Multi-Point Interactions with Immersive Omnidirectional Visualizations in a Dome

Hrvoje Benko *Andrew D. Wilson*
Microsoft Research
One Microsoft Way, Redmond, WA, USA
{benko, awilson}@microsoft.com

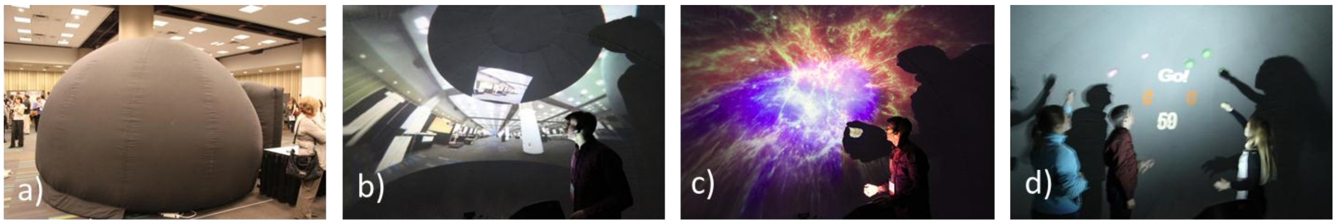

Figure 1: The Pinch-the-Sky Dome experience: a) the inflatable version of our Pinch-the-Sky Dome; b) 360 degree video-conferencing; c) astronomical data from the World Wide Telescope application; and d) a multi-player game.

ABSTRACT

This paper describes an interactive immersive experience using mid-air gestures to interact with a large curved display: a projected dome. Our *Pinch-the-Sky Dome* is an immersive installation where several users can interact simultaneously with omnidirectional data using freehand gestures. The system consists of a single centrally-located omnidirectional projector-camera unit where the projector is able to project an image spanning the entire 360 degrees and a camera is used to track gestures for navigation of the content. We combine speech commands with freehand pinch and clasping gestures and infra-red laser pointers to provide a highly immersive and interactive experience to several users inside the dome, with a very wide field of view for each user. The interactive applications include: 1) the astronomical data exploration, 2) social networking 3D graph visualizations, 3) immersive panoramic images, 4) 360 degree video conferencing, 5) a drawing canvas, and 6) a multi-user interactive game. Finally, we discuss the user reactions and feedback from two demo events where more than 1000 people had the chance to experience our work.

ACM Classification: H5.2 [Information interfaces and presentation]: User Interfaces – Input devices and strategies: Graphical user interfaces.

General terms: Design**,** Human Factors.

Keywords: Freehand interaction, omnidirectional interface, dome, immersive, curved displays, gestures, pinching.

INTRODUCTION

There are increasing amounts of omnidirectional data sources readily available today (e.g., panoramic imagery, astronomical data, earth mapping data, "street view" data); however, the appropriate display options for consuming such data remain scarce due to their inherent immersive and border-less nature.

Omni-directional interfaces, such as CAVE displays [5], room displays [12], cone displays [28], or dome displays [16] offer an interesting solution. While such alternative displays have been extensively explored in research, particularly in virtual reality (e.g., [5, 6, 9, 12, 28]), the interactions within such displays often require the use of expensive tracked devices or intrusive on-body trackers and are often limit control to a single user. Most commercial immersive displays are planetarium domes (e.g., products of Evans and Sutherland[1]). The experiences such planetarium domes are capable of presenting are visually compelling and engaging; however, the people inside are usually passive observers, not able to interact directly with the projected content. In fact, planetariums today are mostly equivalent to domed movie theaters.

We created an immersive dome experience, called *Pinch-the-Sky Dome*, that is both visually engaging and highly interactive (Figure 1). The key differentiation of our work is that the people in the dome interact directly with the experience through simple freehand gestures. Our contribution is not in the design of interaction techniques themselves, as they have been explored in the previous research [8, 18, 19, 20, 21, 30], but instead, in combining them in interesting ways to facilitate a highly engaging, interactive, and novel experience. We leverage the simplicity usually associated with touch-based interfaces and employ gestures

[1] http://www.es.com/products

that act as "touches" in space. Just as a multi-touch interface combine several touches to achieve more complex actions, we combine our gestures in creative ways to offer a richer set of interactions. In designing this experience, we focused on exploring ways to allow the users to interact with immersive content beyond arm's reach through simple hand gestures and speech control, without intrusive trackers often employed in previous virtual reality solutions.

Our solution opens up the possibilities of using such immersive displays for highly interactive tasks such as interactive storytelling, data exploration, multi-player gaming, etc.

This paper describes the implementation of a gesture-based interactive experience with an unusual interactive surface: the dome. First, we describe the user experience inside the installation. Second, we showcase the technology used to facilitate the projection and interactivity in the dome. We then explain the interaction vocabulary we implemented to facilitate data manipulation. Lastly, we discuss user reactions and feedback gathered from several large demonstration events where we exhibited our work.

THE DOME EXPERIENCE

A person enters the dome through the entry gate which is designed to capture outside light (Figure 1a). Inside, the person is immersed in a 360 degree interactive experience. The dome is mostly empty, with a single projector-camera unit located in the middle of the space, leaving plenty of space to accommodate other observers.

The projector uses a very wide angle lens and is capable of projecting an entire hemisphere of content. The projector is angled at 30 degrees from vertical so that the entire projected hemisphere is tilted and more easily observable by the people in the dome. The projector podium also houses a camera, used to sense user interactions around the dome.

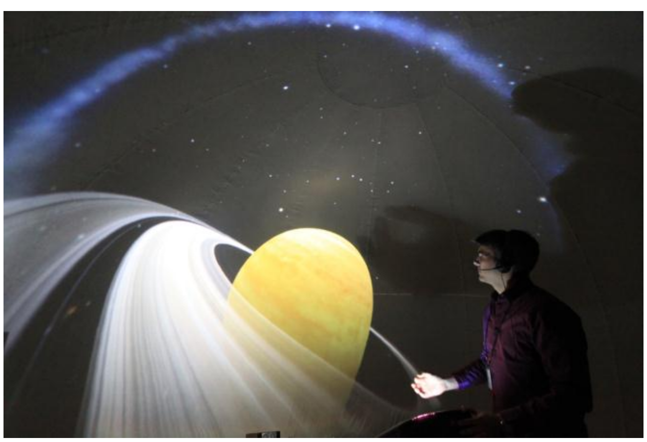

Figure 2. Using a pinch gesture to interact with the projected astronomical content (image courtesy of World Wide Telescope). Note that all of the images in this paper from within the dome were taken with a very wide angle lens. This lens captures more of the projected image, but results in somewhat distorted images.

Currently, the dome provides six omnidirectional applications: 1) astronomical data visualizations, 2) 3D graph visualizations, 3) immersive panoramic images, 4) 360 degree video conferencing, 5) a drawing canvas, and 6) a multi-user interactive game.

We project astronomical imagery from World Wide Telescope[2] [10] and allow the user to explore the sky and the universe by simply moving their hands above the projector. As part of the experience, the users can navigate around the Solar system (Figure 2), visit the far galaxies and the outskirts of the known universe, and observe the incredible imagery of the night sky from the Hubble Space Telescope (Figure 1c).

To manipulate the content one does not need any special devices or tracked gloves. Instead, the user puts their bare hands in front of the projector and makes a *pinch gesture* [18, 30] to move the content around. This simple interaction, illustrated in Figure 2, is the basis of our interaction vocabulary and inspired the name for the overall experience: "Pinch-the-Sky".

Observers can also be virtually transported to several remote destinations by presenting high resolution omnidirectional panoramic images; for example, an Apollo lunar landing site, the lobby of a building (Figure 3), etc. In addition, a live feed from a 360 degree camera located outside the dome can be observed in the dome (Figure 1b). Both the static panoramic images and real-time live video highlight the potential of the dome for omnidirectional video conferencing scenarios with remote participants.

Furthermore, users can explore a custom 3D visualization which presents a social network graph of one of the authors (Figure 4) or use their hand shadows or a laser pointer to draw and scribble on the dome walls.

Lastly, several participants can compete in a multi-user dome game (Figure 5). In this game, two teams compete against the clock to "pop" the bubbles falling from the sky.

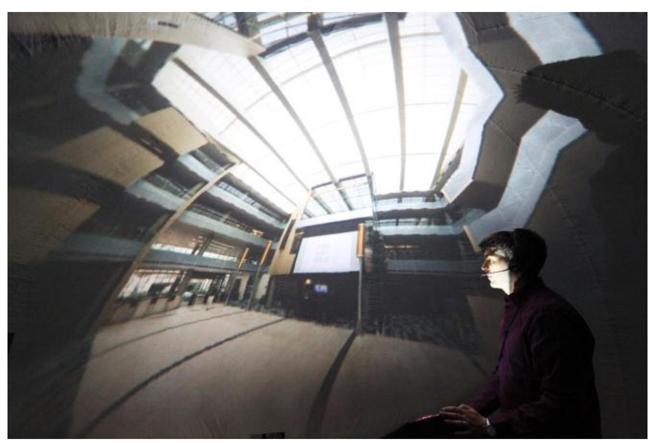

Figure 3: Viewing the omnidirectional panoramic image of a building lobby.

[2] http://www.worldwidetelescope.org

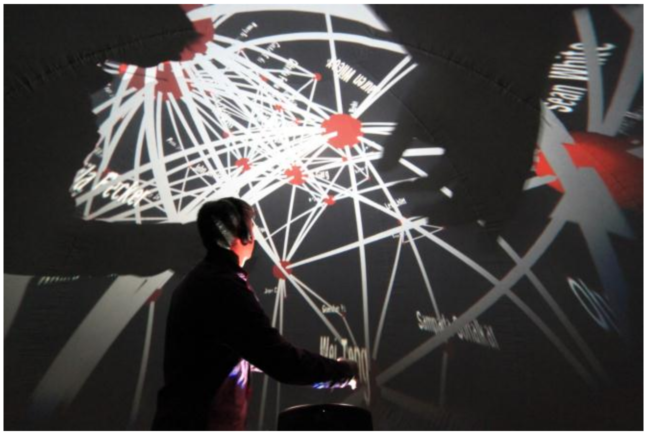

Figure 4: Manipulating a 3D social network graph.

Figure 5: Four players competing in a multi-player game where the object is to pop the falling bubbles using the hand clasping gesture.

Bubbles are popped with a hand clasp gesture (discussed below); the team with more popped bubbles wins.

RELATED WORK

This work builds upon two distinct areas of previous research. The first is the area of immersive dome displays. The second is the area of freehand interactions with virtual content.

Immersive Dome Displays

Much of the research associated with dome visualizations is focused on problems of rendering and projection of content in a highly distorted space such as a dome. Since we are primarily interested in facilitating interactivity in such an environment, a detailed discussion of the rendering and projection aspects is beyond the scope of this paper. We refer the reader to Emmart [6] and Magnor et al. [16] for good overviews of recent research in dome projections, authoring and rendering.

Domes have primarily been used for immersive planetarium visualizations (e.g., [9]) or as immersive 3D scene visualization (e.g., [7]). However, in these works direct interactivity in domes is either not supported, or is supplied through the use of additional input devices (e.g., Fitzmaurice et al. [7] describe the use of handheld tablets to control the 3D experience in the VistaDome). We are not aware of any research in supporting direct gestural interactions to interact with the content of the dome.

Among virtual reality research, the CAVE display [5] is arguably the most widely acknowledged room-sized immersive concept that does not require the user to wear head-worn displays. In CAVE, all sides of the custom-build room are projected with realtime images corresponding to the user's viewpoint to simulate a 3D space. Hua et al. [12] present another effort in enveloping the users with a completely projectable immersive environment. They use head-worn projectors and a room where every surface is covered with retro-reflective material to give multiple users differing perspective views. Our implementation of the omnidirectional projector-camera unit builds upon the work by Benko et al. [3], who presented the first multi-touch sensitive spherical display in which both the projector and the camera were housed in the base of the device.

Freehand Interactions with Virtual Content

Interaction at a distance in immersive virtual environments has been an active research area with most solutions requiring the use of tracked gloves or styli (e.g., [23, 27]). Here we focus on solutions which support freehand interactions without additional trackers.

Kruger et al.'s VIDEOPLACE [14] is probably the earliest example of using freehand gestures to interact with digital content. Interestingly, in that work hands were represented as color-filled outlines. This is fairly analogous to the use of shadows in our work. Since then, researchers have investigated the control of virtual environments through gestures [18, 25] or through a multimodal combination of speech and gestures [15, 13].

Our interactions build on several existing interaction concepts: pinching gestures [8, 11, 18, 30, 31], multimodal speech and gesture interactions [13, 15], as well as laser pointer interactions [19, 20, 21]. Our pinching interactions extend the work of Wilson [30] who proposed using freehand pinching gestures for interacting with a standard desktop in mid-air above the keyboard. Similar pinching interactions have also been demonstrated above an interactive surface [8, 11] or in conjunction with the use of depth sensing cameras [31].

Combining hand gestures with speech commands has been extensively researched in both virtual reality (e.g., [15]) and multimodal input communities (e.g., [13]). We employ this idea with a slight extension where we use a hand gesture as a virtual "push to talk" trigger to activate speech recognition and reduce inadvertent activation.

A large number of computer vision projects have investigated the problem of tracking humans and their actions from video images (e.g., [25, 32]). We refer the reader to [33] for a detailed overview of that space. Interactions in Pinch-the-Sky Dome avoid hard 3D tracking problems by using simple and robust 2D image processing techniques to reason about the spherical space. Our interactions are detected with techniques similar to the standard processing of contacts on a touch-screen. These contacts are transformed to spherical coordinates, thereby avoiding much of the complexities and ambiguities associated with more complex abstractions such as hand or skeletal tracking.

Finally, we take inspiration from the early work of Raskar et al. [24] and Pinhanez et al. [22] where they imagined many interactive surfaces in the environment adapting to users and their context. In particular, Pinhanez et al. [22] used a steerable mirror in front of the projector and camera unit to place a projected interactive image anywhere around the room.

We believe that through the use of omnidirectional projector-camera units similar to the one we used in our dome, we will someday be able facilitate interactions and projections around the room, on every available surface with no more than the user's bare hands. While turning every available surface into a potential projection and interaction surface is a good long-term goal, currently the limited brightness and resolution of today's projectors prevents us from fully realizing this vision without providing an enclosed and relatively dark room, hence our focus here on the dome.

DOME IMPLEMENTATION

Pinch-the-Sky Dome consists of two main parts: the centrally-located projector-camera unit used for display and sensing and the physical dome structure which acts as a display surface.

Wide-Angle Projector-Camera Unit

We placed a custom-made omnidirectional projector-camera unit (Figure 6) in the middle of the dome. This unit is based on the Magic Planet spherical display unit from Global Imagination[3], modified to include an infra-red (IR) camera. The projector-camera unit is 38" high and angled at 30 degrees from vertical so that the entire projected hemisphere is tilted and more easily observable by the people in the dome.

The Magic Planet projector base uses a high-resolution DLP projector (Projection Design F20 sx+, 1400x1050 pixels) and a custom wide-angle lens to project imagery from the bottom of the device onto the dome surface. We removed the spherical display surface of Magic Planet to allow projecting onto the entire hemisphere of the dome surface. The quality of the projected image depends on the size of the dome; the brightness, contrast, and resolution of the projector; and the amount of ambient light that enters the dome. Our 3300 lumens projector is capable of displaying a circular image with diameter of 1050 pixels, or approximately 866,000 pixels.

To enable freehand interactions above the projector in mid-air, we added: an infra-red (IR) sensitive camera, an IR-pass filter for the camera, an IR-cut filter for the projector, an IR illumination ring, and a cold mirror. These components are arranged so that the camera and projector share the same optical axis. The physical layout of these components is illustrated in Figure 7. The modifications are similar to those used in Sphere, a spherical display surface with multi-touch interactions [3].

An IR camera (Firefly MV camera by Point Grey Research[4]) is used for gesture sensing. This camera is able to image the entire area of the projected surface. To ensure that sensing is not affected by currently visible projected data, we perform touch-sensing in the IR portion of the light spectrum, while the projected display contains only light in the visible spectrum. This light spectrum separation approach has previously been demonstrated in many camera-based sensing prototypes (e.g., [3, 17]). A ring of IR LEDs around the lens provides IR light used in sensing.

Because our projector is centrally-located and shares the same optical axis with the camera, we have a lot of flexibility with regards to the environment around the projector.

Figure 6. The projector-camera unit with a wide angle lens and infrared illumination ring. The unit is tilted 30 degrees from the vertical orientation to provide more comfortable viewing in the dome.

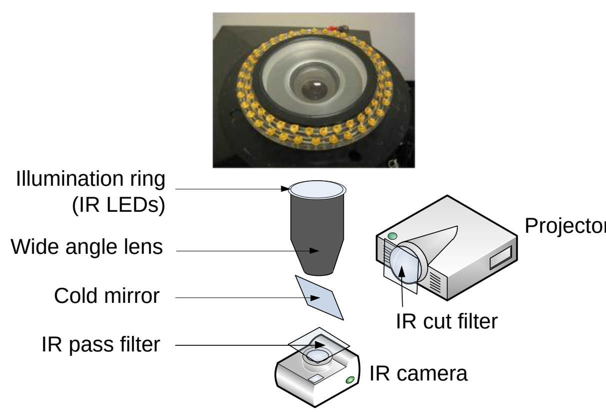

Figure 7. Schematic drawing of our omni-directional projector-camera unit. The detail image shows the wide-angle lens and the IR illumination ring.

[3] http://www.globalimagination.com

[4] http://www.ptgrey.com

For example, our setup can accommodate different sizes of domes and our sensing is always aligned to our projection, without complex calibration routines.

Dome Construction

In our explorations we have employed two dome sizes: a 9ft diameter rigid geodesic dome (Figure 8a) and a 15ft diameter inflatable dome (Figure 8b).

Our 9ft geodesic dome is constructed of cardboard triangles following a 2V design[5], using large binder clips to hold the precisely cut cardboard pieces together. The dome rests on a 30 degree tilted base (matching the tilt of the projector-camera unit), which is built from standard construction lumber and can comfortably accommodate up to 6 observers. We wrapped the base area under the dome with dark fabric to ensure light insulation.

Our second installation uses a 15ft diameter inflatable fabric dome from Go Domes[6]. This implementation can comfortably accommodate up to 12 people and offers a smoother overall display surface. However, this solution is also substantially noisier due to the need to use an air blower to inflate the dome.

Figure 8. Two implementations of our Pinch-the-Sky Dome: a) a 9ft diameter cardboard geodesic dome and b) a 15ft diameter inflatable dome.

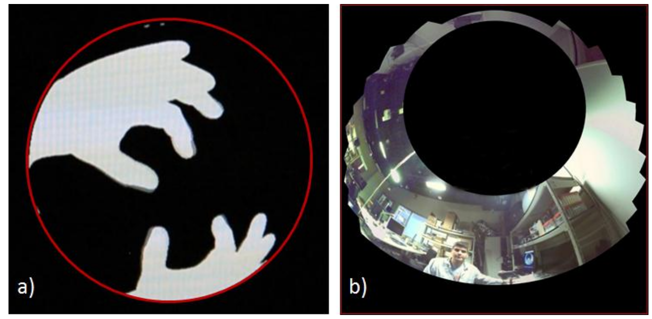

Figure 9: This image shows the distortions present when sensing and projecting in the dome: a) the binary camera image showing user's hands above the projector, and b) the pre-distorted image supplied to our projector which is necessary for correct projection in the dome for the 360 degree video conferencing application. Since our dome is a *tilted dome,* our visualization is uneven in order to appear horizontal in the dome, as seen in Figure 1a.

Projection and Sensing Distortions

The wide-angle lens introduces significant distortions that must be modeled in both sensing and projection. The sensing camera produces a flat radial image that is subsequently mapped onto a spherical surface (Figure 9a). Similarly, the projected imagery must be distorted in order to appear correctly in the dome (Figure 9b).

Many of our visualizations are custom applications written in C# using Microsoft's XNA 3.0 framework and use a custom vertex shader to handle appropriate distortions. In addition, for the astronomical data visualizations, we collaborated with the authors of the World Wide Telescope application to add a custom dome projection mode, which we control from our software.

Once the distortions are appropriately handled, it is trivial to align the camera image and the projected image. This alignment ensures that the actions happening in the sensed image precisely correspond to the content that is being projected. A significant benefit of our approach is that the alignment remains constant regardless of how the environment changes.

Our software runs on a Windows Vista PC with a 2.4 GHz Intel Core2 Quad processor and NVIDIA GeForce 8800 GTS graphics card.

USER INTERACTIONS

The main contribution of our work is in enabling the user to interact with omnidirectional data in the dome using simple freehand gestures above the projector. As with multi-touch touchscreen interactions, which are based on a small set of primitives (i.e., user's touches), we use a small set of mid-air gestures as building blocks for a variety of interactions used in our visualizations.

---

[5] http://www.desertdomes.com

[6] http://www.go-domes.com

Our Pinch-the-Sky Dome interaction vocabulary consists of five different primitives: hand pinch, two hand circle, one hand clasp, speech recognition and interactions with an IR laser pointer.

Before discussing each of these basic interactions in detail, we address a critical problem facing the designers of free-hand gestural interactions which is particularly relevant in the use of an omnidirectional camera.

Gesture Delimiter Problem

The crucial freehand gestural interaction issue is the problem of *gesture delimiters*, i.e., how can the system know when the movement is intended to be a particular gesture or action and not simply a natural human movement through space [1]. More precisely, it is often difficult to precisely know the exact moment the gesture started or ended. For surface interactions, touch contacts provide straightforward delimiters: when the user touches the surface they are engaged/interacting, and lift-off usually signals the end of the action. However, in mid-air, it is not obvious how to disengage from the 3D environment we live in. In our case, the camera's omnidirectional nature makes it even more difficult to step out of the camera frame.

This issue is similar to the classical *Midas touch* problem popularly remembered for the mythical ability of King Midas to turn everything he touched into gold. Little or no difference between a deliberate action and a natural human gesture can result in accidental activations (and in Midas' case, turning his daughter into a gold statue). Therefore, gestures should be designed to avoid accidental activation, allow a reliable means to detect when the interactions begin and end, but remain simple and easy to perform and detect.

Pinch as Mid-Air Touch

We chose the *pinching gesture* [30, 8] as the basic unit of interaction. This can be seen by the camera as two fingers of the hand coming together and making a small hole (Figure 10). The pinching gesture has a beneficial property that the user can feel the exact moment when the pinch begins and ends, making this gesture clearly delimited from other user actions.

This interaction enables the user to reach in front of the projector and literally pinch the content to move it (Figure 11). Furthermore, one can compose pinches in a manner similar to the way multiple touches are composed on a touchscreen. For example, two or more pinches can be used to zoom the content in or out. Throughout our applications we use the combination of one and two pinches to map to different interactions. These are summarized in Table 1.

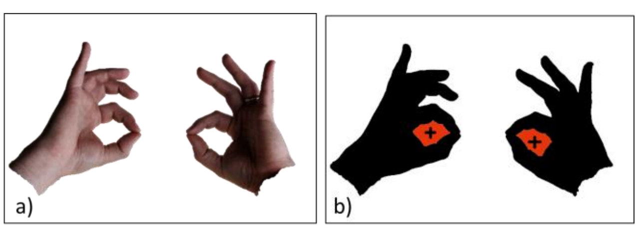

Figure 10: Pinching gestures tracked by our system: a) the image of the user performing two pinches taken from the camera perspective, and b) the binary image showing the areas of detected pinches (highlighted in red). Note: crosshairs mark the points that are reported to the system.

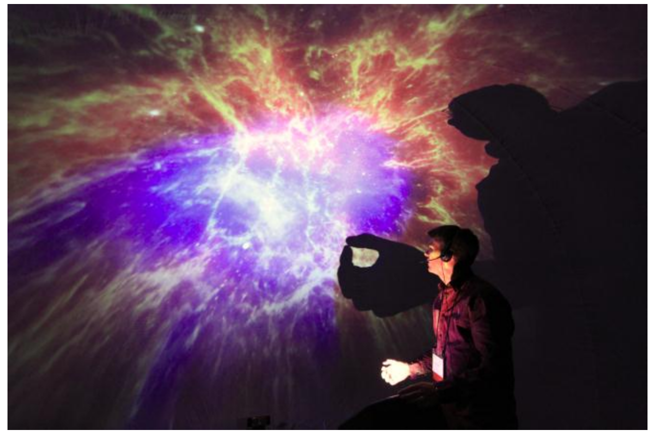

Figure 11: A pinching gesture pans the night sky imagery in World Wide Telescope.

Table 1: Various mappings of one and two pinches facilitate different interactions in our visualizations.

Visualization	1 Pinch	2 Pinches
3D Solar System	Orbit around a planet	Travel towards/away from the planet
Night Sky Images	Pan	Zoom
3D Graph	Rotate	Travel back/forward
360 Panorama/Video	Rotate	Zoom

The similarities between our mid-air pinch interactions and the familiar multi-touch interaction model are probably the most obvious in the 3D graph (Figure 4) and the 360 degree panorama/video (Figure 3). In these applications, the projected content remains directly underneath the users' pinches even while moving the pinch points. This behavior is similar to that of touching an object on a touchscreen to move it. By this token, we might say that our pinching interactions are the analog of touch interactions, but transformed into spherical coordinate space.

Gesture-Invoked Speech Recognition

In Pinch-the-Sky Dome, the navigation between different visualizations is accomplished in a multimodal fashion, where the new visualization is selected by a specific hand gesture in combination with speech input.

In designing this interaction, we wanted to avoid the use of on-screen menus, since they necessarily involve many placement and text orientation choices which can be difficult to resolve in a dome targeted to multiple observers. Speech input provides great flexibility and eliminates the need to select options from an onscreen menu, but an open microphone is often problematic in group scenarios when multiple people might be talking.

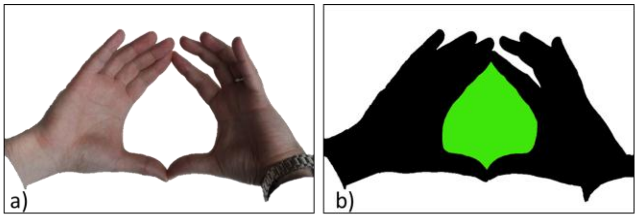

Figure 12: Our *two hand circle* gesture for invoking speech recognition: a) image of the hands, b) processed and binarized image showing the area circumscribed by the user's hands that is recognized as our gesture.

While many virtual environment systems employ a multimodal approach to providing interactivity (e.g., [13, 15]) , we decided to use another freehand gesture to determine when to invoke speech recognition, in order to minimize the number of inadvertent speech recognition errors. This approach provides the user with a gestural "push to talk button". The gesture to invoke speech recognition is a *two hand circle*, which requires the user to put together two hands and make a large circle with their outline (Figure 12). This gesture enables speech recognition and the user can then request to see a new visualization. When the user breaks the gesture (by moving the hands apart), speech recognition is disabled. Speech recognition was implemented using the Microsoft Speech API.

Both the pinching and the two hand circle gestures discussed thus far require the user to be relatively close to the projector. There are two reasons for this requirement. First, the very wide angle of our lens means that the camera does not have sufficient resolution to reliably resolve a hole indicating a pinch at a distance beyond a few feet. Second, the low amount of reflected light at a far distance from our illumination source makes it difficult to reliably detect such gestures. While it is possible to improve our illumination source and thus facilitate the same gestures at a greater distance, we explored two different methods that facilitate such distant interactions even with the current setup.

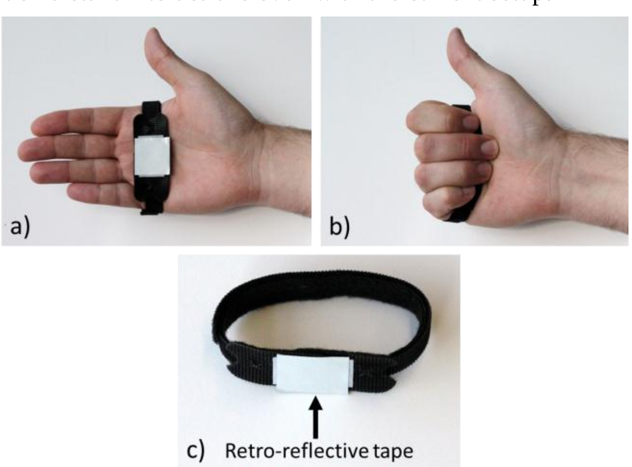

Figure 13: Mid-air hand clasp: a-b) the selection is performed by closing and opening the hand in the same location within 1 sec; c) velcro strap used to hold the retro-reflective tape imaged by the camera.

Hand Clasp as Mid-Air Click

The gestures described thus far facilitate interactions with dome content without requiring the user to wear any tracked object or hold a controller device. Tracked devices can be cumbersome, may be prone to getting lost, or require batteries. Furthermore, in multi-user collaborative scenarios, the need to acquire a tracked device in order to interact with the system can impede the flexibility and the fluidity of interaction. However, we also acknowledge that for many scenarios there are important benefits associated with using tracked physical devices; for example, simplicity and robustness of implementation, reduction of hand movement and fatigue, availability of mode-switching options, differentiation between users, and haptic feedback.

To allow tracking the users' hands at a further distance from the projector, we gave each a simple band with a 1 square inch of retro-reflective tape (Figure 13c). This reflective "token" reflects much more light from our illumination source than the bare hand. These points may be tracked throughout the entire space, from the center of the dome all the way to the dome surface.

In addition to simply tracking the users' hands in space, by quickly closing and opening the hand the user can perform a selection operation (i.e., a mid-air "click"). We termed this gesture a *hand clasp* (Figure 13). This hand clasp gesture is the basic interaction used in our multi-user game where players perform a hand clasp over falling bubbles to pop them (Figure 5).

Our current sensing setup makes it difficult to estimate the distance of the object from the camera. Therefore most of our gestures are best understood in the context of the dome surface and the content projected on it. As future work, it would be interesting to use the brightness of the imaged hands to infer the distance (similar to Hilliges et al. [11]).

IR Laser Pointer Interactions

Another way to interact at a distance in our dome is to use a custom IR laser pointer (5 mW) to point at a specific location on the dome surface (Figure 14). The laser pointer creates an infra-red spot on the surface of the dome which is visible to the camera. While this spot is invisible to the user, the system can project visible light at this location to give the user a visible feedback (i.e., a "cursor"). This point can be tracked and used to manipulate the content in a manner similar to the pinch and hand clasp interactions.

Figure 14: Custom IR laser pointer.

Figure 15: Drawing with the IR laser pointer.

We demonstrate this interaction with a simple drawing application (Figure 15). Our interactions are inspired by the previous research on supporting interactivity with laser pointers [19, 20, 21]; however, we employ an IR laser pointer invisible to human eye in order to be able to track the laser spot, which allows us to smooth the behavior of the subsequently projected cursor or provide control-display gain, as the actual location of the laser spot is not visible to the user.

Using the same logic as for the hand clasp interaction (i.e. briefly depressing the laser pointer button), the user can "click" on a desired item and select it by briefly releasing the button and then pressing it again while pointing the at a same location.

Shadow as a Tool

In our system, because the user is always interacting in front of the projector, shadows on the projected image are inevitable. Such shadows can be considered as both a problem and a unique affordance.

In an environment designed for immersive visualizations, it is preferable to minimize the shadows cast over a presentation as they may reduce the level of immersion and occlude important portions of the visualization.

However, shadows are also very useful. In our multi-user experience, shadows provide a clear indication to other observers as to what gesture is causing the current change in the presentation. In many ways, they act as proxy representations of the user's hands that are directly combined with the projected content. If the user performs a pinch to move an object, the action is very clear to the others in the dome. Furthermore, we often observed that hand shadows are naturally used as a *remote pointer* similar to how one would use a (visible) laser pointer (similar to *Shadow Reaching* [26]). For example, one can use the shadow of their finger as a low-effort means to highlight or point out part of the visualization at a distance (Figure 16). By not requiring the user to actually reach and touch the screen to refer to something, shadows can easily facilitate situations where many things need to be pointed out at various locations around the dome, even at locations out of reach to users (such as the ceiling of the dome).

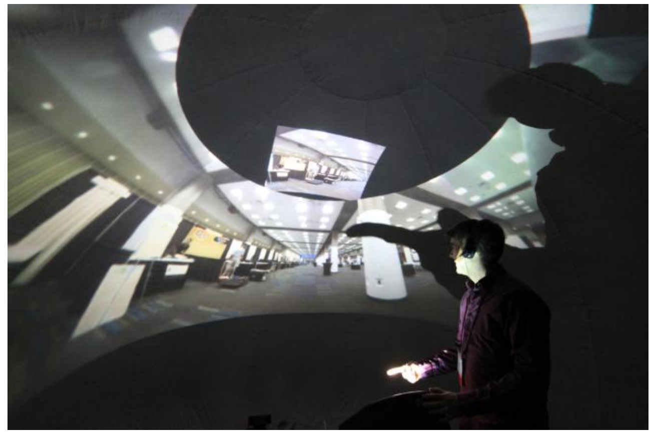

Figure 16: Using a shadow as a remote reference to point at something in the 360 degree video feed.

Lastly, if the user makes a pinching gesture or hand clasp to precisely select an object displayed on the surface of the dome, the shadow provides precise feedback as to where the selection will occur. The shadow effectively enables a three state model of input for mid-air interactions. Buxton [4] noted that most modern interfaces depend on a three-state input model (e.g., a mouse's states are "out-of-range", "tracking" and "dragging"). By seeing their own shadow overlaid on a projected object, the user can precisely know which object they are about to interact with if they make a pinch or a clasp. In essence, the shadow provides the user feedback in a "hover" state for mid-air interactions. This feature is most heavily used in our bubble popping game where each user must position their hand over a projected bubble in order to successfully pop it. If their hand shadow is directly over a bubble, the user can be sure that they are about to engage that particular object.

DISCUSSION AND USER FEEDBACK

We have demonstrated our Pinch-the-Sky Dome on two public occasions. Together, more than 1000 people experienced our demo. The first event was at Microsoft TechFest 2009 (a research showcase event) in which we used the smaller cardboard geodesic dome. The second event was held at the Conference on Human Factors in Computing Systems (ACM SIGCHI 2010) where we used a larger inflatable dome (Figure 17).

The drawing and the multi-person game were implemented after the public demonstrations so most of the user feedback does not directly refer to those scenarios. However, the following discussion refers to all application scenarios.

In general, users commented that our dome provided a compelling immersive experience without much discomfort. As with any immersive experience, some small portion of people experienced *cybersickness*. Cybersickness can be caused by a variety of factors such as the large amount of motion (visual flow), quality of presentation, lag, and field of view issues [29]. In our case, less than 10 people overall (< 1%) left the presentation due to such discomfort.

Figure 17: Pinch-the-Sky Dome was shown as a demo at ACM SIGCHI 2010 and experienced there by more than 500 people.

Users found the notion of pinching to interact in mid-air simple and "magical", but understanding how to perform a pinch was not self-evident. In fact, most users were unable to pinch something on the first try simply because this gesture relies on the camera to observe and track a small hole between one's fingers (Figure 10). Once we explained the basic mechanism behind the pinch detection, users assumed the correct hand orientation and performed pinches without problems. At our prompting, users would often look to the shadow cast by their hand to verify the presence of a hole that can also be imaged by the camera. Similarly, the hand clasp gesture required the users to have their palms facing the camera, which was straightforward and easy to do when explained.

These observations indicate that while we succeeded in creating an easy to detect and easy to perform gestural vocabulary, our gestures were neither self-evident nor easy to learn without some explanation. This was not a serious problem in our demonstrations, as one of the authors always led the presentations, but it would have been problematic if the users were expected to discover this functionality on their own.

Our motivation was to enable multiple observers to easily interact with the content; however, in our experience, most of the presentations were controlled by the single presenter. This might have been due to the short nature of each demo session, where we tried to present as many different applications to the observers as time would allow. Alternatively, it might have been due to the omnidirectional nature of our experiences, as most of our content spanned the entire dome, thus any interaction affected the entire experience. In applications which had more distributed content that could be manipulated independently (e.g., the multi-player game), it was clearly much easier to engage multiple people to interact simultaneously. All of these observations have implications for the creators of dome content, particularly if the interactivity is desired.

CONCLUSIONS

Our Pinch-the-Sky Dome showcases how simple gestural interactions can enhance the immersive experience and how large wide-field-of-view displays provide an immersive perspective of the increasingly available omnidirectional data. To enable the interactions in mid-air, we build upon the concepts from the interactive surface research where simple, clearly delineated actions are composed in a variety of ways to enable a rich set of interactions across applications.

Our work contributes our experience with building, interacting, and presenting the Pinch-the-Sky Dome. We discuss specific implementation details, describe a set of appropriate interactions and their use, as well as contribute the discussion of the use of shadows in such omnidirectional environments.

Ultimately, we would like to be able to place our projector-camera setup in any room and use any surface (walls, tables, couches, etc.) for both projection and interaction, making the idea of on-demand ubiquitous interactive surfaces a reality (similarly to [22, 24]). While we work towards that vision, Pinch-the-Sky Dome offers a glimpse of a highly interactive and immersive experience at your fingertips.

ACKNOWLEDGMENTS

We thank Jonathan Fay of the Microsoft Research World Wide Telescope team and Mike Foody of Global Imagination, Inc..

REFERENCES

1. Benko, H. 2009. Beyond Flat Surface Computing: Challenges of Depth-Aware and Curved Interfaces. In *Proc. ACM MultiMedia '09*. p. 935–944.
2. Benko, H. and Wilson, A. D. 2010. Pinch-the-sky dome: freehand multi-point interactions with immersive omni-directional data. *Extended Abstracts ACM SIGCHI '10.* p. 3045–3050.
3. Benko, H., Wilson, A., and Balakrishnan, R. 2008. Sphere: Multi-Touch Interactions on a Spherical Display. In *Proc. ACM UIST '08.* p. 77–86.
4. Buxton, W. 1990. A three-state model of graphical input. In *Proc. IFIP Tc13 Third International Conference on Human-Computer interaction* (August 27–31, 1990). p. 449–456.
5. Cruz-Neira, C., Sandin, D.J., and DeFanti, T.A. 1993. Surround-screen projection-based virtual reality: The design and implementation of the CAVE. In *Proc. ACM SIGGRAPH '93.* p. 135–142.
6. Emmart, C. 2001. Tools and Techniques for Realtime Dome Production and Education. Computer Graphics for Large Scale Immersive Theaters*. SIGGRAPH '01 Course Notes.*
7. Fitzmaurice, G., Khan, A., Buxton, W., Kurtenbach, G., and Balakrishnan, R. 2003. Sentient data access via a diverse society of devices. *ACM Queue.* p. 53–62.

8. Fukuchi, K., Sato, T., Mamiya, H., and Koike, H. 2010. Pac-pac: pinching gesture recognition for tabletop entertainment system. In *Proc. ACM AVI '10*. p. 267–273.
9. Gaitatzes, A., Papaioannou, G., Christopoulos, D., and Zyba, G. 2006. Media productions for a dome display system. In *Proc. ACM VRST '06*. p. 261–264.
10. Gray, J. and Szalay, A. 2002. The World Wide Telescope: An Archetype for Online Science. *Microsoft Research Technical Report* MSR-TR-2002-75. June '02.
11. Hilliges, O., Izadi, S., Wilson, A. D., Hodges, S., Garcia-Mendoza, A., and Butz, A. 2009. Interactions in the Air: Adding Further Depth to Interactive Tabletops. In *Proc. ACM UIST '09.* p. 139–148.
12. Hua, H., Brown, L. D., and Gao, C. 2004. Scape: Supporting Stereoscopic Collaboration in Augmented and Projective Environments. *IEEE Computer Graphics and Applications.* 24, 1 (Jan. 2004). p. 66–75.
13. Kaiser, E., Olwal, A., McGee, D., Benko, H., Corradini, A., Li, X., Cohen, P., and Feiner, S. 2003. Mutual Disambiguation of 3D Multimodal Interaction in Augmented and Virtual Reality. In *Proc. ICMI '03.* p. 12–19.
14. Krueger, M. W., Gionfriddo, T., and Hinrichsen, K. 1985. VIDEOPLACE—an artificial reality. *SIGCHI Bull.* 16, 4 (Apr. '85). p. 35–40.
15. LaViola, J. 2000. MSVT: A Virtual Reality-Based Multimodal Scientific Visualization Tool. In *Proc. IASTED International Conference on Computer Graphics and Imaging.* p. 1–7.
16. Magnor, M., Sen, P., Kniss, J., Angel, E., and Wenger, S. 2010. Progress in Rendering and Modeling for Digital Planetariums. In *Proc. of EUROGRAPHICS '10.*
17. Matsushita, N. and Rekimoto, J. 1997. HoloWall: Designing a Finger, Hand, Body, and Object Sensitive Wall. In *Proc. ACM UIST '99*. p. 209–210.
18. Mapes, D.P. and Moshell, J.M. 1995. A two-handed interface for object manipulation in virtual environments. *PRESENCE: Teleoperators and Virtual Environments,* 4(4). p. 403–416.
19. Myers, B. A., Bhatnagar, R., Nichols, J., Peck, C. H., Kong, D., Miller, R., and Long, A. C. 2002. Interacting at a distance: measuring the performance of laser pointers and other devices. In *Proc. ACM SIGCHI '02.* p. 33–40.
20. Oh, J.-Y. and Stuerzlinger, W. 2002. Laser Pointers as Collaborative Pointing Devices, In *Proc. Graphics Interface '02.* p. 141–149.
21. Olsen, D. R. and Nielsen, T. 2001. Laser pointer Interaction. In *Proc. ACM SIGCHI '01.* p. 17–22.
22. Pinhanez, C. S. 2001. The Everywhere Displays Projector: A Device to Create Ubiquitous Graphical Interfaces. In *Proc. UBICOMP '01*. p. 315–331.
23. Poupyrev, I., Billinghurst, M., Weghorst, S., and Ichikawa, T. 1996. The go-go interaction technique: non-linear mapping for direct manipulation in VR. In *Proc. ACM UIST '96.* p. 79–80.
24. Raskar, R., Welch, G., Cutts, M., Lake, A., Stesin, L., and Fuchs, H. 1998. The Office of the Future: A Unified Approach to Image-Based Modeling and Spatially Immersive Displays. *In Proc. ACM SIGGRAPH '98.* p. 179–188.
25. Sato, Y., Saito, M. and Koike, H. 2001. Real-time input of 3D pose and gestures of a user's hand and its applications for HCI. In. *Proc. IEEE VR '01*. p. 79–86.
26. Shoemaker, G., Tang, A., and Booth, K. S. 2007. Shadow reaching: a new perspective on interaction for large displays. In *Proc. ACM UIST '07*. p. 53–56.
27. Simon, A. 2005. First-person experience and usability of co-located interaction in a projection-based virtual environment. In *Proc. ACM VRST '05*. p. 23–30.
28. Simon, A. and Göbel, M. 2002. The i-Cone – A Panoramic Display System for Virtual Environments. In *Proc. Pacific Conference on Computer Graphics and Applications.* p. 3–7.
29. Stanney, K.M., Mourant, R.R., and Kennedy, R.S. 1998. Human Factors Issues in Virtual Environments: A Review of the Literature. *Presence 7(4).* p. 327–351.
30. Wilson, A. 2006. Robust Computer Vision-Based Detection of Pinching for One and Two-Handed Gesture Input. In *Proc. ACM UIST '06.* p. 255–258.
31. Wilson, A. 2007. Depth-Sensing Video Cameras for 3D Tangible Tabletop Interaction. In *Proc. IEEE TABLETOP '07.* p. 201–204.
32. Wren, C., Azarbayejani, A., Darrell, T., and Pentland, A. 1997. Pfinder: real-time tracking of the human body. *IEEE Trans. PAMI,* 19 (7). p. 780–785.
33. Yilmaz, A., Javed, O., and Shah, M. 2006. Object tracking: A survey. *ACM Computing Surveys.* 38(4). (Dec. '06), Article #13.