Use Your Head: Tangible Windows for 3D Information Spaces in a Tabletop Environment

Martin Spindler[1], **Wolfgang Büschel**[2], **Raimund Dachselt**[2]
[1] User Interface & Software Engineering Group, University of Magdeburg, Germany
[2] Interactive Media Lab Dresden, Technische Universität Dresden, Germany
{martin.spindler, bueschel, dachselt}@acm.org

ABSTRACT

Tangible Windows are a novel concept for interacting with virtual 3D information spaces in a workbench-like multi-display environment. They allow for performing common 3D interaction tasks in a more accessible manner by combining principles of tangible interaction, head-coupled perspective, and multi-touch techniques. Tangible Windows unify the interaction and representation space in a single device. They either act as physical peepholes into a virtual 3D world or as physical containers for parts of that world and are well-suited for the collaborative exploration and manipulation of such information spaces. One important feature of Tangible Windows is that the use of obtrusive hardware, such as HMDs, is strictly avoided. Instead, lightweight paper-based displays are used. We present different techniques for canonical 3D interaction tasks such as viewport control or object selection and manipulation, based on the combination of independent input modalities. We tested these techniques on a self-developed prototype system and received promising early user feedback.

Author Keywords

Tangible windows; head interaction; head tracking; multiple views; magic lenses; tabletop displays; multi-surface user interfaces; fish tank virtual reality; head-coupled perspective; 3D interaction.

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION

Interacting with 3D content on tabletops has received considerable attention in the past, e.g., [3, 16, 18, 29]. One reason for this is the desire to transfer advantages of touch-based interfaces – which are associated with being natural and intuitive – to a variety of application domains that deal with complex 3D data. Examples include the architectural design of 3D buildings and their surroundings, surgical planning based on 3D imagery, and interactive 3D scientific visualizations. Inevitably, these types of applications implicate challenges that are caused by mapping 3D spaces onto a 2D surface. Therefore, extending the tabletop to a third dimension was and still is a major goal of many research projects. Grossman and Wigdor [15] compiled a taxonomy of prevalent approaches that come from areas as diverse as interactive 3D graphics, e.g., [38], virtual reality, e.g., [25], and augmented reality, e.g., [7]. In their taxonomy, they distinguished between the "actual" and the "perceived" display space and identified unexplored gaps for several combinations of both parameters.

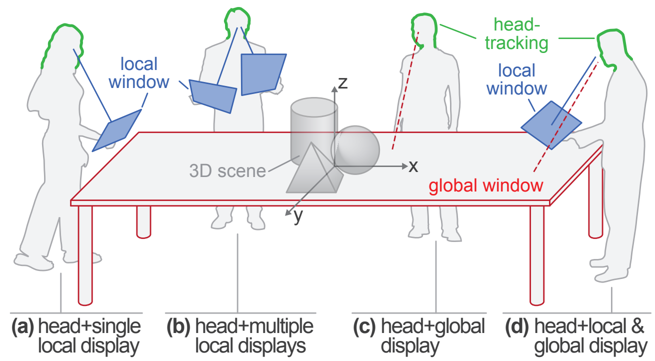

Figure 1. A tabletop (*global display*) serves as the shared view of a virtual 3D scene that is spatially aligned with room and table. This setup is enriched by multiple tracked mobile *local displays* that act as personal views into this scene. The system is also capable of tracking head positions & orientations, which is the basis for head-coupled perspectives on local and global displays.

With Tangible Windows, we fill one of these gaps. In particular, we address stereo spatial augmentation [15] in multi-display tabletop environments. This is the combination of projecting imagery onto physical proxies and the tabletop and using head-coupled perspectives to provide a 3D volumetric perception anywhere in the working volume. Tangible Windows are lightweight paper-based displays like in [30] that are coupled with a user's head for the purpose of auto-perspective. They act as peepholes into a virtual 3D information space or serve as tangible containers for parts of it and are controlled by physically moving them through the volume on or above a tabletop surface (see Figure 1).

The primary contributions of this paper are: (1) The integration of three independent input modalities, i.e., display locations/orientations, head locations/orientations, and touch/pen input, into a unifying interaction model. (2) A set of novel 3D interaction techniques. (3) A technical solution for lightweight spatially aware displays, associated with a virtual 3D information space (or parts of it), supporting multiple displays for both single users and collaborative work. Although various so-

ITS'12, November 11–14, 2012, Cambridge, Massachusetts, USA.

lutions for head-coupled perspectives already exist, e.g., [1, 4, 14, 26, 38], an integration into a single, flexible multi-display system, suitable for co-located collaboration, has not been done before.

The remainder of this paper is organized as follows. After presenting an overview of related work, we discuss the design space from which we derive the Tangible Windows concept. Based on that, we introduce basic 3D interaction techniques, present example application scenarios and summarize our technical setup. We conclude with a discussion of initial user experiences and an outline of future work.

RELATED WORK

Volumetric Displays

Perception of 3D in the real world is accomplished by a number of visual depth cues, such as occlusion, perspective, shading, shadows, parallax, etc., which can be conveyed by 3D displays. 3D display technologies can be categorized as volumetric or geometric (see [12] for an overview). Volumetric displays directly present 3D information by illuminating points in real-world spatial locations, e.g., cubic static-volume implementations with visible gas suspended in glass [11]. Although such technology provides very realistic results, it is still too limited in many ways, e.g., in terms of resolution, mobility, weight and brightness, as well as by high costs.

Geometric Displays & Head-coupled Perspective

Geometric displays render images on one or more 2D displays with a perspective corrected to the user's view (head-coupled perspective). In this context, Ware et al. [38] coined the term Fish Tank Virtual Reality (FTVR) that either provides monocular or stereoscopic views for one perspective (or user) on a vertical display (desktop monitor). Here, shutter glasses are used to multiplex multiple perspectives, either to provide monocular views for two users or a single stereoscopic view. The concept of FTVR has been applied to a variety of form factors. The Responsive Workbench is one variation that uses FTVR on a horizontal tabletop with support for one [25] or two users [1] simultaneously. Unlike our approach, however, their setup does not include the projection on mobile displays. Hancock et al. [17] show that a neutral center of projection combined with parallel projection geometry provides a reasonable compromise for multi-user tabletop setups, which we make use of in our system. In the last years, FTVR has become more popular on consumer electronics, such as the Wii [26] and the iPad [14]. Another example for recent commercial products making use of FTVR is the NettleBox[1] tabletop display. Recently, de Almeida et al. [10] used head-coupled perspective to allow looking "behind" the bezels of multi-display walls.

See-Through Interfaces & Tangible Magic Lenses

The see-through interface, as introduced by Bier et al. [5], is a virtual interface (the Magic Lens) that appears between users and 2D applications for the purpose of local presentation/manipulation of data. It was later extended to 3D [37] and has also been widely adopted in 3D virtual reality (VR) systems. Head-mounted displays (HMD) provide immersive see-through-like stereo views by using small displays located directly in front of the eyes. Although there are lightweight projector-based variants of HMD [7] that help attenuate physical exhaustion that occurs with active HMD [33], HMD systems often suffer from significant mismatches between the virtual and real world.

With spatially-aligned displays these problems are less distracting (but still present) because they are used farther away from the eyes. With their metaDesk [36] system, Ullmer and Ishii introduced a tangible version of the Magic Lens concept that allows users to physically interact with a see-through interface. Here, tracked mobile displays that users can hold in their hands serve as physical peepholes into a VR workspace. As demonstrated by Fitzmaurice [13], moving these displays, e.g., around the user in a donut-like shape, allows users to explore these workspaces. Yee [41] later extended this idea by combining spatial input with pen input. Another prominent example is Boom Chameleon [35], where a single pen-enabled touch screen is mounted on a mechanical arm for the purpose of exploring/annotating a virtual car model.

Recent trends show an increased use of passive (projective) handheld displays, where digital image content is projected onto tracked paper-like screens, e.g., as shown by PaperWindows [22]. This type of passive handheld display technology has been integrated into tabletop environments where it was used, e.g., for the exploration of various information spaces by moving the tangible lenses through the physical 3D space on or above a tabletop surface [30, 31]. With their SecondLight, Izadi et al. [23] presented a more self-contained approach by using switchable diffusers to allow rear-projection on both a tabletop and passive mobile displays above it. This approach also allows for Frustrated Total Internal Reflection (FTIR) touch input on both tabletop and mobile displays, something that our system is not capable of. Chan et al. [9] combine a normal RGB- and an IR-projector to display invisible markers on a tabletop. Using an IR-camera on a mobile device, they can provide 6DOF tracking of the device's position above the tabletop. In comparison, our setup, employing a second projector, provides a bigger interaction volume and allows using handheld displays not only above but also next to the tabletop. Furthermore, these systems have not integrated the use of the head as an additional input channel or address the special requirements of 3D manipulation within a VR environment.

3D Interaction on the Tabletop and beyond

There has been extensive research in the field of multi-touch for 3D interaction, which has proven to be an excellent basis for many powerful 3D interaction techniques, e.g., [3, 16, 29]. Hilliges et al. [21] use an IR-transparent diffuser to facilitate such interaction above the table screen by tracking finger gestures as input. Wilson et al. [40] presented a system for interaction on and above non-instrumented surfaces using consumer depth cameras and a projector setup. A comparable setup, albeit for mobile indoor projections, has recently been shown by Molyneaux et al. [27]. Benko et al. [4] presented the MirageTable, which consists of a curved screen, a depth sensor, and a stereo projector and provides head-coupled perspective through the tracking of shutter glasses worn by the

[1] http://www.nttl.ru

user. Although differing in terms of interaction, these systems share some characteristics with our approach. Tracking with depth cameras, however, still has limited precision and reliability. Thus, we decided for a more traditional optical tracking of IR-reflective markers.

DESIGN SPACE

We envision a system that enables users to directly interact with a virtual 3D world or parts of it in a more accessible manner, compared to traditional setups. Our intent is that users are not required to wear any obtrusive hardware, such as HMD or heavy display/tracking equipment. Instead, all necessary hardware shall be hidden so that users are not aware of its presence. This is to ensure a high degree of immersion combined with an experience as natural as possible.

This goal shall be accomplished by using mobile spatially-aware lightweight displays and additional – much larger – stationary displays that are spatially aligned with the environment. Although vertical wall-sized screens, the CAVE™ and other large 360° displays are possible alternatives, we primarily focus on a horizontal tabletop setup that we will briefly introduce in the following.

General System Design and Components

We propose a multi-display system that consists of two basic types of displays (see Figure 1): a single stationary tabletop screen (*global display*) and one or multiple mobile screens (*local displays*) that provide independent views into a *3D information space*. The system keeps track of the *heads* of *one or multiple users*. The main components of the system are:

3D Information Space According to Bowman et al. [6], typical interaction tasks that users intend to accomplish within virtual 3D environments (3D information spaces) are *navigation*, *selection* and *manipulation*. They commonly involve changing the viewpoint and applying transformations onto individual objects or the whole 3D scene. The purpose of the system shall be the exploration and manipulation of a global *virtual 3D scene* that is (dynamically) aligned with the physical room and is seen from an exocentric (outside-in) view (see Figure 2). This 3D information space can be associated with global and local displays in various ways. For this purpose, a *global coordinate system* is defined with the center of the table surface being the origin and the Z-axis pointing to the ceiling (see Figure 1).

Global Display The horizontal tabletop display features a large interactive screen that usually shows global context information shared by all users. For example, this can be a view into parts of the global 3D scene that virtually resides below, on, or even (partly) above the table surface (see Figure 2).

Local Displays Local displays are spatially-aware lightweight interactive displays that share some similarities with the tangible magic lenses introduced in the PaperLens project [30]. Our system supports different sizes and shapes, such as rectangles and circles. The main objective of local displays is to provide several users with personal views into the virtual 3D world. These can be views that are either in complete synchronicity with the global scene, show some local modifications (e.g., annotations), or provide alternative representations (e.g., wireframe rendering, see Figure 5, right) and perspectives/mappings (e.g., fisheye distortions). They can even show a completely different virtual 3D scene or an object that is exclusively attached to it. We support the simultaneous use of multiple local displays to facilitate co-located collaboration.

Heads Chan et al. [8] found that their head-controlled document zooming was fairly ineffective and that document manipulation and interaction should rather be handled as an active process than a passive process, e.g., by supporting iconic gestures [24]. In contrast, using the head for secondary tasks, e.g., by providing users with the right perspective [28] or by presenting different levels of detail depending on the distance of head and display [20], can help making interaction more natural. One important characteristic of the head is that it becomes more meaningful when it is coupled with a single local display (see Figure 1a) or the tabletop (see Figure 1c), e.g., for the purpose of head-coupled perspectives. These basic combinations in turn can arbitrarily be combined again, such as by coupling one user's head with either multiple local windows (see Figure 1b) or with a local and the global window simultaneously (see Figure 1d).

One vs. Multiple Users The number of users is another defining factor of the system and does not necessarily match the number of local displays in use. In fact, the combination of (typically one) global display and multiple local displays for one or many users opens up additional exciting possibilities. For example, two users could temporarily share a single local display for collaboration purposes. Alternatively, a single user could use two local displays, e.g., for bimanual interaction that could be used for comparison tasks (see Figure 1b). Beyond that, if only one user is present, the tabletop display is not shared anymore and thus becomes a personal view. This allows to provide the only user with special views directly on the tabletop screen that could, e.g., depend on where the user is standing or looking at.

General Types of Input

There are three fundamental types of input that constitute the interaction space of our proposed system: *on-surface input*, *spatial input*, i.e., *with-device input*, and *head input*.

On-Surface Input Interaction can be performed directly *on* the surface of displays. This applies to both local (mobile) displays and the global (tabletop) display. In particular, we consider different variants of touch and digital pen input.

Spatial Input One important feature of local displays is that they are being tracked in physical space with six degrees of freedom (6DOF). This allows users to directly interact *with* a local display by grabbing it with their hands and then moving or rotating it through the physical space. This way, a rich set of interaction techniques becomes available. As a first but non-exhaustive attempt, see the interaction vocabulary described in [31].

Head Input Heads can be considered as input devices with 6DOF. In this respect, they provide an additional interaction modality that is useful, e.g., for the purpose of head-coupled

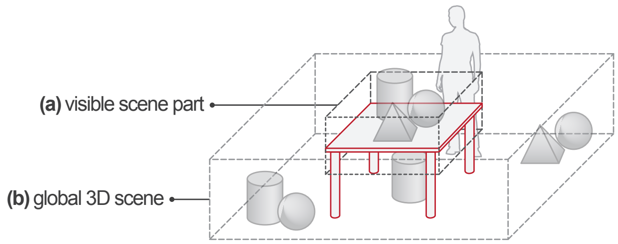

Figure 2. The 3D scene is situated around the table. Only a part of this larger scene is shown on the tabletop at any time.

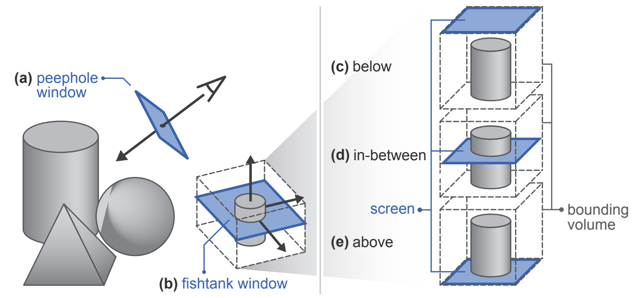

Figure 3. Tangible Windows can be associated with virtual 3D content in two principle ways: as *peephole windows* (a) that serve as physical peepholes into a virtual world, or as *fish tank windows* (b) acting as physical containers for virtual 3D objects that can reside below the display surface (c), above it (e), or somewhere between the two extremes (d).

perspectives [38]. Beyond that, observing head movements also helps to distinguish between users in terms of user ID and user position, which can be a valuable information for systems supporting user collaboration.

THE TANGIBLE WINDOWS CONCEPT

After having introduced the general design space and system components, we will now provide a definition of Tangible Windows and discuss some of their properties.

A Definition of Tangible Windows

We define a Tangible Window as a spatially-aware display that is typically coupled with a user's *head*. This coupling between display and head is particularly intended for the purpose of interacting with a virtual 3D information space in such a way that Tangible Windows are peepholes into it or that they act as physical proxies for parts of it. For this purpose, Tangible Windows serve as physical handles that users can grab and move around with their hands. In this regard, they are not only a tool of direct *representation*, but also a tool of direct *interaction*, seamlessly integrating input and output.

Windows into Virtuality

We distinguish between two basic types of Tangible Windows that differ in how they are associated with virtual 3D scenes (see Figure 3): *fish tank windows* and *peephole windows*.

Fish Tank Windows One way of linking a virtual 3D scene with a Tangible Window is to attach the 3D content so that both always remain aligned with each other ("scene in hand" metaphor [39]). For this purpose, a *fish tank window* provides its own local coordinate system, as illustrated in Figure 3. Moving or rotating a fish tank window through the physical space also moves and rotates the attached 3D content with it. Hence, a fish tank window can be thought of as a visual and physical container for arbitrary virtual 3D scenes.

Peephole Windows Another way of associating a Tangible Window with a virtual 3D world is the *peephole window* that employs the "eyeball in hand" metaphor [39]. Such *peephole windows* serve as windows into the virtual 3D environment that show what virtually exists behind them [13]. Moving or rotating a *peephole window* directly influences its view frustum and view orientation into the virtual 3D space. The see-through effect is usually achieved by maintaining a direct one-to-one spatial mapping between the real and virtual space so that both spaces share the same global coordinate system. However, other mappings are also possible. This may include position and orientation offsets, clipping planes (e.g., useful for slicing), and simulating a mirror.

Mapping 3D Volumes Onto 2D Surfaces

Each Tangible Window employs a canonical 3D bounding volume in the shape of an extruded surface geometry. Typically, with rectangular-shaped displays this will be a box, although we explicitly support different shapes, e.g., circles. The XY-plane of the bounding box always remains aligned with the 2D display and the Z-axis is defined to be orthogonal to the display plane, see Figure 3.

This describes an extent into the third dimension that we can use for mapping arbitrary 3D content onto the 2D surface. Motivated by the *zero-parallax-plane* (see Section Head-coupled Perspectives), three canonical cases are possible: Considering the window surface as a dividing plane held horizontally, we distinguish between objects residing completely *above* it, objects being attached entirely *under* the window, and objects that *intersect* the window plane, i.e., some parts of the 3D content lie below and others above it (see Figure 3).

Head-coupled Perspectives

While VR displays often allow for stereoscopic rendering of 3D scenes, we will not focus on this aspect. Instead, we will use monoscopic views that only address cues that are processed by one eye. Such *monocular cues* involve relative size, texture gradients, linear perspective, occlusion, and motion parallax. Head-dependent control of camera parameters (*head-coupled perspective*) is a technique that allows for reproducing such cues. With it, a more realistic simulation of perspective effects can be created that comes close to what users would expect from their real-world experiences and thus helps making the overall interaction experience more natural.

Technically, this is achieved by setting the *zero-parallax-plane* – a determinant factor for off-axis perspective projections – in such a way that it remains aligned with the window surface. This ensures that no visual shifts occur within this particular plane and thus parts of the 3D object that precisely fall onto the window plane appear to be fixed to it. Opposed to that, the farther away an object part resides from the display plane (in terms of positive or negative Z-distance), the more it is being shifted. For example, when users move their head to the right,

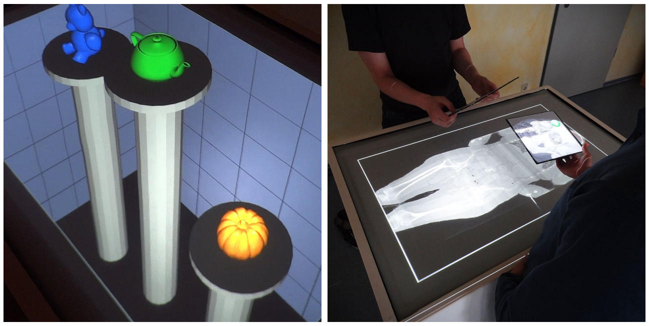

Figure 4. Head-coupled perspective allows users to look at the global 3D scene from arbitrary sides by simply walking around the table (left). One use case is the exploration of medical volume data (right, mockup).

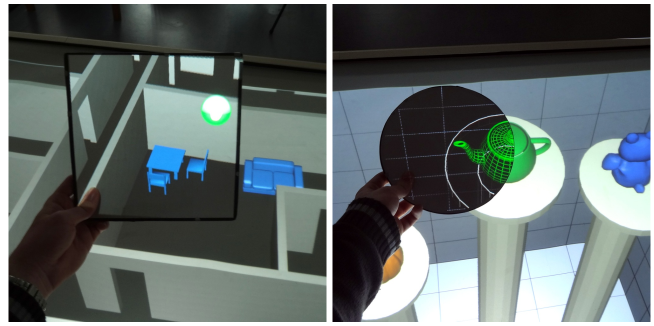

Figure 5. Users can independently explore a virtual 3D world with mobile *peephole windows* that either show photo-realistic views (left) or alternative representations, such as wireframe renderings (right).

objects above the table will appear to move left while those behind it will seem to move right (motion parallax).

We distinguish between two major effects that are based on the head-coupled perspective: the *fish tank VR* effect and the *head-coupled peephole* effect. Both effects are controlled by moving the head and/or a display with respect to each other.

Fish Tank VR Effect The illusion of holding a virtual object physically in the hands by using a fish tank window as a visual container can further be improved by applying the concept of fish tank VR [38]. With it, 3D objects are always shown from the correct viewpoint – the one of the user (observer). While the original fish tank VR technique only addressed conventional desktop monitors, it was later used for horizontal workbench-like VR stereo displays [25], which is similar to what we do with our tabletop. More recently, fish tank VR was shown for the iPad [14] that enabled users to hold 3D objects virtually in their hands. We extended this approach by providing a seamless integration into a multi-display tabletop environment, where every display can provide an individual fish tank view for each user. Beyond that, we support passive displays allowing us to project image content not only onto the front side of a Tangible Window but also onto the back side. This is necessary, e.g., for interaction techniques that employ display flipping.

Head-coupled Peephole Effect Using head-coupled perspectives for *peephole windows* reverts the "eyeball in hand" metaphor into something that could be called "eyeball in place" metaphor. Here, the *peephole window*'s own orientation does not effect the view direction anymore. Instead, the viewer's position defines which part of the scene is visible through the peephole window. This also allows users to hold a *peephole window* less accurately, e.g., when they look at a particular object of the global scene.

INTERACTION TECHNIQUES

We will now illustrate how users can accomplish the most common 3D interaction tasks with Tangible Windows: object selection and manipulation, including moving, copying and deleting objects, as well as viewport control and global navigation. To support this set of basic interaction techniques, we employ the three major input modalities provided by Tangible Windows, *on surface input* (pen/multi-touch input), *spatial input* (moving/rotating a window through the physical space), and *head input* (moving/rotating the head). We will focus on interaction techniques that are mostly based on head-assisted *spatial* input. We believe that such kinesthetic techniques will help users by utilizing their spatial memory, as supported by evidence presented in, e.g., [34]. This implies that while the physical interaction is taking place, users unconsciously acquire spatial knowledge about the overall virtual 3D scene. This supports them in recalling (wayfinding) and coming back (traveling) to a certain region of interest in a natural way.

Global Viewpoint Control on the Tabletop

The tabletop display plays an integral role for the interaction with Tangible Windows and serves as a physical reference that can be used for constraining the interaction to a particular 2D plane. Therefore, defining the tabletop's view is among the most important interaction tasks. With head-coupled perspective enabled, global camera parameters such as viewpoint and angle of aperture are solely controlled by using head positions with respect to the table. This enables the user to explore the global scene from different sides by walking around the table (see Figure 4). Head-coupled perspectives do not work well when two or more users look at the same display (tabletop). In such scenarios, averaged head positions could be used for people standing very close together or the shared tabletop display could be divided into partitions that provide a unique view for each user [19]. For our system, we decided to provide a single movement-independent default view that is a good compromise for most users. As suggested by [17], we opted for orthographic projections with the center of projection being directly above the table.

Scene Exploration (with Mobile Peephole Windows)

Scene exploration as one very basic task of 3D interaction is not only supported by changing the tabletop's view. We also use mobile *peephole windows* that serve as personal peepholes into the global scene to enable multiple users to examine the virtual 3D world independent from each other (see Figure 5). They provide easy viewport control by allowing users to physically move/rotate them through the space above the tabletop, while their physical orientation usually coincides with their view orientation. By this means, for instance, it is possible to orbit around a single virtual object within the overall virtual 3D world by just walking around its fixed virtual

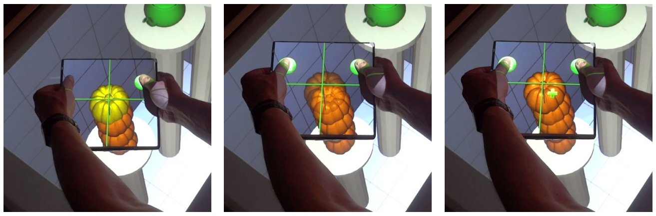

Figure 6. Object selection with a *peephole window*. The object targeted by the selection stick of the 3D cursor-stick gets highlighted (left). Once selected, the object can be moved (center) or copied (right).

center in physical space. Various systems have demonstrated similar capabilities, with the Boom Chameleon [35] being a prominent example. In contrast to others, our solution does not require mechanical arms for tracking, supports multiple displays simultaneously, and provides a seamless integration into a tabletop environment. Besides conventionally rendered views (see Figure 5, left), *peephole windows* can also show filtered representations, such as non-photorealistic or wireframe renderings (see Figure 5, right).

Object Selection (with Mobile Peephole Windows)

A common interaction task is selecting (picking) a particular object that is currently visible in a *peephole window*, e.g., for further inspection or manipulation. For this purpose, a 3D cursor-stick is displayed in the center of the window (see Figure 6). It can be thought of as a virtual stick that extends into the scene and keeps traveling along with the window. The 3D cursor-stick is activated by slightly pushing a pressure-sensitive button on the window and it is deactivated as soon as the user releases the button again. It serves both as a visual cursor and a ray that a user can point to any near or distant object. Objects appearing under the 3D cursor-stick automatically get visually highlighted. For this purpose, a ray is cast into the global scene. This makes it possible to select remote objects, e.g., objects below the table surface, which otherwise would be difficult to reach. We currently only consider the object that is closest to the window. However, more sophisticated approaches are possible, such as considering all available targets or letting users choose from a sorted list, see also [6]. Once an object is visually highlighted, it is a candidate for *snap-to-hand selection* and *snap-to-stick selection*.

Snap-to-Hand Selection Snap-to-hand selection is triggered by double-tapping a button on the window (see Figure 7). Then, a copy of the candidate object is created. The view changes and the copy is displayed in a hand-held *fish tank window*. For users, this has the effect that the object is brought from the distance directly into their hands in an animated way, allowing them to further examine and manipulate it.

Snap-to-Stick Selection Snap-to-stick selection (see Figure 6) is activated by pushing the pressure-sensitive button a little harder than before (remember that the 3D cursor-stick is only visible as long as the user slightly presses this button). Once the candidate object becomes selected, a half-transparent ghost object will be created that gets fixed to the ray at the current hit point and will remain there until the user releases the button again. In a way, this is similar to the real-world example of spiking a strawberry with a chopstick. As we will see next, this "fruit on a stick" metaphor is useful for various manipulation tasks.

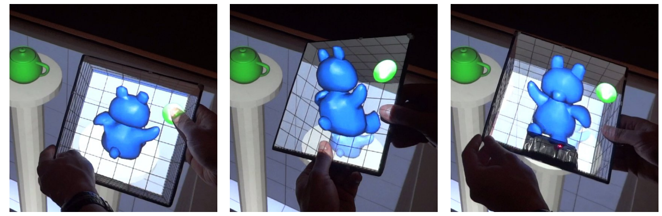

Figure 7. Mobile *fish tank windows* serve as physical proxies for virtual objects that users can inspect by moving/rotating the display. The backside of an object can be inspected by flipping the window.

Object Manipulation (with Mobile Peephole Windows)

Snap-to-stick selection implies that while a user is maneuvering a *peephole window* through the physical space, the object's ghost will be moved along with it, while always keeping the same distance to the window surface. At the same time, the original object is still visible at its original position in the global scene serving as a visual reference (see Figure 6, center). We use this configuration as the basis for the following manipulation tasks that are also inspired by the "drag and drop" metaphor employed on conventional desktop computers. For simplicity, we only address standard manipulation tasks, such as *moving*, *duplicating*, and *deleting* an object.

Object Moving Once the user is satisfied with the new position and orientation of the ghost, s/he only needs to release the button in order to permanently replace it with the original object. Along with this, the object at its original position is being removed and the 3D cursor-stick fades out. At any time, users can cancel the entire move operation by shaking the *peephole window* for more than one second.

Object Copying The above technique can also be used to create a duplicate of an object and to place it somewhere into the global scene. For this purpose, a special modifier button on the display needs to be pressed with the non-dominant hand. This is similar to holding a modifier key when performing a drag and drop operation in a file browser. As long as the button is pushed, a special copy icon appears on the *peephole window* indicating the changed interaction mode. As soon as the modifier button is released, the old interaction mode is restored. This allows users to seamlessly switch between MOVE and COPY operations.

Object Deletion A selected object can be removed from the global scene by dragging and then dropping it into the physical area besides the table. This is accomplished with the move operation as explained above. Once the proxy is released into the void, an icon on the window indicates that the object is to be deleted. In order to prevent unintentional object removals, the user can confirm or cancel the operation by pushing a button.

Beyond these standard operations, other more complex types of manipulation could be realized, too, for example, by using a second Tangible Window as a knife that cuts away parts of the objects or deforms it (two-hand interaction).

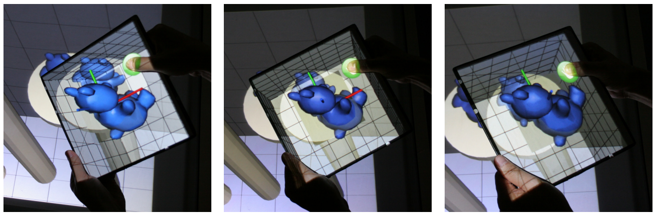

Figure 8. For precise control of how an object is oriented on a display we use *object clutching* that temporarily decouples the object from the window. It is activated by pressing (and holding) a button on the display.

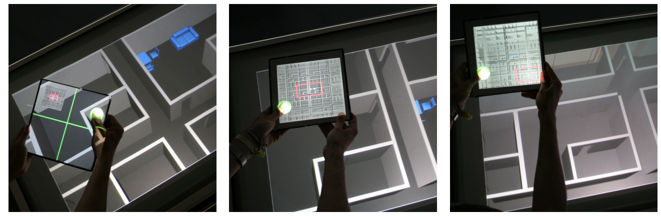

Figure 9. With the help of tangible World in Miniature views, *scene-on-stick-dragging* (left) and *scene-in-hand-dragging* (middle, right) allow users to control what parts of the scene are displayed on the table.

Object Inspection (with Mobile Fish Tank Windows)

After having selected a particular object with the *snap-to-hand selection* technique, a user can examine the selected object from various sides by simply holding, moving and rotating the *fish tank window* with the hands (see Figure 7). Along with head-coupled perspectives this usually results in a convincing experience.

Object Flipping Due to the window's flat nature, usually only one hemisphere of an associated object can be shown on a window's screen, e.g., the front side of a virtual teddy bear. To lessen this problem, we support window flipping, i.e., the object's back side will be displayed on the window after it is flipped (see Figure 7).

Object Clutching For more precise object rotations on the window we use clutching [13], which enables users to control how an object is aligned to the display in terms of orientation. This is achieved by pushing a button on the window. As long as this button is pressed, any rotation of the window will not affect the object's orientation with respect to the physical world, see Figure 8.

Global Navigation on the Tabletop

The global scene can reach far beyond the physical boundaries of the table and thus often only parts of it are displayed on the tabletop (see Figure 2). In order to allow users to explore all parts of the scene, they need to be able to change its overall position with respect to the physical location of the table. This usually involves panning and lifting/lowering the global scene until a particular region of interest shows up. To provide users with such functionality, we utilize the concept of World in Miniature (WIM) [2, 32] – an overview & detail technique that we propose to make *tangible* with Tangible Windows. This is achieved by showing an overview map of the entire scene on the mobile *fish tank window*, where the parts of the scene that are currently displayed on the tabletop are visually highlighted by a semi-transparent box (see Figure 9). We support the following two techniques.

Scene-on-Stick-Dragging During scene exploration with a *peephole window*, a user can select (virtually harpoon) the entire scene by using the *snap-to-stick selection* technique. This is achieved by pointing the 3D cursor-stick to the background of the scene. When the pressure-sensitive button is pushed hard enough, a small overview map appears in the left upper corner of the *peephole window* indicating that the user has control over the position of the global scene (see Figure 9, left). Since the scene is now locked to the handheld window, moving it also moves the scene with respect to the table until the user releases the button again.

Scene-in-Hand-Dragging While *scene-on-stick-dragging* allows for a more precise control of the global scene, it is not suitable for adjustments on a more coarse level, e.g., consider a global scene that consumes 20 m of physical space. For this reason, we developed *scene-in-hand-dragging* that allows users to hold a miniature version of the entire scene physically in their hands by using the *object inspection* technique. Besides providing an overview, the WIM also visually highlights the part of the scene that is currently visible on the table by enclosing it with a semi-transparent box. Similar to *object clutching*, users can temporarily fixate the semi-transparent box in the physical space by pushing and holding a button. While the box is fixated in real-world space, moving the window will move the miniature map with respect to the box (see Figure 9, middle and right). This directly affects what parts of the scene are displayed on the tabletop.

EXAMPLE APPLICATION SCENARIOS

Tangible Windows provide a functionality that is suitable for a variety of application domains. Examples are manifold: a team of architects and urban designers could reshape an historic factory site on the basis of a virtual 3D model; a new car could be designed and then inspected virtually; doctors could plan a surgery with a virtual 3D representation of a patient's body. The key strengths of Tangible Windows are their generality, their support for parallel activities and collaboration, and their natural way of interaction. In order to demonstrate these benefits, we built two simple example applications – the *Virtual Sandbox* and the *Interior Designer* – that we used as a testbed for studying and improving the interaction with Tangible Windows. We also propose Medical Visualization (MedVis) as one particular domain for the application of Tangible Windows.

Virtual Sandbox The Virtual Sandbox is a static virtual 3D space that can exist above, on, and also below a tabletop surface (see Figure 4). It is filled with several simple 3D objects (triangle-based surface geometry) of different sizes and complexity. These objects can be viewed, selected, moved, rotated, and copied individually with Tangible Windows by using the interaction techniques previously described. It does not impose constraints on the location and orientation of the objects. In this way, the Virtual Sandbox acts as a simple virtual playground that allows for arrangement and manipulation of virtual 3D objects.

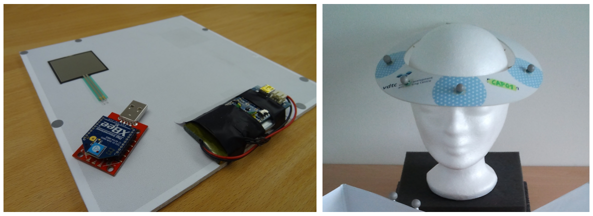

Figure 10. Pressure-sensitive buttons and crown.

Interior Designer Interior design is one of the use cases for our approach. Designers can use Tangible Windows to present different arrangements of furniture to their customers. This may help them to understand the design proposals and facilitate their decision. The Interior Designer scene consists of a large three-dimensional floor plan filled with several pieces of furniture (see Figure 5, left). Compared to the Virtual Sandbox scene, which emphasizes the possibilities of unconstrained three-dimensional interaction, it focuses on a 2D map that extends into the 3rd dimension. Due to its size, not the whole layout can be shown at once. This example thus benefits from the global navigation interaction techniques presented earlier. For example, the bedroom of a virtual apartment in the Interior Designer may be shown on the table display. Should the user decide to examine the kitchen, he or she might opt for using the *scene-in-hand-dragging* to change the viewpoint, because it allows a coarse and quick navigation. The pieces of furniture placed in the scene's rooms can be moved, rotated and copied. Two principle modes are possible on the tabletop: a 2D map of the floor plan, well suited for multiple users, and a 3D view of the rooms on or below the table. Seamless switching between the two modes is supported by lifting and lowering the global scene.

Medical Visualization While the former two examples demonstrate the interaction with Tangible Windows, they do not fully show the potential benefits for collaborative work. One of the domains that we envision to benefit from our approach is Medical Visualization (MedVis). Here, volume data sets often need to be examined by several medical professionals, e.g, for therapy planning (see Figure 4, right). By using multiple displays, we can combine different projections of the same scene. This can be used to provide individual, personalized head-coupled perspectives for each user, displayed on their own peephole window. This may help to present 3D-spatial relations more realistically than normal displays. At the same time, the tabletop can show a general view, e.g., a planar projection or an outline of the patient's body which serves as a frame of reference. Additionally, different visualization techniques can easily be combined. For example, by showing both a direct 3D-volume rendering and a 2D-slice projection (freely defined by the orientation of the Tangible Window), we can support a fast comparison of different views of the same data set using multiple handheld displays. MedVis is also a good use case for the annotation of data, supported by pen input, directly on the Tangible Window.

TECHNICAL SETUP

In the future, Tangible Windows could be implemented as self-contained (active) displays, e.g., by using organic light emitting diode (OLED) technology that provides high-resolution views and high-accuracy multi-touch/pen input. Due to current technical limitations and availability, we chose a passive display approach that uses cardboard as projection material. Besides immediate availability, these passive displays have the advantage of being a straightforward and cost-effective way of integrating and customizing new displays in arbitrary shapes for several users once the system is built.

Principle Setup Inspired by [31], the principle components of our technical setup are: a back-projected tabletop, an infrared (IR) optical tracking system, a ceiling-mounted projector, and several pieces of cardboard serving as mobile displays (Tangible Windows). We extended this setup with the ability to track users' head positions by using tracked crowns.

Tracking of Mobile Displays and Heads Twelve IR cameras (Optitrack FLEX:V100R2) allow for precise determination of position and orientation (6DOF) of mobile displays and crowns at 100 Hz with a tracking error of less than 3 mm. We glued small and unobtrusive IR reflective markers onto all tracked devices. Besides calibration of the overall tracking system, every tracked devices has to be initially registered once. In order to ensure a robust tracking, only four out of six markers needed to be visible for a successful recognition. Unique marker designs ensured a reliable distinction between individual devices.

Mobile Displays Mobile displays were made of circular- and rectangular-shaped pieces of cardboard with edge lengths between 15 and 25 cm. The ceiling-mounted short-throw projector (HD) is used to project arbitrary image content onto them. For this purpose, we wrote an OpenGL-based graphics engine that simulates the physical space above the table surface with each mobile display being represented by a textured polygon. In practice, we achieved an overall projection error of less than 5 mm (measured as the offset of handheld cardboards and the projections onto them).

Touch and Pen Input In order to provide basic touch input on mobile displays, we applied two physical pressure-sensitive buttons (Arduino XBee) to some of them (see Figure 10, left). All touch input currently done in our setup (mainly confirming actions and switching between interaction modes) is realized this way. For digital pen input, we attached Anoto pattern[2] to the tabletop and all mobile displays. Though implemented, we did not use digital pen input in the prototypes.

INITIAL USER EXPERIENCE & DISCUSSION

Although still being in an early stage of implementation, experiences with our system were encouraging. Three computer science students and two members of our institute (22 to 35 years old, 1 female) were invited to test our *Virtual Sandbox* and *Interior Designer* prototypes. After a brief demonstration, they were immediately able to successfully interact with the system. They liked the concept of viewing and manipulating parts of a virtual 3D world by simply grabbing and moving small lightweight paper screens. Users particularly found it easy to explore the global 3D world by looking through a *peephole window*. All five users stated that they were impressed by

[2]Anoto Group AB, http://www.anoto.com

the 3D appearance of virtual objects displayed on the handheld *fish tank windows*. The valuable feedback which we received helped us to identify several new ideas and approaches for further improvements and also revealed limitations of our current implementation.

Limitations

Although users generally appreciated the overall impression of Tangible Windows and in particular the intuitive way of interacting with them, four of the five users complained about unpredictable behavior or a non-responsive system. We identified the following causes for this. First, touch input was preliminary with only two pressure-sensitive buttons on each display that do not provide any haptic feedback. Our experiences suggest that real physical buttons with tactile feedback would have been a much better choice, in particular for techniques that heavily rely on holding a button for a longer time, such as *object clutching*, *object moving*, etc. Second, we implemented all techniques as described in Section Interaction Techniques in a prototypic way, not paying much attention on the interplay between them. This required us to frequently reset the prototype during the tests, which impaired the overall interaction experience. One example for this is the hard-coded mapping of head-coupled perspectives and Tangible Windows. Also, head-coupled perspectives on the tabletop were automatically replaced by a default orthographic perspective as soon as two users (recognized by their crowns) were using the system. However, this did not affect the mobile displays. Fourth, the physical interaction zone was restricted due to our technical setup (single projector, limited tracking volume). This could be mitigated by including additional projectors, an extended tracking setup or using active displays, e.g., the iPad. The latter would also help with the problem of limited buttons and provides a higher display quality. In contrast, passive (projected) displays have several advantages regarding their cost and flexibility. For example, interaction techniques like *object flipping* cannot be easily implemented for active displays without using special hardware. For these cases, alternatives need to be found, e.g., providing a button that triggers *object flipping* or replacing it with *object clutching*.

Precision and Constraints

Users sometimes complained about having problems with adjusting an object's orientation and position. Constraints can be employed to facilitate such object manipulation. Especially the Interior Designer prototype shows how certain application domains may provide inherent constraints. For example, most pieces of furniture are placed on the ground, none are floating in mid-air. Other solutions, e.g., grids, object snapping or alignment guides may also improve the user experience but are beyond the scope of this paper. Another problem arises from a lack of depth perceptibility. As the system does not feature stereoscopic displays, depth clues like projected shadows could be used to support the user, especially when working with parts of the scene floating above the table's surface.

More Permanent Representations

Frequently, users asked for a more permanent representation of virtual content. This is because Tangible Windows are mostly temporary in nature and instantly change their view as soon as they are moved. However, due to their flat shape, Tangible Windows do not lend themselves to be fixed in upright positions. A possible solution would be to provide little support stands that could be used for fasting the displays in arbitrary orientations. Similarly, but for virtual scene parts floating higher above the table surface, tripods could allow for fixation of *fish tank windows* in mid-air. One user proposed fixating and using a *peephole window* similar to a watchmaker's magnifying glass that would provide additional contextual views into the global 3D scene in a more permanent manner, which resembles the Boom Chameleon [35].

Head-coupled Perspectives and Head Input

Head-coupled perspectives were not always perceived as superior to manual perspectives. The best effects were achieved with one eye closed, which is due to the monoscopic display approach. Although users generally liked the fish tank VR-like views, there are situations in which it is preferable to switch off auto-perspectives, e.g., when sharing a view on a mobile display with somebody else. We conclude that the ability for a seamless de- and reactivation of head-coupling is a fundamental requirement of a Tangible Window system. One user suggested to use the head for other interaction purposes, e.g., to fade in text labels or to show more detailed geometry when a user comes closer to a display.

CONCLUSION & FUTURE WORK

With Tangible Windows, we have contributed a novel concept for interacting with virtual 3D information spaces in a tabletop environment that goes beyond conventional multi-touch input. This was achieved by adding a spatial component to the interaction and combining tangible interaction, head tracking, and multi-touch as well as pen input techniques. For this purpose, we built a multi-display environment that features a true unification of physical interaction and output space by using multiple handheld displays providing haptic affordances. They serve as visual proxies or peepholes into a virtual 3D world. Multi-user collaboration and the seamless transition between local and distant control of 3D scenes are especially supported by Tangible Windows. Head input is mainly utilized for auto-perspective effects that help making the interaction more natural and intuitive by approximating what users would expect from real-world experiences. We received promising initial user feedback that makes us confident that Tangible Windows are a valuable tool for many 3D tasks operating above, on, or even below a table surface.

With our contribution of several basic 3D interaction techniques, we have only set first steps into a rich interaction space. We foresee a variety of follow-up work that will extend our techniques to additional collaborative work scenarios and specific application domains, e.g., by addressing the issues discussed in the last section, among them the usage of constraints. While in this work we mainly used Tangible Windows for parallel activities, specific concerns regarding collaboration, such as exchange of data between displays or dynamic mapping of head-coupled perspectives to tangible windows to allow changing their ownership have to be examined. Besides simulating real-world effects with head-coupled perspectives, "unnatural" effects, such as exaggerated zooming or distortion effects (fisheye, etc.), are another promising field of research. Other issues to investigate are accuracy and hand fatigue when working with Tangible Windows over a longer period of time. Finally, technical progress provides new possibilities, e.g., marker-less tracking of users and displays, or affordable volumetric displays.

ACKNOWLEDGEMENTS

This work was funded by the German Ministry of Education and Science (BMBF) project ViERforES-II (01IM1000 2B).

REFERENCES

1. Agrawala, M., Beers, A. C., McDowall, I., Fröhlich, B., Bolas, M., and Hanrahan, P. The Two-user Responsive Workbench: Support for Collaboration through Individual Views of a Shared Space. In *Proc. of SIGGRAPH*, ACM Press (1997), 327–332.
2. Bell, B., Höllerer, T., and Feiner, S. An annotated situation-awareness aid for augmented reality. In *Proc. of UIST*, ACM Press (2002), 213–216.
3. Benko, H., and Feiner, S. Balloon Selection: A Multi-Finger Technique for Accurate Low-Fatigue 3D Selection. In *Proc. of 3DUI* (2007), 79–86.
4. Benko, H., Jota, R., and Wilson, A. MirageTable: Freehand Interaction on a Projected Augmented Reality Tabletop. In *Proc. of CHI*, ACM Press (2012), 199–208.
5. Bier, E. A., Stone, M. C., Pier, K., Buxton, W., and DeRose, T. D. Toolglass and Magic Lenses: The See-through Interface. In *Proc. of SIGGRAPH*, ACM Press (1993), 73–80.
6. Bowman, D. A., Kruijff, E., LaViola, J. J., and Poupyrev, I. *3D User Interfaces: Theory and Practice*. Addison Wesley Longman Publishing Co., Inc., 2004.
7. Brown, L. D., Hua, H., and Gao, C. A Widget Framework for Augmented Interaction in SCAPE. In *Proc. of UIST*, ACM Press (2003), 1–10.
8. Chan, C., Lai, S., and Zhan, Y. Evaluation Of A Head Tracked 3-D Document Visualization. In *Proc. of Human Interface Technologies Conference* (2006).
9. Chan, L.-W., Wu, H.-T., Kao, H.-S., Ko, J.-C., Lin, H.-R., Chen, M. Y., Hsu, J., and Hung, Y.-P. Enabling beyond-surface interactions for interactive surface with an invisible projection. In *Proc. of UIST*, ACM Press (2010), 263–272.
10. de Almeida, R. A., Pillias, C., Pietriga, E., and Cubaud, P. Looking behind bezels: French windows for wall displays. In *Proc. of AVI*, ACM Press (2012), 124–131.
11. Downing, E., Hesselink, L., Ralston, J., and Macfarlane, R. A Three-Color, Solid-State, Three-Dimensional Display. *Science 273*, 5279 (1996), 1185–1189.
12. Favalora, G. Volumetric 3D Displays and Application Infrastructure. *Computer 38*, 8 (2005), 37–44.
13. Fitzmaurice, G., Zhai, S., and Chignell, M. Virtual Reality for Palmtop Computers. *Trans. Inf. Syst. 11*, 3 (1993), 197–218.
14. Francone, J., and Nigay, L. Using the User's Point of View for Interaction on Mobile Devices. In *Actes de la Conference francophone sur l'Interaction Homme-Machine (IHM)*, ACM Press (2011), 25–31.
15. Grossman, T., and Wigdor, D. Going Deeper: A Taxonomy of 3D on the Tabletop. In *Proc. of Tabletop*, IEEE (2007), 137–144.
16. Hancock, M., Carpendale, S., and Cockburn, A. Shallow-Depth 3D Interaction: Design and Evaluation of One-, Two- and Three-Touch Techniques. In *Proc. of CHI*, ACM Press (2007), 1147–1156.
17. Hancock, M., Nacenta, M., Gutwin, C., and Carpendale, S. The Effects of Changing Projection Geometry on the Interpretation of 3D Orientation on Tabletops. In *Proc. of ITS*, ACM Press (2009), 175–182.
18. Hancock, M., ten Cate, T., and Carpendale, S. Sticky Tools: Full 6DOF Force-based Interaction for Multi-touch Tables. In *Proc. of ITS*, ACM Press (2009), 133–140.
19. Hancock, M. S., and Carpendale, S. Supporting Multiple Off-Axis Viewpoints at a Tabletop Display. In *Proc. of Tabletop*, IEEE (2007), 171–178.
20. Harrison, C., and Dey, A. K. Lean and Zoom: Proximity-aware User Interface and Content Magnification. In *Proc. of CHI*, ACM Press (2008), 507–510.
21. Hilliges, O., Izadi, S., Wilson, A. D., Hodges, S., Garcia-Mendoza, A., and Butz, A. Interactions in the air: Adding further depth to interactive tabletops. In *Proc. of UIST*, ACM Press (2009), 139–148.
22. Holman, D., Vertegaal, R., Altosaar, M., Troje, N., and Johns, D. Paper Windows: Interaction Techniques for Digital Paper. In *Proc. of CHI*, ACM Press (2005), 591–599.
23. Izadi, S., Hodges, S., Taylor, S., Rosenfeld, D., Villar, N., Butler, A., and Westhues, J. Going Beyond the Display: A Surface Technology with an Electronically Switchable Diffuser. In *Proc. of UIST*, ACM Press (2008), 269–278.
24. Kjeldsen, R. Head Gestures for Computer Control. In *Proc. of ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, IEEE (2001), 61–68.
25. Krüger, W., Bohn, C., Fröhlich, B., Schüth, H., Strauss, W., and Wesche, G. The Responsive Workbench: A Virtual Work Environment. *Computer 28*, 7 (1995), 42–48.
26. Lee, J. Hacking the Nintendo Wii Remote. *Pervasive Computing 7*, 3 (2008), 39–45.
27. Molyneaux, D., Izadi, S., Kim, D., Hilliges, O., Hodges, S., Cao, X., Butler, A., and Gellersen, H. Interactive environment-aware handheld projectors for pervasive computing spaces. In *Pervasive*, vol. 7319 of *Lecture Notes in Computer Science*, Springer (2012), 197–215.
28. Nacenta, M. A., Sakurai, S., Yamaguchi, T., Miki, Y., Itoh, Y., Kitamura, Y., Subramanian, S., and Gutwin, C. E-conic: A Perspective-aware Interface for Multi-Display Environments. In *Proc. of UIST*, ACM Press (2007), 279–288.
29. Reisman, J. L., Davidson, P. L., and Han, J. Y. A Screen-Space Formulation for 2D and 3D Direct Manipulation. In *Proc. of UIST*, ACM Press (2009), 69–78.
30. Spindler, M., Stellmach, S., and Dachselt, R. PaperLens: Advanced Magic Lens Interaction Above the Tabletop. In *Proc. of ITS*, ACM Press (2009), 77–84.
31. Spindler, M., Tominski, C., Schumann, H., and Dachselt, R. Tangible Views for Information Visualization. In *Proc. of ITS*, ACM Press (2010), 157–166.
32. Stoakley, R., Conway, M. J., and Pausch, R. Virtual Reality on a WIM: Interactive Worlds in Miniature. In *Proc. of CHI*, ACM Press (1995), 265–272.
33. Szalavri, Z., and Gervautz, M. The Personal Interaction Panel - A Two-Handed Interface for Augmented Reality. In *Proc. of EUROGRAPHICS*, ACM Press (1997), 335–346.
34. Tan, D. S., Pausch, R., Stefanucci, J. K., and Proffitt, D. R. Kinesthetic cues aid spatial memory. In *Proc. of CHI Extended Abstracts*, ACM Press (2002), 806–807.
35. Tsang, M., Fitzmaurice, G., Kurtenbach, G., Khan, A., and Buxton, B. Boom Chameleon: Simultaneous Capture of 3D Viewpoint, Voice and Gesture Annotations on a Spatially-aware Display. In *Proc. of UIST*, ACM Press (2002), 111–120.
36. Ullmer, B., and Ishii, H. The metaDESK: models and prototypes for tangible user interfaces. In *Proc. of UIST*, ACM Press (1997), 223–232.
37. Viega, J., Conway, M. J., Williams, G., and Pausch, R. 3D Magic Lenses. In *Proc. of UIST*, ACM Press (1996), 51–58.
38. Ware, C., Arthur, K., and Booth, K. Fish Tank Virtual Reality. In *Proc. of INTERACT and CHI*, ACM Press (1993), 37–42.
39. Ware, C., and Osborne, S. Exploration and Virtual Camera Control in Virtual Three Dimensional Environments. *SIGGRAPH Comput. Graph. 24* (February 1990), 175–183.
40. Wilson, A. D., and Benko, H. Combining multiple depth cameras and projectors for interactions on, above and between surfaces. In *Proc. of UIST*, UIST '10, ACM (New York, NY, USA, 2010), 273–282.
41. Yee, K. Peephole Displays: Pen Interaction on Spatially Aware Handheld Computers. In *Proc. of CHI*, ACM Press (2003), 1–8.