Beyond Flat Surface Computing: Challenges of Depth-Aware and Curved Interfaces

Hrvoje Benko

Microsoft Research
One Microsoft Way
Redmond, WA, USA
+1-425-707-2731

benko@microsoft.com

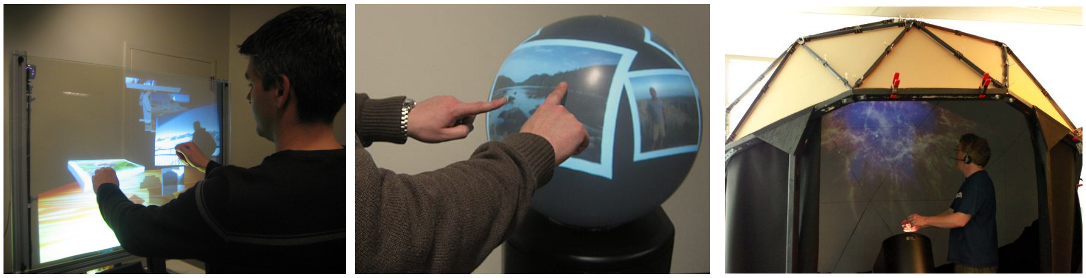

Figure 1: Three non-flat interfaces discussed in this paper that explore the issues of depth-aware or curved interactive surfaces: DepthTouch, Sphere, and Pinch-the-Sky Dome.

ABSTRACT

In the past decade, multi-touch-sensitive interactive surfaces have transitioned from pure research prototypes in the lab, to commercial products with wide-spread adoption. One of the longer term visions of this research follows the idea of ubiquitous computing, where everyday surfaces in our environment are made interactive. However, most of current interfaces remain firmly tied to the traditional flat rectangular displays of the today's computers and while they benefit from the directness and the ease of use, they are often not much more than *touch-enabled* standard desktop interfaces.

In this paper, we argue for explorations that transcend the traditional notion of the flat display, and envision interfaces that are curved, three-dimensional, or that cross the boundary between the digital and physical world. In particular, we present two research directions that explore this idea: (a) exploring the three-dimensional interaction space above the display and (b) enabling gestural and touch interactions on curved devices for novel interaction possibilities. To illustrate both of these, we draw examples from our own work and the work of others, and guide the reader through several case studies that highlight the challenges and benefits of such novel interfaces. The implications on media requirements and collaboration aspects are discussed in detail, and, whenever possible, we highlight promising directions of future research. We believe that the compelling application design for future non-flat user interfaces will greatly depend on exploiting the unique characteristics of the given form factor.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
MM'09, October 19–24, 2009, Beijing, China.
Copyright 2009 ACM 978-1-60558-608-3/09/10...$10.00.

Categories and Subject Descriptors

H5.2. Information interfaces and presentation (e.g., HCI): User Interfaces – Input devices and strategies: Graphical user interfaces.

General Terms

Design, Human Factors.

Keywords

Interactive surfaces, surface computing, spherical displays, multi-touch interactions, depth-sensing cameras, curved interfaces, gestures.

1. INTRODUCTION

Since the pioneering work by Wellner [33], where he imagined many surfaces in our environment becoming interactive and adaptive to the users and their context, research in the area of interactive surfaces has enjoyed stellar growth. Wellner's work was followed by many technological innovations that demonstrated ways of sensing user's touches on the surface: through camera-based tracking of diffuse infra-red illumination (e.g., [23]), frustrated total internal reflection [15], and through capacitive or electrostatic coupling (e.g., [28][11]).

Furthermore, in the past five years, we have seen the emergence of commercial products (e.g., Apple's iPhone[*] and Microsoft Surface[†]) that transitioned multi-touch interactive surfaces from

---

[*] http://www.apple.com/iphone

[†] http://www.microsoft.com/surface

pure research prototypes in the lab, to products with wide adoption and use. Even the upcoming generation of operating systems (i.e., Microsoft Windows 7) will provide native support for multi-touch interactions.

One of the longer term visions of this research follows the idea of ubiquitous computing, where common everyday surfaces in our environment are made interactive (e.g., [24]) and where the user is able to interact with them using multi-touch and whole-hand gestures without specialized gloves or styli (e.g., [39]). However, most of the current interfaces remain firmly tied to the traditional flat rectangular displays of the today's computers and while they benefit from the directness and the ease of use, they are often not much more than *touch-enabled* standard desktop interfaces. In fact, it is hardly surprising that most of the current applications mimic the characteristics of the flat display with two-dimensional (2D or 2.5D) rectilinear user interface elements and concepts, such as rectilinear buttons, windows, scrollbars, etc.

In this paper we make a case for extending the interactive surface vocabulary beyond the 2D interactions that currently dominate our interfaces. We do so by exploring two research directions that push the boundaries of current interactive surfaces: (a) exploring the three-dimensional interaction space above the display and (b) enabling gestural and touch interactions on curved displays. We refer to this space as *non-flat surface computing*.

This paper is organized as follows. First, we review the state of the art in surface computing projects that push the boundary beyond the flat surface interactions. Second, we outline four challenges that researchers and practitioners face when developing compelling experiences with these interfaces. Third, we present three case studies from our own work, which provide some of the initial insights and solutions in this space. First two case studies explore two distinct aspects of non-flat surface computing, while the last one showcases how some of our solution can be tied together to create more impactful holistic experience. Lastly, we offer our vision of what the future might bring if the challenges are resolved.

2. STATE OF THE ART

The research in surface computing has grown substantially in the last five years, and the comprehensive review of all the related work is beyond the scope of this paper. Instead, we restrict our review of the state of the art to projects that push beyond traditional interactive surfaces and explore interactions above the surface and interactions with curved displays.

2.1 Above the Surface Interactions

Most of the interactive touch-sensitive surface systems restrict the user interaction to a 2D plane of the surface and actively disregard the interactions that happen above it. This is usually justified by the system designers' need to reliably detect when the user is in contact with the surface and not accidentally disturb the interface otherwise. Even the interactive surfaces that support interactions with tangible objects, commonly track such objects only when in contact with the 2D plane, leaving the 3D interaction space above the surface largely underutilized. For example, PlayAnywhere prototype [35] allows the user to play a virtual game of chess with a remote opponent. While the user can move real physical chess pieces in front of them, the basic mode of interaction remains two-dimensional.

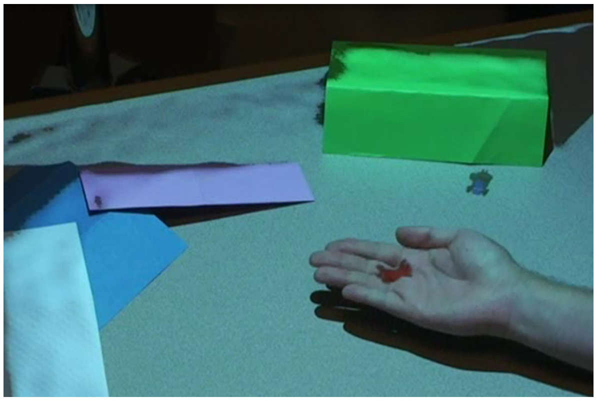

Figure 2: The view of the *Micromotocross* game as seen directly on the tabletop. The user is able to literally reach into the interface, thus altering the virtual terrain and "lifting" a virtual car. (Adapted from Wilson [37])

The interactions in the hover space above the interactive surface have previously been explored within augmented and virtual reality fields with the use of head-tracked displays and tracked gloves, pens or styli, (e.g., [1][10][30]). Such interfaces, demonstrated the range of possibilities when interactive workbenches are augmented with the ability to track user actions above the surface. For example, Starner et al. [30] proposed using 3D reconstruction algorithms from multiple cameras above the tabletop surface to perform simple 3D model acquisition for highly interactive tasks. Their interface enables the user to bring physical objects (props) into the interface and to be able to interact with them to manipulate virtual data. However, most augmented or virtual reality interfaces require the user to wear or hold additional gear making them difficult to setup, initialize, or walk up and use, thus losing the simplicity and the directness that are associated with surface computing interfaces today.

Researchers have also explored using transparent screens to image the actions or documents above the surface. For example, Wilson's *TouchLight* system [34] explored using a holographic screen for touch based interactions, which enabled the system to capture a document or an image through the screen. Izadi et al.'s *SecondLight* project [19] explored using a switchable diffuser screen in combination with rear projectors and a camera to allow for interactions both on and above the surface. While they have not explored freehand interactions, they demonstrated tracking of objects above the screen. Grossman and Wigdor [13] present a useful taxonomy of 3D interfaces on the tabletop and point out areas of promising future work.

So far, only a handful of interactive surface projects have explored freehand 3D interactions without any physical trackers or markers. One of the earliest such projects, *Illuminating Clay* [24], used a laser-range-sensing technology to facilitate manipulations of a morphpable projected surface. The users were able to modify a virtual terrain map, by touching and moving tiny physical particles contained in a sandbox. Probably the best example of terrain modification for interactive purposes is Wilson's *Micromotocross* game [37]. Micromotocross was one of the first interactive surface interfaces that showcased the capabilities of a novel camera device referred to as *depth-sensing camera*, which was used to support interactive modification of the terrain in a car-driving simulation. The user can literally build up the terrain on the tabletop out of whatever physical objects are available (including their hands) and then drive a virtual buggy over such obstacles (Figure 2). The magic of such interfaces lies in the fact that the system does not know anything about the objects and is not trained to track them or recognize them, but simply uses the depth map received by the depth-sensing camera to modify the terrain of the virtual game. The virtual game was then simply projected back onto the tabletop. We further discuss the capabilities of depth-sensing cameras and what interactions they enable as part of our DepthTouch case study in Section 4.1.

2.2 Non-Flat Form Factors

In addition to sensing user's actions above the display, researchers have explored embedding interactive display capabilities into curved and shaped form factors.

Hua et al. experimented with head-worn projected displays and projected their interfaces onto cylindrical surfaces [17]. Popyurev et al. explored a handheld multi-faceted interface concept consisting of 20 displayed faces [26], while Cassinelli and Ishikawa [6] showcased a deformable interactive display where the amount of deformation was used to visualize a different layer in an image. Their *Khronos* display, made of stretchable fabric, was sensitive to significant deformations caused by user's hands.

Holman and Vertegaal recently argued for exploring many existing objects in the environment as potential interactive surfaces [16]. They experimented with using external motion tracking sensors to track interactions with a spherical device, as well as hypothesized what interactions would be enabled if a beverage can was interactive on its surface. Spherical or hemispherical interactive displays have been explored in several interactive projects (e.g., [7][20]); however all such projects used external tracking technologies or handheld controllers in order to interact with the displayed content. There are also several commercially available spherical displays today (e.g., Magic Planet[‡], OmniGlobe[§], and PufferSphere[**]), but none of them offer the touch- or gesture-sensitive interactive capabilities. Our experience with designing a spherical multi-touch sensitive display [2] is discussed in Section 4.2.

In contrast to displays that present data on their curved surfaces, volumetric displays have been used to visualize and interact with 3D data within the display. Grossman and colleagues performed interaction studies on a spherical 3D volumetric display from Actuality Systems, Inc. [14] and found that the two most noticeable interaction difficulties resulted from an inability to: (1) display anything on the volumetric display's surface, and (2) physically reach into the display. To alleviate these problems, they created a set of interactions based on modified ray-casting selection from a distance, and used an external motion tracking system to allow gestural interactions with the 3D data.

There has also been a lot of virtual reality research on multi-faceted immersive displays that surround the user (e.g., CAVE [9]) or planetarium-style immersive displays where the user is located within a hemispherical display (e.g., VisionDome by Elumens [12]). We refer the reader to the work of Bowman et al. [4] as they provide a much deeper discussion of such immersive display technologies that the space permits us here. However, all of the interactions in such environments are constrained to interacting with physical artifacts such as controllers, wands and gloves. Our initial exploration of freehand gestural control of an immersive environment is presented in Section 4.3.

3. CHALLENGES OF NON-FLAT SURFACE COMPUTING

There are many technical challenges in implementing the display and interaction capabilities on non-traditional displays; however, those are specific to the chosen technology, and while important and interesting, they often lack general applicability to a wider research area. While we discuss some specific technical implementation details as part of our case studies in Section 4, we now outline four general challenges that researchers face when trying to create compelling non-flat surface computing interfaces. All challenges discussed in the subsequent sections are open research problems spanning the fields of human-computer interactions, multimedia, computer vision, user interfaces, and virtual and augmented reality.

3.1 Facilitating Direct, Easy, Walk-Up-and-Use Interaction Experience

Much of the appeal of the current touch-sensitive interactive surfaces is due to the directness of such interfaces which do not require the user to wear or hold any additional gear to interact. This *walk-up-and-use* functionally can enable groups of users to interact directly and simultaneously, without needing to take turns or learn complex commands. When extending the surface computing interaction space to the 3rd dimension, whether the interactions happen in the space above the display or by the display itself occupying a volume instead of a plane, it is important to preserve the spontaneous and direct nature of current surface computing interfaces and facilitate as much of the interactions through touch and freehand gesture sensing. Doing so effectively remains challenging: What are the right gestures to use? How to track them without markers or gloves? How to effectively teach such gestures to the user? How to support multiple users? How to provide high precision interaction while keeping the gestures easy and low effort? How to make such interactions seem natural and easy to learn?

Primarily, there is a need to research and design freehand gestures both on the surface and in mid-air. Improvements are needed in gesture tracking, design and learning of gestural languages, as well as design of interfaces that are primarily gesture based rather than mouse and keyboard based. One of the very crucial gestural interaction issues, is the problem of *gesture delimiters,* i.e., how can the system know when the movement is supposed to be a particular gesture or action vs. simply a natural human movement through space. For surface interactions, touch contacts provide the straightforward delimiters: when the user touches the surface they are engaged/interacting, while lift off usually signals the end of the action. However in mid-air, it is not easily possible to disengage from the 3D environment we live in. This issue is similar to the classical *Midas touch* problem. Therefore, gestures need to be designed to avoid accidental activation, but remain simple and easy to perform and detect.

We acknowledge that for many scenarios there are important benefits associated with using tracked physical devices; for example, reduction of hand movement and fatigue, availability of mode-switching buttons, and availability of haptic feedback. However, we feel that there is potentially a large interactivity cost associated with requiring the user to wear or hold a device in order to interact with the system and therefore the application's benefits have to merit imposing such a requirement.

3.2 Facilitating an Ecosystem of Heterogeneous Devices

We do not perceive that non-flat surface computing interfaces will replace the existing computing interfaces. In fact, for many tasks we find standard flat rectangular displays perfectly suitable. However, rather than focusing on a single multi-purpose device, we hope that our workplaces and homes of the future will contain an ecosystem of heterogeneous display devices [12], small and large, flat and curved, each serving a particular purpose. Rather

---

[‡] www.globalimagination.com

[§] www.arcscience.com

[**] www.pufferfishdisplays.com

than the "one size fits all" approach of current desktop computing, having different devices, each well suited to particular tasks, will likely provide a richer and more appropriate "workshop" for information access and manipulation.

This idea, first formulated by Weiser [32], is well familiar to ubiquitous computing researchers and we stipulate that in addition to varying the size, resolution, and portability of the display devices, one should also consider varying their shape, as well as their interactivity and sensing capabilities. Furthermore, we propose that some of the freehand above-the-surface interactions explored in this and related works be used to connect and transition the data between the devices thus creating an interactive "ether" (following the concepts presented by Butz et al. [5] and Rekimoto and Saitoh [27]). For example, it would be interesting to explore the world population data as a chart in a presentation on a vertical screen, then throw it on the spherical display and see it overlaid onto the Earth's globe, and move it to one's handheld device for later retrieval.

Of course, as with any multi-device scenario, this requires the networking and middleware infrastructure to support easy data transition across devices. While demonstrating such ideas as part of a lab prototype is relatively straightforward, taking into account all the real-world issues of permissions, user identification, accessibility, as well as data specification and access remains an open challenge. Here too, having sensors that detect activity in mid air might be beneficial; for example, user-facing cameras could be used to perform facial recognition in addition to gesture tracking, thus authenticating and identifying the users without requiring them to explicitly log into the devices. Furthermore, given dramatic differences between devices, it is important to consider automated ways of picking the most suitable device, or morphing and transforming the data that best suites the presentation on a chosen device.

3.3 Design of Media and Interfaces That Are Compelling From Multiple Directions

Most of the today's media and user interfaces are designed to be viewed and used in one canonical orientation only. This works well for most vertical screens as viewers all share the same *up* direction and are usually able to see the entire screen (albeit with some perspective distortions). However, on horizontal surfaces such as interactive tabletops, the data and interface orientation issues are much more problematic [29]. In fact, it is still an open research question to design a compelling tabletop presentation for multiple people around the table.

However, with non-flat interactive surfaces, this is even further complicated, since each user sees a different view or even a different portion of the display. What does it mean to have a media presentation where different people around a same device get a different view or perspective? See different data? How does one design an interface where not the entire interface is visible at any given time? How does one support multiple users without disturbing one another? Or what are the good awareness cues of actions that happen by other users on the invisible portions of the device?

Similar issues arise when supporting multiple users in the view-dependent interface (i.e., an interface that depends on user's head tracking) as is the case in many above-the-surface interaction prototypes. What are the compelling solutions that do not resort to head-worn glasses?

3.4 Compelling Applications

Lastly, the big challenge is to identify compelling applications that highlight the benefits of such non-traditional displays. While the relative infancy of available research and the low availability of such hardware prototypes make it difficult to discuss useful applications, it is important to start identifying the promising application areas. We believe that the compelling application design for future non-flat user interfaces will greatly depend on exploiting some unique characteristics of the given form factor.

Figure 3. Interacting with *DepthTouch*: user's left hand is touching an object of interest, while his right hand is adjusting the orientation and depth of that object by moving in mid-air above the surface.

In particular, the success of Nintendo's Wii[††] has shown the appeal of activity-based gaming applications, and Microsoft's Project Natal[‡‡] is actively pursuing this direction, by eliminating the controller altogether and making the experience all about hand and body movement. Multi-touch interfaces have also been useful in geospatial map applications and we believe that curved interfaces might provide added benefits for such domains as well. A variety of imaging applications (e.g., medical or geospatial imaging) might benefit from displays that are shaped in appropriate manner to reflect the display content.

While our observations in this paper primarily focus on configurations explored in our prototypes, we envision that the ideas presented here are applicable to a variety of non-flat or curved display form factors that will be available in the future and we hope to inspire interesting application possibilities.

4. CASE STUDIES

We now present three case studies that show our explorations of the non-flat surface computing space.

4.1 DepthTouch

DepthTouch [3] is an interactive system which explored freehand 3D interactions while preserving the "walk-up-and-use" simplicity of a multi-touch surface (Figure 3).

4.1.1 System Implementation

DepthTouch consists of a depth-sensing camera (ZSense depth-sensing camera from 3DV Systems, Ltd. [17]), a transparent vertical display screen (DNP HoloScreen) and a "short-throw" projector (NEC WT610, 1024x768 pixel resolution) (Figure 4). In addition to these components, a desktop PC computer is used for processing the camera data and driving the display.

[††] http://www.wii.com

[‡‡] http://www.xbox.com/projectnatal

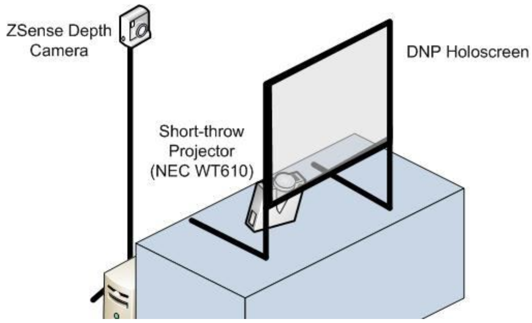

Figure 4. DepthTouch system components.

The enabling technology in DepthTouch is a depth-sensing camera which for every camera pixel, reports not only the color, but also depth value of that pixel. While numerous camera-based interfaces have previously demonstrated ways to influence the virtual world with the shape and gesture of the hand (going back to Kruger et al.'s VIDEOPLACE [21]), depth-sensing cameras present an opportunity to simplify the 3D gesture detection and tracking and thus enable more complex interactions in front of the display.

We acknowledge that other methods of obtaining depth information exist. For example, laser-range scanners have been used in robotics and other fields to acquire accurate depth images, but they are often not fast enough for interactive applications. Correlation-based stereo is another well known approach which suffers from the need of precise calibration, high computational costs, and it often fails on regions with little or no texture.

However, cameras that can directly compute depth information, such as ZSense camera[§§] by 3DV Systems [17], are not susceptible to drawbacks of such related approaches. ZSense camera computes a depth-map image (8bit, 320x240 depth image at 30Hz) by timing the pulsed infra-red light released by the camera and reflected of the objects in front of it: the more light gets returned, the closer the object is at that particular pixel. By measuring the depth of the object or the user directly, one can easily segment it from the background and track it in mid air making depth sensing cameras very suitable for above the screen interactions.

The motivation behind the use of the transparent screen is both practical and fun: it allows for the depth-sensing camera to be placed directly behind the screen and it further enhances the three-dimensionality of the interface as the surface is not just a 2D plane, but rather a window that looks at a 3D virtual scene embedded in a real world. The camera location behind the screen minimizes situations in which one hand occludes the other and allows for tracking of the user's hands by relatively easy segmentation of the range data (Figure 5).

4.1.2 DepthTouch Interactions

The DepthTouch prototype enables the following three types of interactions: (a) perspective view manipulation based on the user's head position, (b) touch-based 2D interactions in the surface plane, and (c) mid-air freehand 3D interactions above the surface.

body hands

Figure 5: Segmenting the user's body using depth values. Top row shows the ZSense depth-sensing camera and the depth image acquired through our display. The bottom row shows the segmented body image and segmented hands in front of the body.

Providing effective feedback for mid-air gestures or 3D visualizations without resorting to head-worn glasses is challenging. While we do not provide a truly stereoscopic view, as that would require that our user to wear some kind of glasses, we provide a correct perspective 3D view to the user based on the position of their head. In addition to the motion parallax obtained by continuous tracking of the head, we enhance the user's depth perception by providing real-time virtual shadows between the objects and the virtual plane at the bottom of the screen.

The screen also behaves in a manner similar to other multi-touch screens. When the user is touching the object on the screen, they can select it and move it in the surface plane by dragging it around.

Lastly, we also allow for fine manipulation of the object rotation and depth by performing mid-air interactions with the second hand, while keeping the object selected with the first hand. The object can be rotated in place by moving the second hand in plane above the surface or brought closer or further in depth by moving the second hand closer or further away from the user's body. We currently do not use the 3D orientation of tracked hand points, but map the object rotation to the simple hand movement in plane.

4.1.3 Research Implications of Depth-Sensing Interactions

There are a number of open research issues facing depth aware interfaces. What interaction metaphors are suitable for this form factor? How does the lack of tangible feedback impact the user's mid-air interactions? What are the "killer" applications? What is the best suited media, or how can different media properties be effectively utilized on such interfaces? So far, the best applications we encountered focused either on 3D physics based interactions or on 3D terrain modifications, which are both very interesting from computer gaming perspective, but might have limited potential with other kinds of applications.

Furthermore, the problem of gesture delimiters (as discussed in Section 3.1) remains a very pertinent one. In DepthTouch, we resolve this by requiring the user to be touching a particular object on the screen with one hand in order to perform depth-based interactions with the other hand. This solution, while adequate,

---

[§§] In June 2009, Microsoft Xbox announced the use of a different depth-sensing camera (code-named Project Natal) for enabling more immersive game play in video games, by allowing the players to control the game through their body movement alone.

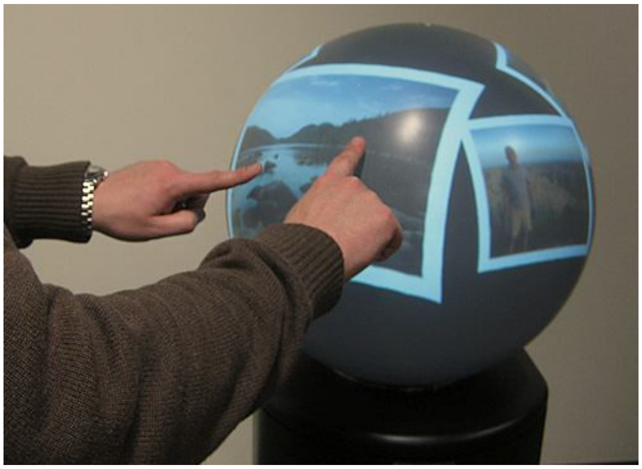

Figure 6: Interacting with a picture on Sphere, a multi-user, multi-touch spherical display prototype built on top of Global Imagination's Magic Planet display.

has a high cost of always requiring a bimanual action. A completely different gestural approach is presented in Section 4.3.

Lastly, we do not believe that all depth-aware interfaces will necessarily all be three-dimensional. In fact, some very compelling depth-based interactions could be mapped to two-dimensional media. However, if 3D is desired, facilitating more than a single user with correct depth cues and potentially providing stereoscopic views is currently not possible without requiring the users to wear head worn displays.

4.2 Sphere

We now focus on our explorations of interactions on curved surfaces, and in particular describe a spherical multi-touch sensitive display called S*phere* [2]. The promise of curved, deformable, or organic-looking displays opens up numerous novel uses and interaction possibilities; however, most of the current applications are ill-suited for such non-traditional surfaces.

In the next several sections, we argue that the design of compelling applications for non-flat user interfaces greatly depends on the designers' ability to overcome inherent interaction challenges and exploit some unique characteristics of such unusual display form factors. We motivate our position with observations and experience with designing interactions and applications for our Sphere prototype.

4.2.1 System Implementation

Our multi-touch-sensitive spherical display, Sphere (Figure 6), is built on a podium version of the commercially available Magic Planet display[***]. The Sphere's surface is an empty plastic ball coated with a diffuse material that serves as a passive curved projector screen. Touch-sensing is performed with an infra-red camera built into the base of the device right next to the projector that is able to image the entire displayable portion of the spherical surface (360 degrees horizontally and approximately 270 degrees vertically) (Figure 7).

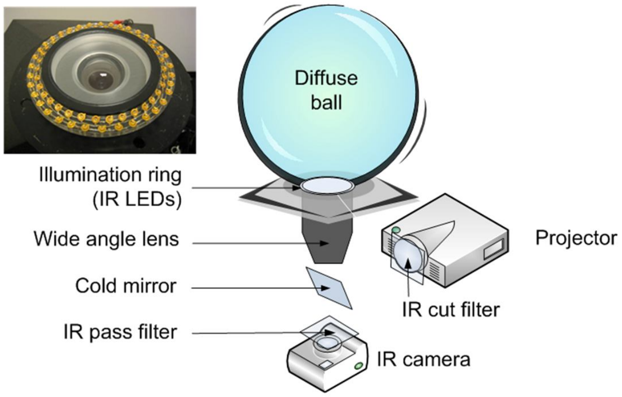

Figure 7: Schematic drawing of Sphere's hardware components that enable multi-touch sensing through the same optical axis as the projection on the spherical surface. The inset picture shows the IR illumination ring consisting of wide-angle LEDs fitted around the wide-angle lens.

The wide-angle lens introduces significant distortions that need to be accounted for in both sensing and projection. The sensing camera is imaging a flat radial image that is subsequently mapped onto a spherical surface to report touch contacts in a 3D Cartesian coordinate system. The projection of data onto the spherical surface requires the use of the inverse mapping, i.e., the data in 3D Cartesian coordinates need to be flattened into a flat radial image for the projector. This means that displayed objects need to be pre-distorted in order to appear undistorted when projected. The reverse mapping is needed for camera sensed image. By performing these distortions in real time, we are able to present the user with highly interactive applications and enable multi-touch tracking of contacts on the surface.

This novel hardware configuration permits the enclosure of both the projection and the sensing mechanism in the base of the device (sharing the same wide angle lens), and also easy 360-degree access for multiple users, with a high degree of interactivity and without any shadowing or occlusion problems. For more details on Sphere's implementation, please refer to [2].

4.2.2 Unique Properties of Spherical Displays

We have developed several prototype Sphere applications such as painting, photo viewer, globe and panoramic visualizations, interactive game concepts, as well as some new multi-touch interactions that facilitate data sharing around the display. We now discuss some unique characteristics of spherical displays and explain how those can be used to design more compelling applications on such unusual form factors.

4.2.2.1 Borderless, but Finite Display

Spherical displays present a difficult design challenge as they require a user interface to be thought of as a continuous surface without borders. Standard flat displays often require an opposite mental model, the content can often stretch beyond the borders of the display, i.e., the display can be thought of as a window into the larger digital world. But for a spherical display, such "off-screen space" usually does not exist; rather, any data moved far enough in one direction will eventually make it full circle around the display. This characteristic can be exploited for interesting effects. For example, we implemented a "potter's wheel" metaphor in our painting application (Figure 8a) where the entire canvas can rotate in place, thus allowing the user to continuously paint all around the display without changing his location.

This characteristic of a borderless, but finite display also create difficulties when application needs to facilitate zooming (e.g., zooming in a global mapping application, such as Virtual Earth). With flat displays, zooming mental model assumes that a lot of content transitions into the off-screen area. Given the lack of off-screen area in a borderless display, standard zooming techniques introduce zippering problems on the opposite side of a display. A

[***] Magic Planet display is made by Global Imagination, Inc.

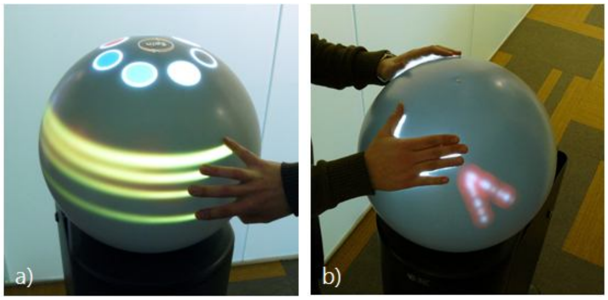

Figure 8: Two interactive applications that exploit the spherical nature of the interface: (a) potter's wheel painting application and (b) spherical pong game where the entire "field" of the game is not visible to any single player.

better metaphor for zooming on a sphere would be to implement a "fish-eye" effect and provide simultaneous focus and context areas thus preserving the benefits of a continuous surface while providing more details in some areas.

4.2.2.2 Non-Visible Hemisphere

Unlike true 3D volumetric displays [14], the diffuse nature of the spherical surface makes it impossible for users to see inside the display and ensures that each user, at any given time, can see at most one half (one hemisphere) of the display. While not being able to see the entire display simultaneously may be a disadvantage for some applications, we believe that in many scenarios this presents a unique benefit. For example, not being able to see all your opponent's actions makes our Sphere pong game (Figure 8b) simultaneously challenging and very engaging.

4.2.2.3 Visible Content Changes with Head Position

Around the spherical interface, even small changes in head position may reveal new content or hide previously visible content. In our pong game, this means that while the user can hope to gain some advantage by shifting their position and peeking at the opponent's actions, they are simultaneously leaving another part of their interface unattended, i.e. vulnerable. Such actions are also socially obvious and participants can rely on standard social cues to ensure "pseudo privacy" for their actions or content.

4.2.2.4 No Master User Position or Orientation

In contrast to horizontal tabletop displays for which orientation of displayed content is often a difficult problem, spherical displays do not have a "master user" position. In many ways, spherical displays offer an egalitarian user experience, with each viewer around the display possessing an equally compelling perspective.

4.2.2.5 Smooth Transitions between Vertical and Horizontal, Near and Far, Shared and Private

A spherical display can be thought of as a continuously varying surface that combines the properties of both vertical and horizontal surfaces. The top of the display can be considered a shared, almost horizontal, flat zone, while the sides of the sphere can be thought of as approximating multiple vertical displays. While this is also true of a cuboid or a cylindrical display, spherical displays offer continuously smooth transitions between all such areas. The top shared portion of the display can be used for content of interest to all participants, such as the circular menu we designed to switch between all our applications (Figure 9). A similar radial interface was explored by Shen et al. [29] on a flat tabletop display. Furthermore, the menu is operated by rotating, rather than directly selecting, which further reinforces the rounded nature of the interface.

Figure 9: Invoking a shared circular menu on top of Sphere using a bimanual invocation gesture.

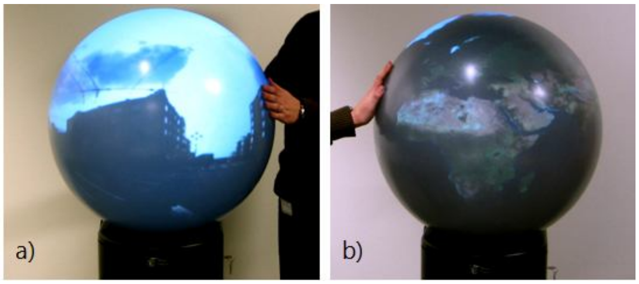

Figure 10: Examples of Sphere omni-directional media visualizations: (a) panoramic walk down Seattle city street; (b) visualization of the Earth as a globe.

4.2.3 Research Implication s of Omni-Directional Interfaces

Omni-directional media – such as cylindrical maps of any spherical object or 360° panoramic images – are well suited for display on Sphere. Examples we explored were a live-stream from an omni-directional video conferencing camera, omni-directional images of a city captured by a camera mounted on a car roof (Figure 10a), and the Earth's surface (Figure 10b).

However, the fact that omni-directional media usually spans the entire display surface presents interesting implications for multi-user, multi-touch collaborative scenarios. Allowing more than one person to touch the data often results in an interaction conflict (e.g., multiple people trying to spin the globe in multiple directions at the same time). While restricting interactions to a single touch does mitigate some of the problems (e.g., the first touch assumes control), such a solution is often confusing to the other users who might not be able to see the action being performed. While this issue should be investigated further, in our current system, users are left to socially mitigate such situations: either taking turns or allowing one person to "drive" the interaction.

All of the interfaces discussed in this paper depend on a projector-camera combination to enable interesting interactions. While flexible eInk or organic LED displays (e.g., [8]) should become available in the future, currently, the major limiting factor for presenting really compelling media is the resolution and brightness constraints of the available projectors.

The lack of resolution is particularly troubling, as projectors have not kept up the resolution when compared to the LCD displays. In fact, the standard projection resolution of 1024x768 pixels is in stark contrast with the 2560x1600 now available on the mainstream LCD panels. While projectors offer us the ability to project onto large surfaces, much of the data described above deserves close inspection where the lack of pixel density becomes very visible and seriously limits the data density that can be presented.

Enabling multi-touch sensing on a spherical surface was done by a powerful combination of a camera and a projector that share the same wide angle lens in the base of our Sphere device. To explore the implications of the scale of the device itself on possible interactions, we have also experimented with drastically different sizes of hemispherical devices ranging from a small handheld device to a large room-sized immersive display. Our next case study presents our research in one of those directions.

4.3 Pinch-the-Sky Dome

Our final example project integrates the research in above-the-surface depth-aware interactions within a large curved display. In this project, we explored a large immersive experience in a prototype called *Pinch-the-Sky Dome*.

4.3.1 System Implementation

Pinch-the-Sky Dome consists of the same projector-camera unit as in the base of the Sphere device, but without the plastic spherical ball on top. By removing the ball, the projector is able to project an image spanning the entire 360 degrees and filling the surrounding space. We have built a tilted geodesic dome (9ft diameter at roughly 30 degree tilt) that surrounds the projector and serves as a large hemispherical projection surface (Figure 1). This setup presents a highly immersive experience to several users inside the dome, with a very wide field of view for each user.

In addition to the omni-directional data sources from the Sphere project, we incorporated the astronomical data from WorldWide Telescope[†††] into our dome and allowed the user to explore the sky and the universe by simply moving their hands above the projector. The main focus of this work is in enabling the user to interact with omni-directional data in the dome using simple freehand gestures above the projector without requiring any special gloves or tracking devices (Figure 11).

4.3.2 Gestural Interactions

The difficulty with allowing the user to use freehand gestures for interacting with the data is the same notion of delimiting actions discussed in Section 3.1. Since our projector-aligned camera is able to image the entire dome that made it difficult to decide when the user is actively engaged with the system and when they are simply watching or interacting with others in the dome. In essence, we wanted to have a simple and reliable way to detect when the interactions begin and end (i.e., the equivalent of a mouse click in a standard user interface).

To enable this, we designed the basic unit of interaction to be a *pinching gesture* (adapted from [36]) which can be seen by the camera as two fingers of the hand coming together and making a little hole (Figure 12). This enabled us to literally "pinch the sky" and move it around to follow the hand, or introduce two or more pinches to zoom in or out similar to more standard multi-touch interactions available on interactive surfaces. One of the significant benefits of choosing this particular gesture is that since the user can have a precise control of when they release the pinch, they can perform rather precise manipulation tasks. For example the user can get the image to a desired state and then simply release it without causing any extra disturbance to the state of the system. This behavior is consistent to the user's expectation of how a computer mouse-based interaction would perform a similar task.

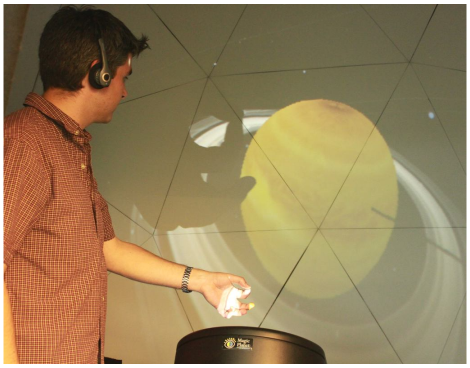

Figure 11: Interacting with freehand gestures in our Pinch-the-Sky Dome.

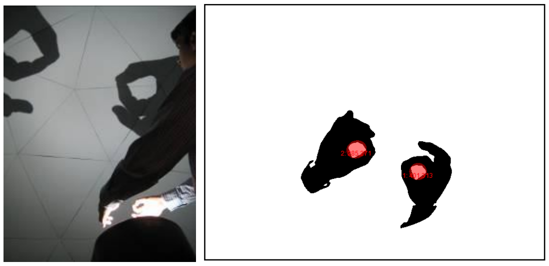

Figure 12: The detection of pinching gestures above the projector (left) in our binarized camera image (right). Red ellipses mark the points where pinching was detected.

Ultimately, by using this projector-camera setup, we would like to enable simply placing it into any room and being able to use any surface (walls, tables, couches, etc.) in the room to both project on and interact on, making the idea of on-demand ubiquitous interactive surfaces a reality. While Pinhanez et al. [24] explored similar ideas while researching interactions with a steerable projector, they were unable to simultaneously project on a variety of surfaces in the environment, which we are able to do. However, currently the low brightness and the low resolution of available projectors prevents us from making this vision into a viable solution today, which is why we have prototyped it in an enclosed immersive dome.

5. VISION OF THE FUTURE

Given that majority of our day-to-day interactions with the physical world requires us to operate in 3D space and handle 3D objects of various shapes, sizes, and forms, it is somewhat surprising that we feel the need to make a case for exploring non-flat interfaces. We understand that there are clear benefits of flat rectangular computer displays, and we do not feel that those will be replaced soon with curved alternatives. However, we also believe that with the improvements in sensing technologies, the interactions will move away from being purely surface-bound and involve people's movement and physical objects above or in front of the display. The directness and ease of use of current multi-touch interactive surfaces already highlight the promise of the *natural user interface* where the only experience the user needs to start interacting is their *real life* experience.

In addition, the success of Nintendo's Wii Remote controller and the recently announced Microsoft's Xbox Project Natal point at the future where standard human movement and interaction with

---

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

physical objects will be a significant way of interacting with digital content.

While many rich sensors are already available (such as the aforementioned depth-sensing cameras) most of the interaction models we currently rely on distilling our actions into point-based actions. For example, while we might use the entire palm of the hand to interact on the interactive surface, the system approximates our action with a single contact point and all of the information about the shape and contour of our hand is basically discarded. This is a direct consequence of the dominant computer mouse interaction model. We believe that, in order to fully utilize the rich interaction space, it is important to facilitate *full-hand* interactions, which incorporate such information as gesture movement, contour, pressure, and depth into the interaction model. Wilson et al. show a promising direction to bring this idea to reality by implementing physics-based interactions on a multi-touch surface [38].

Furthermore, we believe that the most compelling applications for non-flat interactive surfaces will embrace and exploit some of the unique properties such displays embody and we illustrated this in our case study of Sphere. We stipulate that most of the upcoming non-flat, 3D, or even deformable displays will carry a different set of unique properties, and targeting applications that build on top of such characteristics will be critical in the adoption of those interfaces in the future.

6. CONCLUSION

In this paper, we presented an overview of our research in the area of gestural interactions with non-flat surface computing interfaces. We summarized the state of the art, presented four challenges facing researchers in this space, as well as discussed three projects that provide some initial explorations of non-flat surface computing.

We are most interested in exploring the enabling sensing and interaction technologies that will make whole-hand and multi-touch interactions on such surfaces possible. We strongly believe that most displays will soon be bi-directional, i.e., they will display images to the user and also sense the user's actions on their surface and as such provide interesting gestural interaction opportunities. We also believe that in addition to standard rectangular flat displays, the displays of the future will start taking shape and be aware of user's actions above them. However, much work remains to be done to find and develop compelling application for such displays, beyond gaming and high-visibility advertising displays.

We hope that rather than the "one size fits all" approach of current desktop computing, our workplaces and homes of the future will contain an ecosystem of heterogeneous display devices, small and large, flat and curved, each serving a particular purpose, and that interacting with them will require not much more than a touch of a finger or a movement of a hand.

7. ACKNOWLEDGMENTS

We would like to thank Andrew D. Wilson, Jonathan Fay, Ravin Balakrishnan, Steven Feiner, Eyal Ofek, Billy Chen, and Mike Foody. We are also grateful for the generous support from 3DV Systems, Ltd. and Global Imagination, Inc.

8. REFERENCES

[1] Benko, H., Ishak, E.W., Feiner, S. (2005). Cross-Dimensional Gestural Interaction Techniques for Hybrid Immersive Environments. In *Proceedings of IEEE Conference on Virtual Reality (VR)*. p. 209–116.

[2] Benko, H., Wilson, A., and Balakrishnan, R. (2008). Sphere: Multi-Touch Interactions on a Spherical Display. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*. p. 77–86.

[3] Benko, H., Wilson, A. (2009). DepthTouch: Using Depth-Sensing Camera to Enable Freehand Interactions On and Above the Interactive Surface. *Microsoft Research Technical Report MSR-TR-2009-23.* March, 2009.

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

[5] Butz, A., Höllerer, T., Feiner, S., MacIntyre, B., and Beshers, C. (1999). Enveloping users and Computers in a Collaborative 3D Augmented Reality. In *Proceedings International Workshop on Augmented Reality (IWAR).* p. 35–44.

[6] Cassinelli, A. and Ishikawa, M. (2005). Khronos projector. In *ACM SIGGRAPH 2005 Emerging Technologies*. p. 10.

[7] Companje, R., van Dijk, N., Hogenbirk, H. and Mast, D. (2007). Globe4D, Time-Traveling with an Interactive Four-Dimensional Globe. In *Proceedings of ACM Conference on Multimedia (MULTIMEDIA)*. p. 959–960.

[8] Chen, Y., Au, J., Kazlas, P., Ritenour, A., Gates, H. and McCreary, M. (2003). Flexible Active-Matrix Electronic Ink Display. *Nature.* 423. p. 136.

[9] Cruz-Neira,C., Sandin, D.J., DeFanti, T.A., Kenyon, R.V., and Hart, J.C. (1992). The CAVE: Audio Visual Experience Automatic Virtual Environment. *Communications of the ACM*. 35, 6. (June 1992). p. 65–72.

[10] Cutler, L.D., Fröhlich, B., and Hanrahan, P. (1997). Two-handed Direct Manipulation on the Responsive Workbench. In *Proceedings of the ACM Symposium on Interactive 3D Graphics (I3D)*. p. 107–114.

[11] Dietz, P. and Leigh, D. (2001). DiamondTouch: A Multi-User Touch Technology. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*. p. 219–226.

[12] 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.

[13] Grossman, T. and Wigdor, D. (2007). Going Deeper: a Taxonomy of 3D on the Tabletop. In *Proceedings of the IEEE International Workshop on Horizontal Interactive Human-Computer Systems (TABLETOP).* p.137–144,

[14] Grossman, T., Wigdor, D. and Balakrishnan, R. (2004). Multi-Finger Gestural Interaction with 3D Volumetric Displays. In *Proceedings of ACM Symposium on User Interface Software and Technology (UIST)*. p. 61–70.

[15] Han, J. (2005). Low-cost multi-touch sensing through frustrated total internal reflection. In *Proceedings of ACM Symposium on User Interface Software and Technology (UIST)*. p. 115–118.

[16] Holman, D. and Vertegaal, R. (2008). Organic user interfaces: designing computers in any way, shape, or form. *Communications of the ACM* 51, 6 (Jun. 2008), p. 48–55.

[17] Hua, H., Brown, L. D., Gao, C., and Ahuja, N. (2003). A New Collaborative Infrastructure: SCAPE. In *Proceedings of IEEE Virtual Reality (VR).* p. 171–179.

[18] Iddan, G. J. and Yahav, G. (2001). 3D Imaging in the Studio. *SPIE*, vol. 4298. p. 48.

[19] Izadi, S., Hodges, S., Taylor, S., Rosenfeld, D., Villar, N., Butler, A., and Westhues, J. (2008). Going beyond the display: a surface technology with an electronically switchable diffuser. In *Proceedings of ACM Symposium on User interface Software and Technology* (*UIST*). p. 269–278.

[20] Kettner, S., Madden, C. and Ziegler, R. (2004). Direct Rotational Interaction with a Spherical Projection. In *Proceedings of Creativity & Cognition Symposium on Interaction: Systems, Practice and Theory*.

[21] Krueger, M. W., Gionfriddo, T., and Hinrichsen, K. (1985). VIDEOPLACE—an artificial reality. *SIGCHI Bulletin* 16(4) (Apr. 1985). p. 35–40.

[22] Kruger, R., Carpendale, S., Scott, S. and Greenberg, S. (2003). How People Use Orientation on Tables: Comprehension, Coordination and Communication. In *Proceedings of ACM Conference on Supporting Group Work (SIGGROUP)*. p. 369–378.

[23] Matsushita, N. and Rekimoto, J. (1997). HoloWall: Designing a Finger, Hand, Body, and Object Sensitive Wall. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*. p. 209–210.

[24] Pinhanez, C. S. (2001) The Everywhere Displays Projector: A Device to Create Ubiquitous Graphical Interfaces. In *Proceedings of the International Conference on Ubiquitous Computing (UBICOMP)*. p. 315–331.

[25] Piper, B., Ratti, C., and Ishii, H. (2002) Illuminating Clay: A 3-D Tangible Interface for Landscape Analysis. *In Proceedings of Conference on Human Factors in Computing Systems (CHI)*. p. 355–362.

[26] Poupyrev, I., Newton-Dunn, H., Bau, O. (2006). D20: Interaction with Multifaceted Display Devices. *In Proceedings of Conference on Human Factors in Computing Systems (CHI)*. Extended Abstracts. p. 1241–1246.

[27] Rekimoto, J. and Saitoh, M. (1999). Augmented Surfaces: A Spatially Continuous Work Space for Hybrid Computing Environments. In *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI)*. p. 378–385.

[28] Rekimoto, J. (2002). SmartSkin: An Infrastructure for Free-hand Manipulation on Interactive Surfaces. In *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI)*. p. 113–120.

[29] Shen, C., Vernier, F., Forlines, C. and Ringel, M. (2004). DiamondSpin: An Extensible Toolkit for Around-the-Table Interaction. In *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI)*. p. 167–174.

[30] Starner, T., Leibe, B., Minnen, D., Westeyn, T., Hurst, A., and Weeks, J. (2003). The Perceptive Workbench: Computer-Vision-Based Gesture Tracking, Object Tracking, and 3D Reconstruction for Augmented Desks," *Machine Vision and Applications*, vol. 14, p. 51–71.

[31] Taylor, B. T. and Bove, V. M. (2009). Graspables: grasp-recognition as a user interface. In *Proceedings of Conference on Human Factors in Computing Systems (CHI)*. p. 917–926.

[32] Weiser, M. (1991). The computer for the 21st century. *Scientific American.* 265, 3. December 11, 1991. p. 94–104.

[33] Wellner, P. (1993). Interacting with paper on the DigitalDesk. *Communications of the ACM. 36, 7* (Jul. 1993). p. 87–96.

[34] Wilson, A. (2004). TouchLight: An Imaging Touch Screen and Display for Gesture-Based Interaction. In *Proceedings of the Conference on Multimodal Interfaces (ICMI)*. p. 69–76.

[35] Wilson, A. (2005). PlayAnywhere: A Compact Tabletop Computer Vision System. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*. p. 83–92.

[36] Wilson, A. (2006) Robust Computer Vision-Based Detection of Pinching for One and Two-Handed Gesture Input. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*. p. 255–258.

[37] Wilson, A. (2007) Depth-Sensing Video Cameras for 3D Tangible Tabletop Interaction. In *Proceedings of the IEEE International Workshop on Horizontal Interactive Human-Computer Systems (TABLETOP)*. p. 201–204.

[38] Wilson, A. D., Izadi, S., Hilliges, O., Garcia-Mendoza, A., and Kirk, D. 2008. Bringing physics to the surface. In *Proceedings of ACM Symposium on User interface Software and Technology (UIST)*. p. 67–76.

[39] Wu, M. and Balakrishnan, R. (2003). Multi-Finger and Whole Hand Gestural Interaction Techniques for Multi-User Tabletop Displays. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*. p. 193–202.