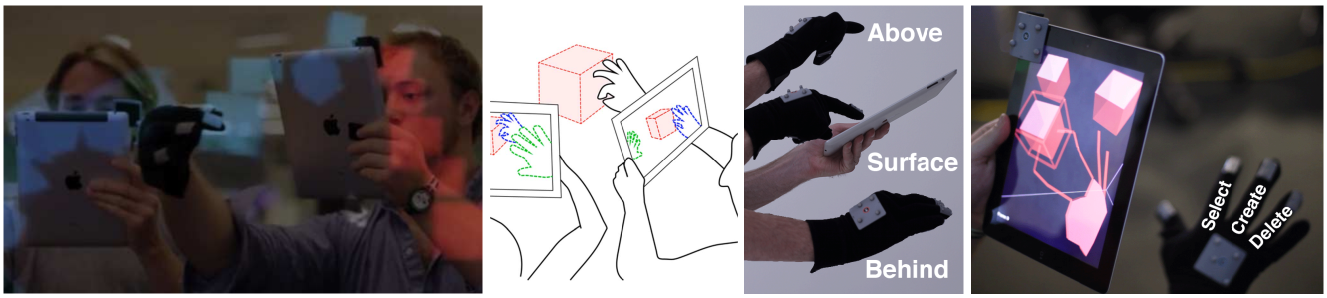
T(ether): Spatially-Aware Handhelds, Gestures and Proprioception for Multi-User 3D Modeling and Animation

Dávid Lakatos[1], Matthew Blackshaw[1], Alex Olwal[1,3,4], Zachary Barryte[1], Ken Perlin[2], Hiroshi Ishii[1]

[1] Tangible Media Group, MIT Media Lab
Cambridge, MA, USA
{dlakatos, mab, olwal, zbarryte, ishii}@media.mit.edu

[2] Media Research Lab, NYU
New York, NY, USA
perlin@mrl.nyu.edu



a) Collaborative 3D manipulation and animation in the physical space. b) Gesture space c) VR viewport and pinch mapping

Figure. 1: a) T(ether) is a system for spatially-aware handhelds that emphasizes multi-user collaboration, e.g., when animating a shared 3D scene. b) Gestural interaction *above*, on the *surface*, and *behind* the handheld leverages proprioception and a body-centric frame of reference. (c) The UI provides a perspective-correct VR view of the tracked hands and 3D objects through the head-tracked viewport, with direct control through the spatial 3D UI.

ABSTRACT

T(ether) is a spatially-aware display system for multi-user, collaborative manipulation and animation of virtual 3D objects. The handheld display acts as a window into virtual reality, providing users with a perspective view of 3D data. T(ether) tracks users' heads, hands, fingers and pinching, in addition to a handheld touch screen, to enable rich interaction with the virtual scene.

We introduce gestural interaction techniques that exploit proprioception to adapt the UI based on the hand's position *above*, *behind* or on the *surface* of the display. These spatial interactions use a tangible frame of reference to help users manipulate and animate the model in addition to controlling environment properties. We report on initial user observations from an experiment for 3D modeling, which indicate T(ether)'s potential for embodied viewport control and 3D modeling interactions.

Author Keywords

3D user interfaces, Spatially-aware displays, Gestural interaction, Multi-user, Collaborative, 3D modeling, VR.

[3] KTH, Royal Institute of Technology, Stockholm, Sweden
[4] Google [x], Mountain View, CA, USA

SUI '14, October 04 – 05, 2014, Hololulu, HI, USA.

ACM 978-1-4503-2820-3/14/10…$15.00.
http://dx.doi.org/10.1145/2659766.2659785

ACM Classification Keywords

H.5.2 User Interfaces: Interaction styles; I.3.6 [Methodology and Techniques]**:** Interaction techniques.

INTRODUCTION

We are seeing an increasing amount of devices that have the capability for advanced context and spatial awareness thanks to advances in embedded sensors and available infrastructure. Recent advances have made many relevant technologies available in a portable and mobile context, including magnetometers, accelerometers, gyroscopes, GPS, proximity sensing, depth-sensing cameras, and numerous other approaches for tracking and interaction. Previous work has extensively explored the use of spatially aware displays, but primarily focuses on single-user scenarios and how the display's tracked position in the 3D space can be used to interact with virtual contents.

In this paper, we introduce T(ether), a prototype system that specifically focuses on novel interaction techniques for spatially aware handhelds. It leverages proprioception to exploit body-centric awareness, and it is specifically designed to support concurrent and co-located multi-user interaction with virtual 3D contents in the physical space, while maintaining natural communication and eye contact.

We report on initial user observations from our 3D modeling applications that explores viewport control and object manipulation.

RELATED WORK

The concepts of using tracked displays as viewports into Virtual Reality (VR), as introduced by McKenna [8] and Fitzmaurice [4], has inspired numerous related projects.

Spatially-Aware Displays

The Personal Interaction Panel [15] is a tracked handheld surface that enables a portable stereoscopic 3D workbench for immersive VR. Boom Chameleon's [16] mechanically tracked VR viewport on a counter-balanced boom frees the user from holding the device, but limits motion with mechanical constraints. Yee [17] investigates spatial interaction with a tracked device and stylus. Collaborative AR has been explored with head-mounted displays (HMDs) [14] and on mobile phones [6]. Yokokohji et al. [18] add haptic feedback to the virtual environment observed through a spatial display. Spindler et al. [13] combine a large tabletop with projected perspective-correct viewports. The authors present several interesting concepts but also describe interaction issues with their implemented passive handheld displays due to lack of tactile feedback, constrained tracking and projection volume, and limited image quality. T(ether) focuses specifically on supporting rich interaction, high-quality graphics and tactile feedback. We therefore extend the stylus, touch and buttons used in the above-mentioned projects, with proprioceptive interaction techniques on and around active displays that form a tangible frame of reference in 3D space.

Gestural Interaction and Proprioception

Early research in immersive VR demonstrated powerful interactions that exploited 3D widgets, remote pointing and body-centric proprioception [3, 9, 11]. Advancements in tracking and display has allowed the use of more complex gestural input for wall-sized user interfaces (UIs), shape displays [8], augmented reality [10], and volumetric displays [5]. T(ether) emphasizes proprioceptive cues for multi-user interactions with unhindered, natural communication and eye contact.

Multi-user Interaction

Related work on multi-user 3D UIs with support for face-to-face interaction [1, 5] focuses on workspaces with support for a small number of users, while T(ether) emphasizes a technical infrastructure to support large groups of users for room-scale interaction with full body movement for navigation.

INTERACTION TECHIQUES

T(ether) extends previous work through an exploration of gestures that exploit proprioception to advance the interaction with spatially aware displays. By tracking the user's head, hands, fingers and their pinching, in addition to a handheld touch screen, we enable multiple possibilities for interaction with virtual contents. Head tracking relative to the display further enhances realism in lieu of stereoscopy by enabling perspective-correct rendering [8].

For body-centric, proprioceptive interaction, we use the tablet to separate the interaction into three spaces:

- *Behind.* Direct manipulation of objects in 3D.
- *Above*. Spatial control of global parameters (e.g., time).
- *Surface*. GUI elements, properties and tactile feedback.

The available functions in each of these spaces are mutually exclusive by design, and the switch between them is implicit. The view of the interactive virtual 3D environment is shown on the display when the user's hand is behind the tablet, while the GUI appears when the hand is moved above or in front of it.

We use a 6DOF-tracked glove with pinch-detection for 3D control and actuation in the spirit of previous work [10]. Our initial user observations indicate that pinch works well also in our system. Pinching an object maps to different functions based on whether the thumb pinches the index (select), middle (create) or ring (delete) finger (Figure 1c).

Behind: Direct manipulation of virtual 3D shapes

Create. Pinching the middle finger to the thumb adds a new shape primitive. The shape is created at the point of the pinch, while the orientation defaults to align with the X-Y plane of the virtual world. The distance between the start and release of the pinch determines object size. When the user begins creating a shape, other entities in the scene (objects, hand representations and other users' positions) become transparent to decrease visual load and for an unhindered view of the current operation. T(ether) currently supports lines, spheres, cubes and tri-meshes.

Select. As the user moves their hand "behind" the screen, the "cursor" (a wire-frame box) indicates the closest entity, and allows selection of objects, or vertices of a mesh.

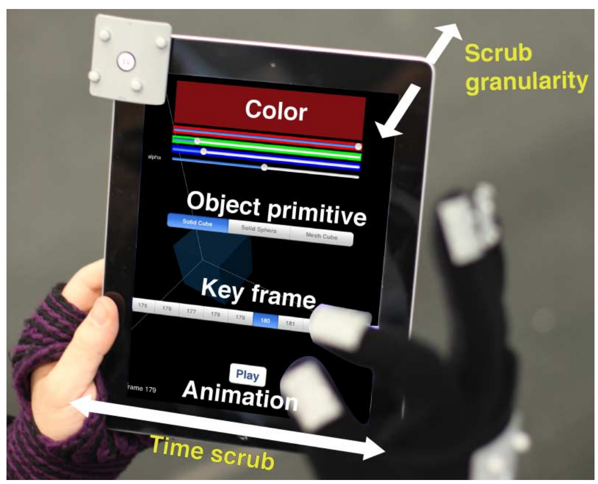


Figure 2: T(ether) adapts the spatial UI for the most relevant interactions based on the location of the user's hand. In our 3D modeling and animation application, gestures for navigating time are available *above* (yellow) the display, while settings and GUI controls are available on its *surface* (white).

Manipulate. After selection, the user can pinch the index finger to the thumb for 1:1 manipulation. Objects are translated and rotated by hand movement while pinched. Transformations are relative to the starting pinch pose. Users can select and manipulate vertices to deform meshes.

Delete. Pinching of ring finger and thumb deletes entities.

Above: Spatial 3D parameter control

A key-frame based animation layer built into our system allows users to animate virtual objects. Key frames are recorded automatically when a user modifies the scene. The user can animate an object by recording its position in one key frame, transforming it and moving the current key frame to match the desired duration of the animation. The user has access to the key frame engine through the pinch gesture above the screen, as shown in Figure 2.

The user can scrub through key frames by pinching the index finger and moving it left (rewind) or right (fast forward) relative to the tablet. The user can adjust the granularity of scrubbing by moving the pinched hand away from the tablet. By anchoring hand motions relative to the tablet, the tablet becomes a tangible frame of reference. Similarly to how the ubiquitous "pinch to zoom" touch gesture couples translation and zooming, we couple the time scrubbing and its granularity in order to allow users to rapidly and precisely control key frames.

Surface: GUI and Tactile Surface for 2D Interaction

Object properties. A UI fades in when the hand moves from *behind* to *above* the screen. Here, users configure settings for new objects, such as primitive type (cube, sphere or mesh cube) and color.

Animation. The 2D GUI also provides control over the animation engine and related temporal information, such as indication of the current key frame and scrubbing granularity. Users manipulate animation playback through different controls, such as the on-screen Play/Stop button.

Annotation. Freehand content can be draw on the tablet's plane and will be mapped to the virtual environment based on the tablet's pose [11]. The user can annotate the scene and create spatial drawings by simultaneously moving the tablet in space while touching the surface.

IMPLEMENTATION

Our handheld display software is implemented using C++ and the Cinder low-level OpenGL-wrapper with our custom Objective-C scene graph, to allow native Cocoa UI elements on Apple iPad 2 (600 g). We obtain the position and orientation of tablets, users' heads and hands through attached retro-reflective tags that are tracked with 19 cameras in a G-speak motion capture system (http://www.oblong.com/), covering a space of 14×12×9 ft. Our gloves use one tag for each finger and one for the palm. We enable capacitive pinch-sensing with a woven conductive thread through each fingertip.

Our server software is implemented in Node.js (http://nodejs.org/) and handles tag location broadcasts and synchronization of device activity (sketching, model manipulation, etc.) and wirelessly transmits this data to the tablets (802.11n). System performance is related to scene complexity, but in our experiences with user testing and hundreds of objects and multiple collaborators, frame rates have been consistently above 30 Hz.

INITIAL USER OBSERVATIONS

To assess the potential of T(ether), we conducted an experiment to explore its 3D modeling capabilities.

3D Modeling

Participants. We recruited 12 participants, 19–40 years old (3 females), from our institution that were compensated with a $50 gift card. All were familiar with tablets, 8 had used traditional CAD software, and none had experience with T(ether). Session lasted approximately 40-90 min.

Procedure. In a brief introduction (10–15 min), we demonstrated T(ether)'s gestural modeling capabilities. Once participants got familiar with the gestural interaction, we introduced them to the on-surface GUI for modifying object properties. Participants received training (15–30 min) in the Rhinoceros (Rhino3D) desktop 3D CAD software (http://www.rhino3d.com/), unless they were experts in it.

Conditions. Participants performed three tasks, first with T(ether) and then in Rhino3D. In the sorting task, participants sorted a random mix of 10 cubes and 10 spheres into two groups. In the stacking task, participants were instructed to create two cubes of similar size and stack and align them on top of each other. Then they repeated this task for 10 cubes. In the third task, participants recreated a random 3D arrangement of 6 cubes and 3 spheres with some of the objects stacked.

Observations. Participants were able to perform all functions in both interfaces. Using the body for "walking through data" was "a very appealing" approach to viewport manipulation and was considered easier than in traditional CAD. Some participants especially appreciated that they "regained peripheral awareness", since the "body is the tool" for viewport control. Shape creation and manipulation was generally "easy" and "straight-forward". They enjoyed the "unprecedented" freedom of the system, although some of them commented that the alignment relative to other objects was "tricky" and suggested inclusion of common features from traditional CAD, such as grids, snapping and guided alignment operations.

Discussion

Our experiment confirmed that with little training, participants could indeed perform basic 3D modeling tasks in our spatial UI. The observations especially highlight how participants appreciate the embodied interface and viewport control for navigating the 3D scene in the physical space.

While more complex 3D modeling would benefit from widgets, constraints and interaction techniques found in traditional CAD, we believe that the experiment illustrates the potential of spatially aware handhelds, as discussed in previous work [8, 4, 16], while leveraging modern, high-resolution widely available multi-touch displays, and a massively scalable infrastructure.

LIMITATIONS AND FUTURE WORK

Our system is currently using an untethered tablet to support multi-user interaction and mobility. Similarly to previous work [8, 4, 15, 11, 17] and handheld mobile augmented reality systems, there is, however, a risk for fatigue when using a handheld device as a viewport and interaction surface. This could be of particular importance for 3D modeling scenarios, where participants may be expected to interact for extended time. We believe that these issues will be partially addressed through advances in hardware with increasingly lighter handhelds or by using projection surfaces [13]. Mid-air interaction can, however, also affect precision and the quality of interaction, issues that require additional investigation to assess their impact on our scenarios. THRED [12] indicates that carefully designed bi-manual mid-air interaction does not necessarily need to result in more pain or fatigue than a mouse-based interface. If mobility is not required, then counterbalanced mechanical arms could also be introduced [16].

In future work we would like to extend collaborative spatial modeling by integrating advanced functionality from Open Source tools like Blender (http://www.blender.org/) and Verse (http://www.quelsolaar.com/verse/). State-of-the-art software and hardware for location and mapping, e.g., Project Tango (https://www.google.com/atap/projecttango), are natural next steps to implement our techniques without infrastructure. Similarly, mobile depth cameras and eye tracking would enable improved perspective tracking and detailed shape capture of hand geometry. This could, e.g., enable more freeform clay-like deformation of virtual contents. Gaze tracking could also improve multi-user scenarios by rendering collaborators' field-of-view and attention. For improved feedback from virtual content, we believe that the passive feedback from the physical tablet surface could be complemented with techniques like TeslaTouch [2], instrumented gloves, and passive or actuated tangible objects in the environment. In fact, some of our study participants already used physical objects in the space for reference when placing and retrieving virtual content. Physical objects not only have the benefit of tactile feedback, but also improve legibility for collaborators with or without a personal T(ether) display.

We believe that much potential lies in further exploring massive collaborative scenarios with a large number of participants and complex scenes. Our network-distributed architecture would also make it straightforward to explore our techniques for remote collaboration scenarios, with distributed teams for various types of applications, such as architectural visualizations, augmented reality and virtual cameras for movie production.

CONCLUSIONS

Today's interfaces for interacting with 3D data are typically designed for stationary displays that limit movements and interaction to a single co-located user. T(ether) builds on previous research for spatially aware handheld displays, but with an emphasis on gestural interaction and proprioception in its use of the display as a tangible frame of reference. T(ether) was also designed for multi-user, collaborative, concurrent and co-located spatial interaction with 3D data and focuses on technology that minimizes interference with human-human interaction.

ACKNOWNALEDGMENTS

We thank the members of the Tangible Media group and the MIT Media Lab. Alex Olwal was supported by the Swedish Research Council.

REFERENCES

1. Agrawala, M., Beers, A., McDowall, I., Fröhlich, B., Bolas, M., and Hanrahan, P. The two-user Responsive Workbench: support for collaboration through individual views of a shared space. Proc. SIGGRAPH '97, 327-332.
2. Bau, O., Poupyrev, I., Israr, A., and Harrison, C. TeslaTouch: electrovibration for touch surfaces. Proc. UIST '10. 283-292.
3. Bowman, D., and Hodges, L., F. An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. Proc. I3D '97, 35-39.
4. Fitzmaurice, G.W. Situated Information Spaces and Spatially Aware Palmtop Computers. Com. ACM (1993), 36(7), 38-49.
5. Grossman, T. and Balakrishnan, R.. Collaborative interaction with volumetric displays. Proc. CHI '08, 383-392.
6. Henrysson, A., Billinghurst, M., and Ollila, M. Virtual object manipulation using a mobile phone. Proc. ICAT '05, 164-171.
7. Leithinger, D., Lakatos, D., DeVincenzi, A., Blackshaw, M., and Ishii, H. Direct and gestural interaction with relief: a 2.5D shape display. Proc. UIST '11, 541-548.
8. McKenna, M. Interactive viewpoint control and three-dimensional operations. Proc. I3D '92, 53-56.
9. Mine, M. R. 1996. Working in a Virtual World: Interaction Techniques Used in the Chapel Hill Immersive Modeling Program. Technical Report.
10. Piekarski, W., Thomas, B. H. Through-Walls Collaboration. IEEE Pervasive Computing 8(3), 42-49.
11. Poupyrev, I., Tomokazu, N., Weghorst, S., Virtual Notepad: handwriting in immersive VR. Proc. VR '98. 126-132.
12. Shaw, C. Pain and Fatigue in Desktop VR. Proc. GI '98. 185-192.
13. Spindler, M., Büschel, W., and Dachselt, R. Use Your Head: Tangible Windows for 3D Information Spaces in a Tabletop environment. Proc. ITS '12, 245-254.
14. Szalavári, Z., Schmalstieg, D., Fuhrmann, A. and Gervautz, M. Studierstube: An environment for collaboration in augmented reality. Virtual Reality (1998), 3:37-48.
15. Szalavári, Zs,, Gervautz, M. The Personal Interaction Panel – a Two-Handed Interface for Augmented Reality. Computer Graphics Forum (1997), 16(3).
16. Tsang, M., Fitzmaurice, G, Kurtenbach, G., Khan A., and Buxton, B. Boom chameleon: simultaneous capture of 3D viewpoint, voice and gesture annotations on a spatially-aware display. Proc. UIST '02, 111-120.
17. Yee, K-P. Peephole displays: pen interaction on spatially aware handheld computers. Proc. CHI '03, 1-8.
18. Yokokohji, Y., Hollis, R. L., and Kanade, T. What you can see is what you can feel - Development of a visual/haptic interface to virtual environment. Proc. VRAIS '96, 46-53.