Towards Spatially Aware Tangible Displays for the Masses

Martin Spindler
University of Magdeburg, Germany
spindler@ovgu.de

Wolfgang Büschel
University of Magdeburg, Germany
bueschel@st.ovgu.de

Raimund Dachselt
University of Dresden, Germany
dachselt@acm.org

ABSTRACT

Spatially aware tangible displays are a promising approach for extending the interaction space of conventional tabletops from the 2D surface to the 3D physical space above it. This is achieved by utilizing the position and orientation of lightweight tracked displays, which allows for a natural and intuitive way to interact with complex information spaces. Technical solutions for such systems already exist in research labs. However, they usually are expensive, complex, and difficult to maintain and thus are inappropriate for a broad audience. With our work, we want to address this issue. In a long term, we envision a low-cost tangible display system that supports both active displays (e.g., the iPad) and passive displays (e.g., paper screens and everyday objects such as a mug). In this paper, we take the initial step into this direction by presenting a proof of concept system that supports active displays. Along with the application of a consumer depth camera (Kinect), this reduces the costs by more than an order of magnitude.

Categories and Subject Descriptors

H.5.2 [**User Interfaces**]: Input devices and strategies.

General Terms

Design, Human Factors.

Keywords

Tangible Magic Lenses, PaperLens, low-cost system, spatial interaction, Kinect.

1. INTRODUCTION

With the recent advances in commercially available high-performance tablets and smartphones, a new interface paradigm (post-PC) has quickly changed the way we interact with computers. The most prominent feature of such post-PC devices is the absence of any additional input hardware, such as keyboard and mouse. Instead, they are optimized for display-space. As we all know, this works because we can directly multi-touch the visual elements *on* the display with our fingers. This makes it possible to design devices that are ultra-portable, so we can take them around with us as our constant companions – be it for daily work or for leisure. As recent trends show, combining mobile displays with each other or even with larger stationary displays, such as HD-TVs or tabletops, offers exciting new possibilities not only in terms of an enlarged presentation space but also as an increased interaction space.

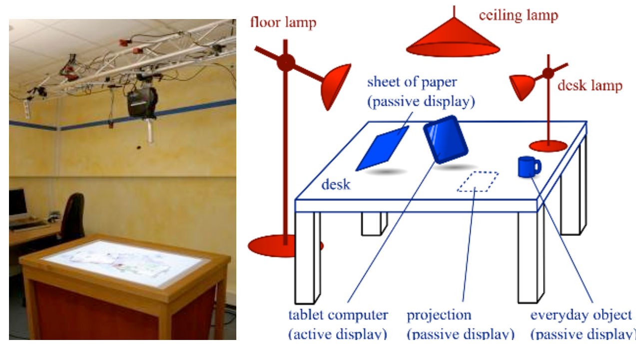

Figure 1. The technical setup of PaperLens [11] is not suitable for the average living room (left). We therefore propose an enhanced concept (right) that is easier to setup, less expensive, robust, modular, and unobtrusive.

Building and studying such multi-display environments is the subject of current research in modern HCI labs, where researchers can experiment with technically complex and costly hardware installations that are usually not suitable for the average living room. Often, a large interactive tabletop or wall-mounted display is central to these installations. They serve as a global display that can be shared by multiple users. Besides investigating techniques for interacting *on* a tabletop, a recent research goal is to extend the interaction space to the physical space above its surface.

Our PaperLens project [11] is such a system. From a technical point of view, it provides a rather complex solution for projecting digital imagery onto lightweight handheld paper-based projection screens (passive displays) that are tracked in 3D space with six degrees of freedom (6DOF). This requires an expensive tracking system, consisting of at least six Optitrack FLEX:V100R2 cameras and a short-throw projector that are attached to the ceiling. Together with a self-tailored interactive tabletop this sums up to a price of over $20.000. This is aggravated by the complicated setup (see Figure 1, left).

In terms of interaction, PaperLens utilizes the concept of spatially aware tangible displays (Tangible Magic Lenses) that users can interact with by grabbing and moving them around in 3D space (spatial input). In this way, users can explore large information spaces in a very natural and intuitive way. For example, by lifting and lowering a tangible display they can zoom in/out a Gigapixel image that is displayed on the tabletop. Our experiences show that spatial input is another powerful input channel that integrates particularly well within multi-display environments. This is also supported by many demo sessions and user studies, e.g., [13], where we received very positive user feedback, not only from average users or children, but also from domain experts (e.g., biologists or radiologist). In this process, we often have been asked when and how these techniques could be available to them.

In this paper, we address this question by proposing a low-cost tangible display system that is easy to setup, robust, affordable, and unobtrusive (see Figure 1, right). We do this in the hope that it will bring spatial interaction a step further in becoming widely available beyond research labs. The remainder of this paper is organized as following. First, we briefly review related work and outline typical application domains. We then assess different options regarding display strategies that we use as a basis for our technical concept.

2. RELATED WORK

The approach presented in this paper continues our research into Tangible Magic Lenses [11], [12]. The underlying concept of Tangible Magic Lenses used in tabletop environments has been proposed by Ullmer and Ishii [14]. Passive displays are presented in PaperWindows [5] by Holman et al. There, a traditional WIMP interface is projected on paper that is spatially tracked. Other prototypes include Chameleon by Fitzmaurice [3], which allows the exploration of 3D-situated information spaces through a spatially aware palmtop computer. Yee's Peephole Display [16] combines the navigation in two-dimensional virtual workspaces with digital pen input.

Various approaches for tracking objects have been used in the past. Visual markers were used in [1]. Zhang et al. [17] use a Hough transform based approach to visually track a panel without markers. In [7], a pattern is projected onto tangible displays. These are equipped with light sensors connected to a microcontroller that computes the position from the detected pattern. IR tracking is, for instance, used in [5], [11], and [12].

Among the first works to utilize a depth camera for touch detection is [15] by Wilson, who uses a Kinect sensor to capture raw depth images and extracts touch events by thresholding operations based on a known model, e.g. the distance to a flat surface. In [4], Harrison, Benko and Wilson propose a wearable system for multitouch interaction, based on a depth-camera and a projector. Finger tips are detected using a gradient model. Surface reconstruction from Kinect depth data, although for different purposes, was done by Clark et al. [2], to provide real-time, user-configurable environments for an AR racing game. Extensive real-time mapping of scenes and simultaneous tracking of a single depth camera is presented by Newcombe et al. in their KinectFusion system [10].

3. APPLICATION DOMAINS

The combination of stationary displays (tabletops or wall-mounted) and mobile tangible displays provides several benefits. As a multi-display environment, it enables the simultaneous use of global and local views and supports collaboration. By employing active displays, e.g., tablets, users can also take their data with them, seamlessly alternating between mobile usage and a fixed interactive space. A multitude of application domains in fields like science or education can be conceived. In the following, we present two of them as an example.

3.1 Exploration of Scientific 3D Datasets

One particular field of application for Tangible Magic Lenses is the exploration of large three-dimensional data. Besides geological or biological data, prime examples are medical volume data sets acquired from MRI or CT.

In a collaborative interactive space consisting of a large tabletop computer and multiple tangible mobile displays, such data sets can be understood as residing in the space above the tabletop. The tabletop acts as a global view, providing reference for the users. For example, it can show an outline of the patient or a specific slice of the data set. The mobile displays provide local, personal views. When a user moves a display through the interaction volume, arbitrary, user-defined cutting planes can be computed in real-time and displayed [11]. This form of direct interaction allows for fast and flexible exploration of the whole dataset or specific structures within.

There are several use cases, which may benefit from such a system. Regularly, physicians have to collaborate for diagnostic or therapy planning purposes, for instance in tumor board meetings. By using tangible mobile displays they can access the data simultaneously, with each of them being provided with their own view. According to their specific fields of expertise, different abstraction levels or visualization parameters can be employed. Also, the experts can particularly benefit from the use of active displays. High-quality screens, like those of the latest iPads, provide the necessary resolution to review even small details of patient data. The possibility to leave the interactive space and still access the data allows for greater flexibility and to continue the work immediately using the normal interaction capabilities of tablet computers.

A similar use case would be doctor-patient consultations. Here, mobile displays can help to present the diagnosis and intended therapy in a way that improves the patient's understanding. Such setups can also be used for educational and training purposes. Students can learn the location, structure and appearance of organs within the human body in a more flexible and interactive manner than with textbooks, complementing established learning resources. A mobile display system with its collaborative capabilities also lends itself to be used by study groups.

3.2 Information Visualization

In many fields of scientific research, large amounts of data have to be visualized and examined. Such complex datasets usually cannot be presented in a single image without the risk of cluttering. This can be mitigated by filtering or presenting multiple views on the data. Tangible mobile displays can support this by providing both additional display space and a new means for interaction, combined in one tangible object.

By utilizing the space above a tabletop, mobile displays can make use of six additional degrees of freedom. These are the location and rotation of the tangible with respect to the interaction space. Also, higher-level gestures (e.g. flipping, shaking) can be defined and used for interaction. For instance, the relative height of the display may encode the zoom factor or, for a semantic zoom, a specific level of abstraction. Other examples include changing the parameters of a fisheye lens by rotating the display. Also, touch interaction directly on the tangible can be employed. A more exhaustive overview of the interaction vocabulary of mobile displays in information visualization can be found in [12].

Tangible mobile displays support the focus and context concept. The fixed tabletop may serve as a reference, showing the contextual background in a main view. The mobile displays on the other hand provide additional physical display space that can be used to show local views into the information space. The support for multiple independent tangibles also makes them a suitable tool for collaboration between users or the visual comparison of portions of the data. As described in the other example application domain, users can carry active displays around, taking their data out of the fixed interactive space and allowing them to continue their work elsewhere.

4. PASSIVE AND ACTIVE DISPLAYS

We distinguish between two principle types of spatially aware tangible displays: *passive* and *active displays*. In the following, we will review important properties of both types and discuss how they impact a tangible display system in terms of interaction and technical realization.

4.1 Passive Displays

Passive displays are (non-instrumented) lightweight mobile projection mediums that are made of paper, cardboard, acrylic glass, porcelain, cloth, etc. This also includes everyday objects, such as mugs, playing cards, or the surface of a table. Their display functionality is realized by projecting the relevant information onto their surface. One advantage of passive displays is their flexibility in terms of form factors. They can be made very thin and lightweight (e.g., by using cardboard), do not feature annoying display frame borders, can show image content on their front and back side (e.g., useful for flipping), allow for arbitrary shapes (e.g., discs), can be extended into the third dimension (e.g., as cylinders or cubes), are inexpensive, and usually they are easy to reproduce.

As a downside, passive displays exhibit only a rather limited mobility. This is because they work only within technically complex environments that are usually stationary. These installations are necessary to precisely determine the position and orientation of passive displays in 3D space and to provide the infrastructure for projecting images onto them. Often passive displays also suffer from poor image quality in terms of resolution and noticeable shifts between object and projection space that are caused by imprecise tracking. Also, curved surfaces or materials with poor reflective properties may limit the projection quality. Beyond that, occlusion (i.e., shadows) can be a problem, e.g., for passive displays positioned over each other.

In terms of natural interaction, passive displays have a huge potential. For example, they allow for changing their shape, e.g. by bending and folding a sheet of paper, they support the interaction *with* them, e.g., by lifting/lowering them, and users can interact *on* their surface, e.g., by pen and finger input. These properties make them suitable for techniques aiming at interacting with everyday objects that are augmented with digital content [1]. For this purpose, a tangible display system must solve three fundamental technical challenges: tracking of objects, projecting digital content onto them, and providing a (gestural) user interface for these objects, e.g., by using finger or pen input.

4.2 Active Displays

Using active displays, e.g., smart phones and tablets, can solve many disadvantages of passive displays. They feature high quality displays (e.g., the Retina display of the 2012's iPad) and thus do not require complicated projector setups. This also implies that device tracking is solely used for spatial interaction and therefore can be less accurate. This is an important benefit and allows for the application of less obtrusive tracking technology, e.g., marker-less approaches. Another advantage of many active displays is that they provide precise multi-touch capabilities out of the box. Beyond that, they are often instrumented with a variety of useful sensors, e.g., accelerometers, near field communication (NFC), and compasses, that add further degrees of freedom to the interaction. In this way, active displays address two technical challenges of a tangible display system. They provide a built-in display solution and a multi-touch interface.

Despite all these advantages, active displays are less flexible than passive displays in terms of form factors. They usually are heavier and thicker, have noticeable display frames, are less variable in shape, and support only a front display. Although with technical progress this might change in the future, e.g., by using OLEDs, a seamless integration of everyday objects and the digital world will hardly be possible with active displays only. We therefore believe that a fully functional tangible display system should support both active and passive displays.

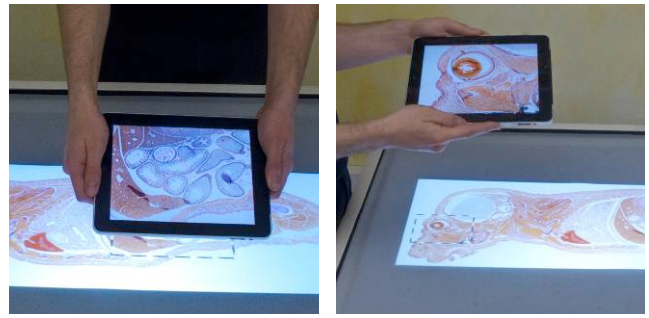

Figure 2. The iPad is used to explore a high-resolution image of a cut through a rat. Lowering and lifting the device controls the zoom factor, whereas horizontal movements allow for panning. Due to the marker less tracking provided by the Kinect, no markers need to be attached to the iPad.

5. TECHNICAL CONCEPT

As our goal is to make tangible displays widely available beyond research laboratories, a tangible display system should be affordable, flexible, extendable, robust, and easy to setup and maintain. We therefore envision a setup that extends the concept of LuminAR Bulb [8]. LuminAR Bulb combines a Pico-projector and a camera in a single device with a compact form factor. It can be screwed into standard light sockets everywhere. These bulbs do not need to provide all functionality, but rather can be specialized to a specific task. A Tracking Bulb, for example, could solely address the tracking of active displays. In this way, using multiple bulbs that wirelessly communicate with each other can expand a workspace that better fits the user's purposes (see Figure 1, right).

While developing such a system is the long-term goal of our research, in this work we address the short-term goal of building a prototypic (but affordable) technical solution for *active* displays, in our case the iPad. For this purpose, we build on PaperLens [11] that is a fully functional passive tangible display system. We extend this system with the capability to stream spatial information between iOS devices in real-time and use this as input for the interaction with them. In terms of tracking, we replace the complex marker-based approach (see Figure 1, left) with a marker-less one that only uses affordable consumer hardware: the Mircosoft Kinect. The use of one iPad and a single Kinect enables us to build a simple tangible active display system for less than $900 (not including the tabletop display).

With the advent of inexpensive depth-sensing cameras like the Microsoft Kinect, many technically demanding computer vision tasks have been made affordable. Such tasks include motion capturing, simultaneous localization and mapping (SLAM), and 3D scanning. In particular, the Kinect employs a marker less 6DOF tracking, which supports our design goal of concealing as much of the technical aspects as possible from the user, e.g., there is no need to glue markers on the iPad.

Although the tracking of objects is feasible with only one depth-sensing camera, several disadvantages can limit the usefulness of such a solution. This includes occlusion (e.g., a user or other displays occlude a tracked display) and unfavorable viewing angles that let the perceived appearance of the tracked display collapse to a line in the worst case. By using two or more sensors, these problems cannot only be mitigated, but they also allow for an increased tracking stability as well as the coverage of a larger interactive space. Unfortunately, sensors operating with the structured light approach like it is done with the Kinect are prone to increased noise when the patterns of two or more Kinects are projected onto the same surface. This eventually can lead to a complete failure of tracking and limits the use of multiple sensors to setups with little or no overlapping of the covered area.

For our prototype, we define a global coordinate system that has its origin in the middle of the tabletop with the Z-axis pointing up. Although we experimented with two Kinects that both provide their own local coordinate system, due to the reasons described above, we decided to use a single Kinect for our prototype. It is attached to a fixture at the ceiling approx. 1.20 m above the table surface. During the setup process, the transformation between the Kinects' image coordinate systems and their own local coordinate systems has to be found by internal calibration. This is done only once. In addition to that, for each depth-camera the transformation between local space and world space has to be computed. For this purpose, the known size and geometry of the tabletop as well as its capability to display calibration patterns can be facilitated.

During operation, in each frame the iPads need to be recognized and then their spatial locations and orientations have to be determined. In order to do this, a designated server (PC) is fed with the depth images provided by the Kinect. For segmentation, we use the known shape of the iPad, which is a rectangle. Once segmented, the planar equations for each iPad can easily be computed by utilizing well-established algorithms like RANSAC. After transforming these planes into the global coordinate system, the position and orientation of each iPad is wirelessly streamed to all devices via a self-tailored, VRPN-based protocol. Figure 2 depicts an example of using this to control a client application running on the iPad.

Early results show that we can achieve tracking accuracies of about 2 cm at a rate of 20 Hz, but there still is plenty of room for improvement, e.g., by accelerating the computation via graphics hardware. Although in an early stage, we experimented with integrating internal sensors of the iPads, in particular accelerometers and compasses. We are confident that sensor fusion with the information acquired from the depth cameras will considerably improve the precision and reliability of the overall tracking process.

6. CONCLUSION

In this paper, we presented a concept for a low-cost system that aims at bringing spatially aware tangible displays to the masses. Our approach uses active mobile displays like the iPad that are tracked without markers by off-the-shelf consumer hardware, in our case Microsoft's Kinect. This has several benefits compared to earlier solutions. Our approach allows for considerably easier setups with fewer components. With lower costs, it can be made available for new application areas outside of specialized laboratories. Beyond that, by using mobile active displays, users can collaborate on complex interaction tasks and then take the data with them, e.g., to use it beyond the boundaries of a fixed environment.

The employment of depth sensors also allow for a broad range of other applications, such as hand gesture detection and the tracking of passive displays. Although not the focus of this paper, this even involves multi-touch recognition on everyday objects such as a mug.

Acknowledgements

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

7. REFERENCES

[1] D, Bandyopadhyay, R. Raskar, H. Fuchs. 2001. Dynamic Shader Lamps: Painting on Movable Objects. In *Proc. of ISAR'01*. IEEE Computer Society, 207-216.

[2] A. Clark and T. Piumsomboon. 2011. A realistic augmented reality racing game using a depth-sensing camera. In *Proc. of VRCAI '11*. ACM, 499-502.

[3] G. W. Fitzmaurice. 1993. Situated information spaces and spatially aware palmtop computers. *Commun. ACM* 36, 7 (July 1993), 39-49.

[4] C. Harrison, H. Benko, and A. D. Wilson. 2011. OmniTouch: wearable multitouch interaction everywhere. In *Proc. UIST '11*. ACM, 441-450.

[5] D. Holman, R. Vertegaal, M. Altosaar, N. Troje, and D. Johns. 2005. Paper windows: interaction techniques for digital paper. In *Proc. of CHI '05*. ACM, 591-599.

[6] S. Junuzovic, K. Inkpen Quinn, T. Blank, and A. Gupta. IllumiShare: Sharing Any Surface. *Proc. CHI*. 2012, ACM.

[7] J. C. Lee, S. E. Hudson, J. W. Summet, and P. H. Dietz. Moveable interactive projected displays using projector based tracking. In *Proc. of UIST*. 2005. ACM, 63-72.

[8] N. Linder and P. Maes. LuminAR: Portable robotic augmented reality interface design and prototype. In *Adjunct Proc. of UIST*. 2010. ACM, 395-396.

[9] L. H. Nakatani and J. A. Rohrlich. Soft Machines: A Philosophy of User-Computer Interface Design. In *Proc. of CHI.* 1983. ACM, 12-15.

[10] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohli, J. Shotton, S. Hodges, and A. Fitzgibbon. 2011. KinectFusion: Real-time dense surface mapping and tracking. In *Proc. of ISMAR '11*. IEEE Computer Society, 127-136.

[11] M. Spindler, S. Stellmach, and R. Dachselt. 2009. PaperLens: advanced magic lens interaction above the tabletop. In *Proc. of ITS '09*. ACM, 69-76.

[12] M. Spindler, C. Tominski, H. Schumann, and R. Dachselt. 2010. Tangible views for information visualization. In *Proc. of ITS '10*. ACM, 157-166.

[13] M. Spindler, M. Martsch, and R. Dachselt. Going Beyond the Surface: Studying Multi-Layer Interaction Above the Tabletop. In *Proc. of CHI*. 2012. ACM.

[14] B. Ullmer and H. Ishii. 1997. The metaDESK: models and prototypes for tangible user interfaces. In *Proc. of UIST '97*. ACM, 223-232.

[15] A. D. Wilson. 2010. Using a depth camera as a touch sensor. In *Proc. of ITS '10*. ACM, 69-72.

[16] K. Yee. 2003. Peephole displays: pen interaction on spatially aware handheld computers. In *Proc. of CHI '03*. ACM, 1-8.

[17] Z. Zhang, Y. Wu, Y. Shan, and S. Shafer. 2001. Visual panel: virtual mouse, keyboard and 3D controller with an ordinary piece of paper. In *Proc. of PUI '01*. ACM, 1-8.