Olwal, A. Unencumbered 3D Interaction with See-through Displays.
Proceedings of *NordiCHI 2008 (Nordic Conference on Human-Computer Interaction)*, Lund, Sweden, October 18-22, 2008, pp. 527-530.

Unencumbered 3D Interaction with See-through Displays

Alex Olwal

School of Computer Science and Communication, KTH (Royal Institute of Technology), Stockholm, Sweden

alx@csc.kth.se

ABSTRACT

Augmented Reality (AR) systems that employ user-worn display and sensor technology can be problematic for certain applications as the technology might, for instance, be encumbering to the user or limit the deployment options of the system. Spatial AR systems instead use stationary displays that provide augmentation to an on-looking user. They could avoid issues with damage, breakage and wear, while enabling ubiquitous installations in unmanned environments, through protected display and sensing technology.

Our contribution is an exploration of compatible interfaces for public AR environments. We investigate interactive technologies, such as touch, gesture and head tracking, which are specifically appropriate for spatial optical see-through displays. A prototype system for a digital museum display was implemented and evaluated. We present the feedback from domain experts, and the results from a qualitative user study of seven interfaces for public spatial optical see-through displays.

Categories and Subject Descriptors

H.5.1 **[Multimedia Information Systems]**: Artificial, augmented, and virtual realities; H.5.2. **[User Interfaces]**: Graphical user interfaces, Input devices and strategies; I.3.6 **[Methodology and Techniques]**: Interaction techniques.

General Terms

Experimentation, Human Factors

Keywords

Interaction, interface, augmented reality, mixed reality, public display, spatial display, see-through, 3D, touch, gesture, pose

1. INTRODUCTION

Augmented Reality (AR) systems often use head-mounted-displays and user-worn sensors to merge the real world with computer generated imagery. User-worn technology can, however, be unsuitable for many scenarios and we instead focus on installations where it is acceptable to introduce some infrastructure in return for a walk-up-and-use system.

Spatial optical see-through AR systems use stationary displays in the environment, unlike traditional mobile AR systems where the display system is worn. They could be ideal in public, unmanned environments, since issues with damage, breakage and general wear can be avoided by making the installed equipment physically inaccessible to the users. There are two important criteria that need to be met for spatial optical see-through systems. First, the display component must remain optically transparent, to ensure unmediated augmentation. Second, the installed interactive technology must work *on* or *through* the surface, without blocking the view through the display and without mechanical parts on the user's side. Traditional input devices, such as mice, trackballs, joysticks and keyboards, are thus outside the scope of this work.

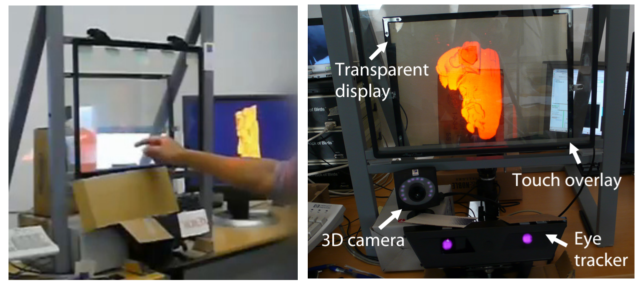

Figure 1. We explore seven interfaces for public see-through displays. They are designed to provide untethered input that is compatible with transparent see-through 3D displays for spatial augmented reality. The interfaces may however also be applied to traditional 3D displays and virtual reality.

A touch-sensitive surface avoids traditional input devices and is a possible solution for interaction on stationary spatial AR displays. Interactive film[1] using projective capacitance allows touch input through a solid (e.g., glass) surface and can be combined with a holographic display material[2], for a transparent touch-sensitive surface. Wilson's TouchLight [12] uses a holographic display material and stereo cameras for multi-touch interaction based on optical flow, while Bimber [2] combines holograms with autostereoscopic displays, touch-sensitive surfaces and haptic devices. Paradiso and colleagues present techniques for tracking hands using custom laser range-finders, and acoustic tap tracking [10]. Multi-touch input using frustrated total internal reflection (FTIR) works inherently with transparent displays [5], in contrast to most other camera-based techniques, which were designed for diffuse surfaces.

While the desired system should avoid external technology that is part of the installation, we find it compelling to also investigate the support for interaction with the user's personal mobile phones. The advantages include a richer set of input [11][4] and the user's familiarity with the physical affordances of the device.

2. MUSEUM SCENARIO

The Vasa museum in Stockholm has a mission to preserve and make accessible artifacts from the world's only salvaged 17th-century battleship, which sunk in 1628. The preservation process involves obtaining 3D-scanned models of the sculptures, which are deteriorating with time. The goal of making a larger part of

[1] Rearpro Interactive Screens. http://www.rearpro.com/. September 2008. Interactivity Window. http://www.interactivity.com.au. September 2008. iWindow. http://www.iwindow.be/. September 2008.

[2] dnp Holo Screen. http://www.dnp.dk/. September 2008.

the museum collection available to visitors motivates AR and VR visualizations, since issues such as space constraints, fragility, restoration procedures and loans to other institutions limit the accessibility of artifacts. We were partly inspired by The Virtual Showcase project [1], which effectively demonstrates how virtual information can complement real artifacts with virtual graphics. The project however depends on user-worn equipment in the described implementation.

In this work we focus on manipulation of virtual models in a real space, which allows more freedom regarding interaction, since the geometry can be freely scaled and rotated.

Three employees at the Vasa museum were invited for an informal interview session to assess applicability and usefulness of the prototyped system in a public scenario. They were 31, 39 and 47 year-old males, and had worked with information technology in the museum for 3.5, 13 and 12 years, respectively.

One participant expressed a general interest in interactive technologies by describing ambitions to move beyond the traditional notion of "the passive visitor" and "dead presentations" associated with museums today. The participants emphasized that the motivation for digital technology was the potential for a deeper experience and understanding. It was however pointed out that visitors still very much like to touch and feel artifacts. The limited number of interactive exhibits was both attributed to the tradition of the "passive visitor", but also to limitations in conservation technology. The artifacts are too fragile for interaction and visitors are generally less interested in copies.

Two participants had negative personal experiences with touch screens in museums, due to poor calibration and distracting offsets between mouse cursor and touch location. They pointed out that touch-screen computers and their user interfaces have too much of a "computer look-and-feel".

In a brainstorming session of ideas for a "dream" exhibit, participants mentioned things related to physical sensation ("pick up things", "feel the weight/shape/texture of an artifact"), activities ("go aboard", "load a cannon", "shoot a rifle", "cook food") and visuals ("the ship in full sails", "the wood in original color", "the crew running around on deck", "what the inside looks like", "what it looked like when it was salvaged"). They all argued that a richer experience would most likely lead to a better understanding for how and why things worked in a certain way during the time period.

3. EXPERIMENTAL SETUP

ASTOR, an autostereoscopic spatial optical see-through display [9], simultaneously provides monochromatic 3D augmentation and a clear view of the real environment behind the display (See Figure 1). We complemented this system with a range of input devices compatible with a spatial optical see-through system and discuss how each supports various levels of tracking and actuation. While the interfaces are applicable to AR scenarios, we focus on interaction with virtual content in this study. As the primary goal for our public scenarios is to support inspection, the priority was to support yaw/pitch rotations and zoom (uniform scaling) of a 3D model in this work.

3.1 Touch-input with buttons and gestures

We fit a transparent single-touch overlay (Magic Touch KTMT-1921) in front of the display, which allows users to employ simple gestures for direct manipulation of the content.

The user can enter one of three modes ("rotate", "move" and "zoom") using on-screen buttons at the top of the screen. Yaw/pitch is controlled in "rotate mode", as the user drags the finger horizontally/vertically across the display. Relative motion allows the user to reposition for unlimited rotation. Dragging gestures translate the object in 2D (in the plane of the display) in "move mode". "Zoom mode" allows the user to drag the finger to the right or upwards for enlargement, or in the opposite direction to shrink the object. Two additional buttons in the top-left ("previous model") and top-right ("next model") corners allow the user to cycle among different 3D models.

We also included the *Rubbing* [7] touch-screen technique, which uses small repetitive strokes along a diagonal for "rubbing" actions, where the slope of the diagonal distinguishes between two actions. In our application, rubbing along the diagonal with a positive slope enlarges the model, while rubbing along the diagonal with a negative slope shrinks it, making it possible to support both zoom and rotation (or movement) in a single mode.

3.2 Head movement to control rotation

To remotely detect and track the pose of the user's head we use a remote eye tracker (Tobii Technology X50) that recovers the 3D position of the user's eyes. It provides precise indication of the user's viewpoint from a distance, such that it can be mounted behind the glass.

Actuation in gaze-only-controlled applications could be achieved using blinking and dwell time, for instance, but we found them inappropriate for our interface, as they would require training. We thus chose to not support clutching or mode switching for the eye tracker, and instead rely on the absolute position of the user's head. Similar to previous work on head movement for controlling rotation [6][3], we map the average eye position to yaw/pitch rotations based on the user's horizontal/ vertical movement. Since the possible head motion is small, we amplify the effect of movement on rotation, such that the model rotates more than the user moves [6].

3.3 Hand gestures to control rotation

A depth-sensing camera (3DVSystems Zcam prototype) provides us with a clean, segmented image of the user's head, body and hands (with an estimate of the number of visible fingers). We experimented with numerous interactions, and found a rotation-only mapping to work most satisfactory and robustly. As the user holds up one or more fingers, the system starts tracking the hand and maps horizontal/vertical movement to yaw/pitch rotation. Clutching is supported by closing the hand and repositioning it.

3.4 Remote control using a mobile device

The lack of physical affordances in the previous techniques could be a limitation, which led us to also implement support for interaction using the user's personal mobile device.

3.4.1 Control through buttons

A general mobile device (in our experiments, a Sony Ericsson K800i) can be used to rotate the model with cursor keys or using a joystick, to move it using the 4/6/2/8 buttons (for ←, →, ↑, ↓, respectively), and using the star (*) and pound (#) keys for zooming in/ out. Two soft buttons under the display are mapped to roll. The activation or select button shows the next 3D model.

3.4.2 Touch-screen device

For touch-screen devices with a scroll wheel and numeric keypad (we used a Sony Ericsson W960i), we map yaw/pitch rotations to horizontal/vertical finger movement on the touch screen, and roll to scroll wheel movement. Translation and scaling is performed similarly to the button-only devices, using the numeric keypad.

3.4.3 Tracked mobile device

The on-board mobile phone camera can be used to track the device relative to the environment (e.g., using fiducials, natural features or optical flow). By placing some identifiable features, for instance, underneath the display, the user could point at it with the camera, and control the 3D model using relative motion with the device [4]. Our LightSense system [8] instead uses an exocentric approach, where an installed camera tracks the 2D position of a mobile device's (Sony Ericsson K850i) photo light, and we map horizontal/vertical phone movement in the 2D plane to yaw/pitch rotations.

4. EVALUATION

Initial feedback acquired from informal individual sessions with three colleagues, led to our first interface modifications. We then gathered informal group feedback from the museum staff from our interview session, by having them test out the interfaces. While both the head movement and hand gestures were found problematic for ergonomic and practical reasons, the amplified rotation from head movement was appreciated, as well as the interesting potential of being able to provide access for certain disabled users. The mobile joystick was found to be easy to understand and use, and considered good for continuous manipulation, while buttons were preferred for mode switching. Interacting with the tracked device (using the photo/video light) was problematic since the device (being held in front of the user) obstructed the user's view of the scene and the light created reflections in the glass. The smooth manipulation made possible by the continuous input from the mobile touch screen was appreciated. The participants generally found the touch-screen techniques to be self-explanatory, the easiest to use and their definite favorites. The only concern for the stationary display was regarding fatigue and a display at an inclination (e.g., at 45°) was suggested. They were positive about this type of continuous touch-screen interaction, which was not dominated by discrete button presses, and thus was significantly different from the systems they had seen in other exhibition contexts. On-screen buttons were considered clear, easy to understand, and straightforward to extend with language-independent graphical symbols.

A pilot study with three students allowed for final adjustments before we ran a qualitative user study with 12 new participants.

4.1 Participants

Four female and eight male uncompensated participants between 20 and 32 years old (average age 26, standard deviation 3.73) were recruited. Eight were university students and four were engineers. None of them had experience with our system or interfaces. Four participants owned or had owned a touch-screen device and ten participants were positive about touch-screen devices, while two were neutral. Ten had used 3D programs a few times, one used them often, while one participant had no experience. Ten had used 3D map applications (e.g., Google Earth), four had used a 3D modeling application, and two had used a 3D viewer.

4.2 Procedure

We divided the evaluation into seven interface categories based on hardware device and interaction. Although several of them could be used in combination or in parallel, we evaluated each of the interfaces individually to assess their respective potential.

1) Touch-screen with on-screen buttons
2) Touch-screen with rubbing for zooming
3) Mobile device with buttons
4) Mobile device with touch screen
5) Tracked mobile device
6) Head movement
7) Hand gestures

The first four interfaces support all operations (rotation, movement, zooming, cycling through models), while the last three support only rotations.

The participant first provided background information by answering questions in a questionnaire. The experimenter then gave a brief introduction and motivation for the interfaces. The participant's task was to complete transformations (primarily rotation and zoom) of 3D models on the autostereoscopic display using the seven different interfaces, the order of which was randomized for the participants. While some interfaces supported more transformations than others, our primary goal in this preliminary study was to get feedback on user experience with the input devices and their associated interaction, rather than measuring performance. Each interface was preceded by a demonstration from the experimenter and the participant was free to discuss it during the test. The participant filled out a questionnaire after each interface, ranking it according to seven criteria (See 4.3 Results, below) on a seven-point Likert scale (-3 to 3) and were also encouraged to provide written comments. There was no time limit, but each interface test typically lasted for about five minutes, including the questionnaire. After all interfaces had been tested, a final questionnaire gathered comparative feedback for all of them. The test lasted from approximate 45 minutes to one hour.

4.3 Results

The questionnaire following each interface allowed the participants to rank them according to learnability, usability, comfort, preference, speed, effiency and fatigue. All interfaces, except head movement and hand gestures received high positive scores over all criteria. Figure 2 shows neutral to vaguely positive median scores for head movement, whereas the hand gestures were vaguely positive to vaguely negative. The mode values show negative usability, comfort and efficiency for hand gestures.

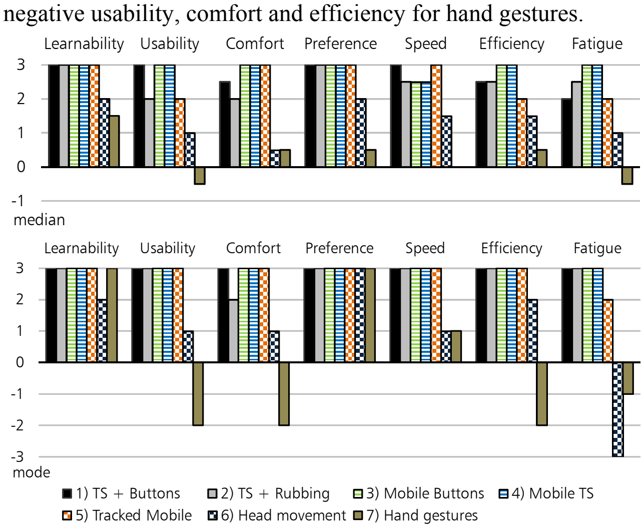

Figure 2. Median and mode plots of qualitative user feedback.

The participants also picked the best and worst interface for each category. They were allowed to name more than one technique or to not specify any technique at all for a category. The familiar

Olwal, A. Unencumbered 3D Interaction with See-through Displays.
Proceedings of *NordiCHI 2008 (Nordic Conference on Human-Computer Interaction)*, Lund, Sweden, October 18-22, 2008, pp. 527-530.

interfaces (touch screen, and mobile device with buttons or touch screen) got the best scores for all categories except in experience. The head movement and hand gestures were picked most frequently as the worst interfaces. (See Figure 3.)

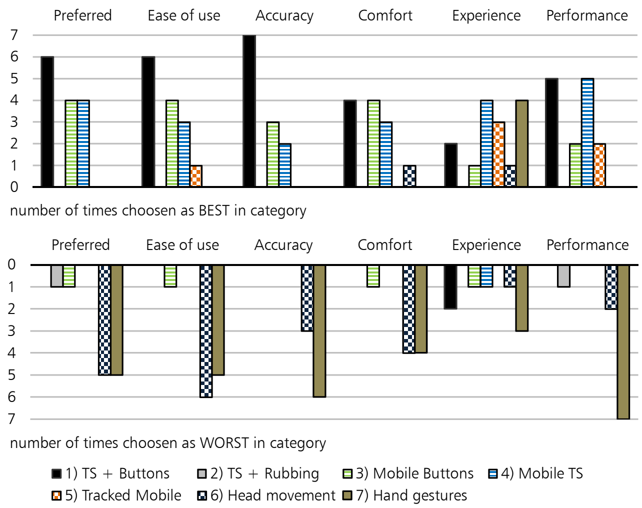

Figure 3. The number of times an interface was chosen as the best (top) or worst (bottom) in a category.

4.4 Discussion

The study indicates that the touch screen interface was the most popular, which could be related to the available tactile and graphical feedback, robust sensing, and familiarity (four participants owned or had owned a touch-screen device). The head and hand tracking in the system was a bit sensitive to external factors, such as the user's distance and position relative to the device, or the wearing of contact lenses or glasses in the case of the eye tracker. While the tracking robustness and range of these devices will likely be improved with upcoming products, several participants liked their novelty. Some of the negative feedback might also be related to the lack of head movement actuation, and the somewhat artificial hand gesture actuation (holding up one or more fingers).

Familiarity and discoverability can explain why the button-based touch-screen interaction received higher scores than rubbing. The buttons at the top of the screen are self-explanatory, while rubbing has to be learnt. Rubbing, on the other hand, requires less mode switching, but this performance advantage [7] was not of significant importance for the tasks in this study.

There were several verbal and written comments on the comfort and convenience in using the phone as a wireless touchpad or joystick without needing to stretch out one's hand. As long as connectivity and multi-user issues are not a problem, a remote device might be a useful alternative approach for rich input. However, the tracked device had issues with reflections from its light (which could be addressed using other tracking techniques), occlusion of displayed content, and fatigue.

The mostly positive or neutral individual results suggest that most techniques (after some tweaking) will work in situations where alternate techniques are not available. It is also interesting that the simplest interface, with the least sophisticated hardware, an ordinary single-touch glass surface, provided the best user experience. It seems to be well-suited for basic manipulations on interactive public displays. A natural extension would upgrade the single-touch device to a multi-touch device based on FTIR [5].

5. CONCLUSIONS AND FUTURE WORK

We have presented seven interfaces (based on hardware devices and interaction techniques) that are compatible with a public spatial optical see-through display system and we have provided qualitative feedback based on interaction with virtual content. The results of this study suggest that our interface based on a touch-sensitive surface, or a touch-screen mobile device, works well in a spatial optical see-through configuration. The experiences from this study should now be applied to AR, where the user would interact with both real and virtual content.

We have focused on evaluating each specific interface in isolation in this work, but are interested in looking at combinations and variations of the approaches, which could improve performance and user experience for all devices.

The study indicates that individual use of the implemented head and gesture tracking techniques might be less appropriate for direct manipulation by novice users. However, we see an interesting potential in their use as complementary remote and unobtrusive sensing technology that can assist other modalities interpreting user actions. While the emphasis on novice users is especially relevant for public installations, future work includes interfaces for skilled expert users, where remote head and gesture sensing could be significantly more beneficial. Such interfaces may support more complex scenarios, which could identify additionally challenging and interesting interface issues in spatial see-through systems.

6. ACKNOWLEDGMENTS

We thank the School of Industrial Engineering and Management at KTH, and Jonny Gustafsson in particular, for making the 3D display technology available to us. Thanks to Liu Yuhang and Tahir Mustafa Madni Ikram for administration of the qualitative study, to the user study participants, the Vasa museum, and to Steven Feiner and Yngve Sundblad for valuable feedback and comments on the paper. 3DVSystems kindly provided the Zcam prototype and Ericsson provided the cell phones used in the study.

REFERENCES

[1] Bimber, O., Encarnação, L. M., and Schmalstieg, D. 2003. The virtual showcase as a new platform for augmented reality digital storytelling. EGVE '03, vol. 39, 87-95.
[2] Bimber, O., 2004. Combining optical holograms with interactive computer graphics. IEEE Computer 37, 1, 85-91.
[3] Chung, J. C. 1992. A comparison of head-tracked and non-head-tracked steering modes in the targeting of radiotherapy treatment beams. SI3D '92, 193-196.
[4] Hachet, M. and Kitamura, Y. 2005. 3D Interaction with and from handheld computers. IEEE VR 2005 Workshop: New Directions in 3D User Interfaces, 11-15.
[5] Han, J. Y. 2005. Low-cost multi-touch sensing through frustrated total internal reflection. UIST '05, 115-118.
[6] Hix, D., Templeman, J. N., and Jacob, R. J. 1995. Pre-screen projection: from concept to testing of a new interaction technique. CHI '95, 226-233.
[7] Olwal, A., Feiner, S., and Heyman, S. 2008. Rubbing and Tapping for precise and rapid selection on touch-screen displays. CHI '08, 295-304.
[8] Olwal, A., 2006. LightSense: enabling spatially aware handheld interaction devices. ISMAR '06, 119-122.
[9] Olwal, A., Lindfors, C., Gustafsson, J., Kjellberg, T., Mattsson, L. 2005. ASTOR: an autostereoscopic optical see-through augmented reality system, ISMAR '05, 24-27.
[10] Paradiso, J. A. 2003. Tracking contact and free gesture across large interactive surfaces. Comm. of the ACM 46, 7, 62-69.
[11] Watsen, K., Darken, R., and Capps, M. 1999. A Handheld Computer as an Interaction Device to a Virtual Environment. IPTW '99.
[12] Wilson, A. D. 2004. TouchLight: an imaging touch screen and display for gesture-based interaction. ICMI '04, 69-76.