 1 standard deviation above and below the mean.

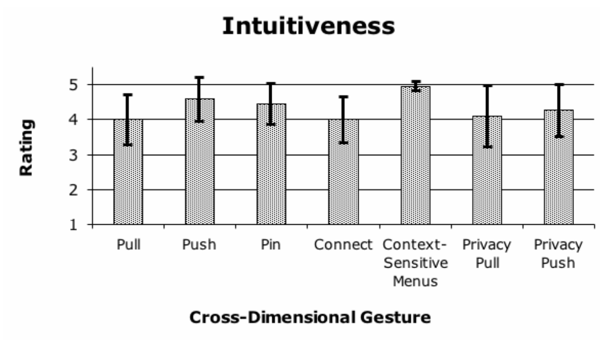

Figure 10: Mean and standard deviation scores for intuitiveness of cross-dimensional gestures. Participants ranked each gesture on a 5-point scale (1-confusing, 5-intuitive). Vertical error bars show 1 standard deviation above and below the mean.

curve for privacy gestures." Some participants, however, stated "it was very easy to remember and use" and "the shield gesture wasn't too difficult to perform."

Overall, the participants were quite stimulated (mean score, 4.83) and fairly satisfied (mean score, 3.84) with the entire experience. We received overwhelmingly positive comments, such as "gigantically interesting" and "it was a very new and cool experience for me." Most made comments expressing their frustrations early on, later followed by positive statements about intuitiveness and ease of use after having practiced the gestures several times. One participant wrote, "Overall, I didn't feel like the gestures were too tough to learn. Within one or two tries, I seemed to pick them up." and "I was really amazed at what was possible."

Our implementation incorporated simple animations to provide continuity and visual feedback. The study participants pointed out that this visual feedback was essential to making hybrid interactions feel like seamless transitions (particularly for cross-dimensional push and pull). Furthermore, many user study participants suggested that additional visual (and audio) feedback should be incorporated into the interactions, especially when gestures were performed incorrectly.

6. Additional Considerations

We have considered the idea of making the rotation a two-handed or two-fingered technique (similar to Wu and Balakrishnan's freeform rotation technique [27]), but in our experience, we have noticed that rotation is clearly specified by fixing the center of rotation to the center of the object and using one additional point to specify the angle. Hence we have decided that it is most simply performed as a one-handed gesture. Furthermore this frees up the other hand to execute additional simultaneous object transformations.

Our gestures are designed to require only one 3D tracked hand. While this is sufficient for the gestures presented here, additional gestures would be possible if both hands were tracked. While we currently track 3D gestures using the tethered P5 glove, which is available only in a right-handed version, we plan to experiment with video-based gesture tracking, which would eliminate the need for a glove and allow for a much more flexible configuration.

7. Conclusions

We developed a set of interaction techniques for hybrid user interfaces that exploit 2D touch-sensitive displays and 3D AR user interfaces. Our system uses synchronized 2D and 3D gestures to facilitate the seamless transition of data and interactions across the dimensional boundary. In addition, our techniques are designed for multi-user collaborative environments, and thus address some of the need for privacy in these shared spaces. Initial user study feedback suggests that users found the gestures fairly intuitive and easy to perform, after a relatively short learning period, possibly due to their close similarity to physical world interactions. While our techniques are presented within an archaeological scenario, we believe that they are applicable to a broad range of hybrid environments.

8. Acknowledgments

We wish to thank Sajid Sadi and Avanindra Utukuri for their support with the Essential Reality P5 glove; Lawrence Wang for his work on DiamondTouch code; Ian Morris, Trinity Jackman, and the Stanford Archaeology Center for images and data from the Monte Polizzo excavation; and Peter Allen, Alejandro Troccoli and Benjamin Smith for the 3D archaeological site models used in VITA. Mitsubishi Electric Research Labs generously donated the DiamondTouch table. This work was funded in part by NSF Grants IIS-01-21239 and IIS-00-82961, Office of Naval Research Contracts N00014-04-1-0005, N00014-99-1-0394, N00014-99-1-0683, and N00014-99-1-0249, and gifts from Alias Systems and Microsoft Research.

9. Video

Work presented in this paper can be previewed in a digital video (DivX encoded) available for download from: www.cs.columbia.edu/graphics/projects/HybridInteraction.

References

[1] Benko, H., Ishak, E. W., and Feiner, S., "Collaborative mixed reality visualization of an archaeological excavation," *Proc. ISMAR 2004 (IEEE and ACM Int. Symp. on Mixed and Augmented Reality)*, pp. 132–140, 2004.

[2] Brown, L., Hua, H., and Gao, C., "A Widget Framework for Augmented Interaction in SCAPE," *Proc. Symp. on User Interface Software and Technology (ACM UIST 2003)*, pp. 1–10, 2003.

[3] Butz, A., Höllerer, T., Feiner, S., MacIntyre, B., and Beshers, C., "Enveloping Users and Computers in a Collaborative 3D Augmented Reality," *Proc. IWAR '99 (Int. Workshop on Augmented Reality)*, pp. 35–44, 1999.

[4] Cutler, L. D., Fröhlich, B., and Hanrahan, P., "Two-Handed Direct Manipulation on the Responsive Workbench," *ACM Symposium on*

Interactive 3D Graphics, 1997.

[5] Dietz, P. and Lehigh, D., "DiamondTouch: a Multi-User Touch Technology," *Proc. ACM Symp. on User Interface Software and Technology (UIST '01)*, pp. 219–226, 2001.

[6] Feiner, S. and Shamash, A., "Hybrid User Interfaces: Breeding Virtually Bigger Interfaces for Physically Smaller Computers," *Proc. ACM Symp. on User Interface Software and Technology (UIST '91)*, pp. 9–17, 1991.

[7] Herndon, K. P., Zeleznik, R. C., Robbins, D. C., Conner, D. B., Snibbe, S. S., and van Dam, A., "Interactive Shadows," in *Proc. Symp. On User Interface Software and Technology (ACM UIST 1992)*: ACM Press, 1992, pp. 1–6.

[8] Hinckley, K., "Synchronous Gestures for Multiple Persons and Computers," presented at Proc. Symp. on User Interface Software and Technology (ACM UIST 2003), Vancouver, BC, 2003.

[9] Ishii, H., Kobayashi, M., and Arita, K., "Iterative design of seamless collaboration media," *Communications of the ACM*, vol. 37, pp. 83–97, 1994.

[10] Ishii, H. and Ullmer, B., "Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms," in *Proc. Conference on Human Factors in Computing Systems (CHI '97), (Atlanta, March 1997)*: ACM Press, 1997, pp. 234–241.

[11] Kaiser, E., Olwal, A., McGee, D., Benko, H., Corradini, A., Li, X., Cohen, P., and Feiner, S., "Mutual disambiguation of 3D multimodal interaction in augmented and virtual reality," *Proc. ICMI 2003 (Fifth Int. Conf. on Multimodal Interfaces)*, pp. 12–19, 2003.

[12] Krueger, M., "VIDEOPLACE and the interface of the future," in *The art of human computer interface design*, B. Laurel, Ed. Menlo Park, CA: Addison Wesley, 1991, pp. 417–422.

[13] Kumar, S., Cohen, P., and Levesque, H. J., "The Adaptive Agent Architecture: Achieving Fault-Tolerance Using Persistent Broker Teams," *Proc. Int. Conf. on Multi-Agent Systems*, pp. 159–166, 2000.

[14] Low, K. L., Ilie, A., Welch, G., and Lastra, A., "Combining head-mounted and projector-based displays for surgical training," *Proc. IEEE VR 2003*, pp. 110-117, 2003.

[15] Mine, M., R., Brooks, F., P., Jr., and Sequin, C., H., "Moving Objects in Space: Exploiting Proprioception In Virtual-Environment Interaction," in *Proc. ACM SIGGRAPH '97*. Los Angeles, CA, 1997, pp. 19–26.

[16] Nishino, H., Utsumiya, K., Kuraoka, D., and Yoshioka, K., "Interactive Two-Handed Gesture Interface in 3D Virtual Environments," *Proc. of the ACM Symposium of Virtual Reality Software and Technology (VRST'97)*, pp. 1–8, 1997.

[17] Olwal, A., Benko, H., and Feiner, S., "SenseShapes: Using statistical geometry for object selection in a multimodal augmented reality system," *Proc. IEEE and ACM Int. Symp. on Mixed and Augmented Reality (ISMAR 2003)*, pp. 300–301, 2003.

[18] Pierce, J. S., Forsberg, A., Conway, M. J., Hong, S., Zeleznik, R., and Mine, M. R., "Image Plane Interaction Techniques In 3D Immersive Environments," in *Symposium on Interactive 3D Graphics*: ACM SIGGRAPH, 1997.

[19] Regenbrecht, H. T., Wagner, M. T., and Baratoff, G., "MagicMeeting: A Collaborative Tangible Augmented Reality System," *Virtual Reality - Systems, Development and Applications*, vol. 6, pp. 151–166, 2002.

[20] Rekimoto, J., "Pick-and-Drop: A Direct Manipulation Technique for Multiple Computer Environments," *Proc. UIST '97*, pp. 31–39, 1997.

[21] Rekimoto, J., "SmartSkin: an Infrastructure for Free-hand Manipulation on Interactive Surfaces," *Proc. ACM SIGCHI '02*, pp. 113–120, 2002.

[22] Rekimoto, J. and Saitoh, M., "Augmented Surfaces: A Spatially Continuous Work Space for Hybrid Computing Environments," *Proc. ACM CHI '99*, 1999.

[23] Schmalstieg, D., Fuhrmann, A., Hesina, G., Szalavari, Z., Encarnação, L. M., Gervautz, M., and Purgathofer, W., "The Studierstube Augmented Reality Project," *Presence: Teleoperators and Virtual Environments*, vol. 11, 2002.

[24] Streitz, N., Geißler, J., Holmer, T., Konomi, S., Müller-Tomfelde, C., Reischl, W., Rexroth, P., Seitz, P., and Steinmetz, R., "i-LAND: an interactive landscape for creativity and innovation," *Proc. ACM SIGCHI '99*, pp. 120–127, 1999.

[25] Tandler, P., Prante, T., Müller-Tomfelde, C., Streitz, N., and Steinmetz, R., "Connectables: dynamic coupling of displays for the flexible creation of shared workspaces," *Proc. Symp. on User Interface Software and Technology (ACM UIST '01)*, pp. 11–20, 2001.

[26] Wexelblat, A., "An Approach to Natural Gesture in Virtual Environemnts," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 2, pp. 179–200, 1995.

[27] Wu, M. and Balakrishnan, R., "Multi-Finger and Whole Hand Gestural Interaction Techniques for Multi-User Tabletop Displays," *Proc. Symp. on User Interface Software and Technology (ACM UIST 2003)*, pp. 193–202, 2003.

[28] Zeleznik, R. C., Herndon, K. P., and Hughes, J. F., "SKETCH: An Interface for Sketching 3D Scenes," *Computer Graphics*, vol. 30, pp. 163–170, 1996.