SView'n'MTouch: Interacting with Distant Objects in Immersive AR Using a Transparent Plane Prop

Joon Hyub Lee, Sang-Gyun An, Yongkwan Kim, Hyung-Gi Ham, Seok-Hyung Bae
Department of Industrial Design, KAIST
Daejeon, Republic of Korea
joonhyub.lee | sang-gyun.an | yongkwan.kim | hyunggi.ham | seokhyung.bae @ kaist.ac.kr

Figure 1. A small, flat, transparent panel has captured popular imagination for years and could serve as a haptic peripheral affording on-surface and through-surface interactions.

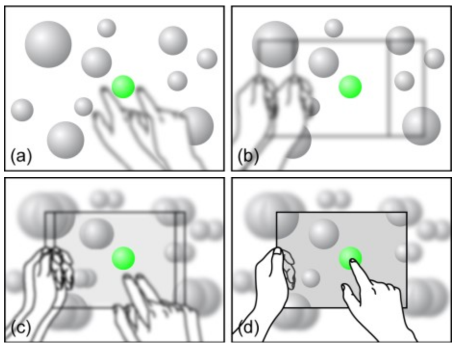

Figure 2. (a) In immersive AR, lack of haptic feedback and presence of binocular parallax make direct interactions, such as selection, with a distant virtual object difficult. With SView'n'MTouch, (b) the user stereoscopically views through a transparent plane prop held in the non-dominant hand, (c) brings the dominant hand closer to the prop surface to trigger the animated flattening, and (d) touches the monoscopic projection of the object on the prop surface, with haptic feedback and without binocular parallax.

KEYWORDS

AR; remote interaction; haptic feedback; binocular parallax

ABSTRACT

In HMD-based immersive AR, interacting with virtual objects often requires indirect pointing methods such as ray casting. We present SView'n'MTouch, which brings parallax-free direct touching of remote objects, analogous to that of smartphone-based mobile AR, to HMD-based immersive AR, using a physical plane prop that is conceptually transparent. The stereoscopic viewing mode lets the user inspect the AR scene in full stereoscopic depth, while the monoscopic touching mode lets the user touch on anything that appears through the plane prop, with haptic feedback and without binocular parallax. Our animated flattening algorithm ensures smooth mode transitions. We show that a lightweight transparent plane prop can complement HMD-based immersive AR in the future.

INTRODUCTION

A small, flat, transparent panel—like the one Tony Stark carries with him in the *Iron Man* movies that he touches, types, and swipes on and holds up and interacts with distant objects through—has captured popular imagination for years (Figure 1). Such a device could serve as a handy interface to a larger interactive world in the near future, when HMD-based immersive AR becomes commonplace and the user's view of reality becomes seamlessly augmented with many virtual objects. These virtual objects can be simple annotations that the user can quickly glance at or more complex spatial UI elements that the user can interact with to accomplish complex spatial tasks, such as IoT [3] and robotics [4] programming and in-situ 3D sketching [1].

CHI'19 Extended Abstracts, May 4–9, 2019, Glasgow, Scotland UK

ACM ISBN 978-1-4503-5971-9/19/05.
https://doi.org/10.1145/3290607.3313259

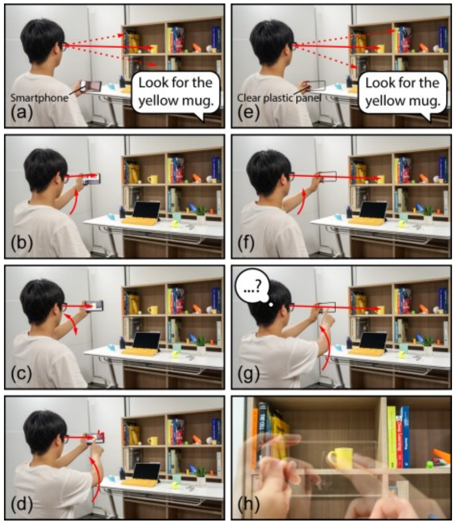

Figure 3. In our formative study, with a smartphone, (a) after fixating on the target, (b) the user raised the smartphone but (c) had to adjust its aim before (d) touching on its surface. With a clear plastic panel, (e) after fixating on the target, (f) the user raised the panel (g) while maintaining the fixation (h) but could not select unambiguously via touch due to binocular parallax.

Interacting with virtual objects via methods such as ray casting can be indirect and demanding, whereas interacting with them as they appear [8] via touch interactions through a transparent panel can be more direct and effortless [9]. However, when the user points a finger at an object through a transparent panel, the user must focus his or her two eyes on either the finger or the object, resulting in the other appearing as duplicate and overlapping images due to binocular parallax [5].

Interestingly, with the considerably less immersive smartphone-based mobile AR, where the user views only through the small screen, of which the perspective can be misaligned from the user's own [2], the user can directly touch on any virtual objects appearing on the screen, with haptic feedback and without binocular parallax.

We propose *SView'n'MTouch*, a novel interaction technique that brings parallax-free direct touching of remote objects, analogous to that of smartphone-based mobile AR, to HMD-based immersive AR (Figure 2). We use a lightweight plane prop that acts both as a clear plastic panel the user can see through stereoscopically (Figure 2b) and as a perspective-corrected, monoscopic viewport the user can directly touch on without being affected by binocular parallax (Figure 2d). Our animated flattening algorithm makes the transition seamless and comfortable to the user (Figure 2c).

FORMATIVE STUDY

How would a user immersed in an AR scene search for and select a virtual object through a touch surface, to reveal more information about it or to perform actions on it? We conducted a formative study to identify requirements of direct touch interactions with remote objects in immersive AR.

We imagined that our office was a reasonably crowded AR scene; the office supplies were interactive virtual objects; and a smartphone, with its default camera app open, and a clear smartphone-sized plastic panel were the touch surfaces. Then, we role played in pairs: One (instructor) chose an object and told the other (user) standing next to him to find and select it by touching it on the surface without telling the user where it was, e.g., "Look for the yellow mug!" The task was to be performed as quickly as possible, with no specific instruction as to how, to draw out natural behaviors.

Common Behaviors. The user did not search for a specific object through the small frame of the phone or the panel. Instead, the user fully utilized the wide field of view of both eyes, freely rotated the head, quickly glanced at different parts of the office, and then, upon finding the object, fixated both eyes on the target (Figure 3a, e) before lifting up the phone or the panel.

Smartphone Behaviors. When the user lifted the phone, the user had to shift focus suddenly from the object to the phone. The perspective mismatch between the image displayed on the phone and the rest of the user's sight made this jump perplexing (Figure 3b). Also, the aim of the phone was often imprecise, and the target appeared off-center, so the user had to adjust the phone to bring the target back to the center (Figure 3c) before touching it to complete the task (Figure 3d).

Transparent Panel Behaviors. In contrast, when the user lifted the transparent panel, the user did not have to shift focus or adjust the aim of the panel and could retain the binocular fixation on the target (Figure 3f). However, when the user tried to touch it while fixating on the target (Figure 3g), the user saw duplicate images of the finger due to binocular parallax (Figure 3h). This parallax made selection nearly impossible in cases where the objects were too small and/or too distant.

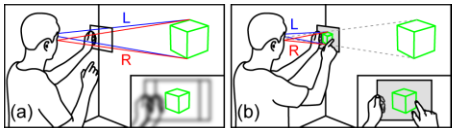

Figure 4. (a) When the user just holds the plane prop, it acts like a clear plastic panel, and the user can stereoscopically view through it, (b) and when the fingertip comes closer to the prop, it acts like a perspective-corrected, monoscopic viewport, and the user can touch it without experiencing binocular parallax.

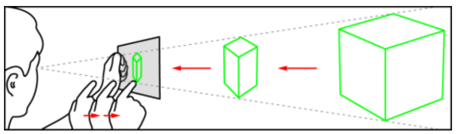

Figure 5. The object is gradually flattened and continuously brought closer to the prop surface as the fingertip approaches the surface of the plane prop.

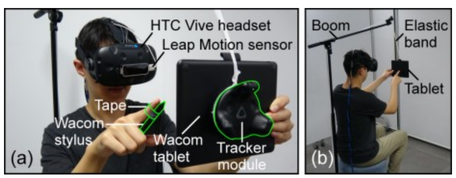

Figure 6. (a) In our proof-of-concept implementation, (b) the tablet was suspended from a boom to reduce its weight.

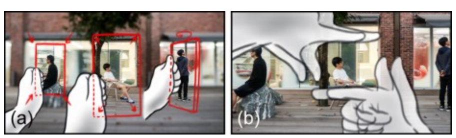

Figure 7. In the AR camera, the user can (a) move the plane prop farther to zoom in or closer to zoom out and tilt it to change the aspect ratio of the photo to be taken. (b) Doing so only with bare hand gestures can be affected by binocular parallax.

Here, one instinctive behavior we observed was the user closing one eye to eliminate the hindering parallax. However, this significantly reduced the user's field of view and disabled stereoscopic depth cues. We also found that some people cannot voluntarily close one eye, and even those who can do so find it very straining after a short while.

Our formative study indicated to us the following: First, a user should be able to freely glance at different parts of the AR scene outside the bounds of the touch surfaces, find the interaction target, and binocularly fixate on the target before taking actions on it. Second, the touch surface propped up in between the user's eyes and the target should not break visual continuity, so that the user can retain this fixation. Third, monoscopy, e.g., looking at the image captured by the rear-facing camera of the phone or closing one eye, eliminates the hindering effects of binocular parallax.

SVIEW'N'MTOUCH

SView'n'MTouch enables parallax-free direct touching of remote objects, analogous to that of smartphone-based mobile AR, in the context of HMD-based immersive AR, using a lightweight plane prop. In our technique, the plane prop transitions between two distinct modes: stereoscopic viewing and monoscopic touching (Figure 4). Our animated flattening algorithm (Figure 5) can make the transition between the two modes visually seamless and comfortable to the user.

Stereoscopic Viewing ("SView") Mode

When the user simply holds the plane prop with the non-dominant hand, the viewing mode is enabled. In this mode, the prop acts like a clear plastic panel, and the user can look through it stereoscopically (Figure 4a). Note that, in this mode, the prop is merely a transparent layer over the immersive AR scene, so virtual objects look the same whether they appear inside or outside the bounds of the prop, so a user can first stare at a distant target object with both eyes and then lift the prop to the eye level, all while not taking the eyes off the target.

Monoscopic Touching ("MTouch") Mode

When the user brings the fingertip of the dominant hand within a certain distance (e.g., 2 cm) from the surface of the plane prop, the touching mode is enabled. In this mode, the AR scene is inversely projected onto the prop surface in the perspective of the midpoint between the two eyes ("mid eye") [6]. In this regard, the prop acts like an opaque screen simulating transparency with a monoscopic, perspective-corrected image of the background [2], and the user can touch, swipe, and pinch on the prop surface without being affected by binocular parallax (Figure 4b).

Animated Flattening Transition

Even when the perspectives match, a jump between the stereoscopic and monoscopic vision requires a vergence depth shift that can be eye straining, so we animate the transition in between. In the *SView* mode, scene objects retain their original volumes, but during the transition to the *MTouch* mode, they are progressively flattened via homographic projection onto the prop surface (Figure 5). We map this animated flattening to the fingertip-to-prop distance so that objects are progressively flattened as the fingertip approaches the prop surface (e.g., within 8 cm) and are un-flattened as the fingertip retracts.

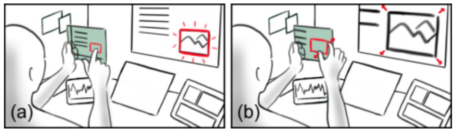

Figure 8. In AR web browsing, (a) the user can select small and distant 2D UI elements in 3D space (e.g., a hyperlinked picture) and (b) perform touch gestures (e.g., pinch to zoom).

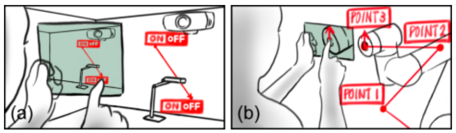

Figure 9. In AR IoT and robotics programming, (a) the user can draw lines between smart objects to define functional relationships (e.g., "turn off the desk lamp when the projector is on") and (b) control movements of a robot arm (e.g., "move through point 1, 2, and 3, and then go up"), while paying attention to both the target and the context.

Figure 10. In AR in-situ 3D sketching, (a) the user can perform perspective drawings in the monoscopic drawing mode (e.g., sketch a new door handle) and (b) inspect the 3D sketch and decide where to sketch next in the stereoscopic viewing mode (e.g., check if the new handle looks right in the context of the entire car sketch and then move on to the side mirror).

IMPLEMENTATION & APPLICATION

Our proof-of-concept implementation used an HTC Vive VR headset, a Vive tracker module, a Leap Motion hand-tracking sensor, and a Wacom Intuos S Bluetooth tablet (Figure 6a). The tracker module attached to the back of the tablet tracked its position and orientation. The hand-tracking sensor attached to the front of the VR headset detected the hand posture and the fingertip-to-tablet distance. The tablet stylus taped underneath the index finger precisely registered finger hover and touch events. The tablet was suspended from an overhead boom with an elastic band, which canceled some of the total weight (340 g) while also letting the user freely maneuver the tablet (Figure 6b).

However, in the near future, HMD-based immersive AR could become commonplace, and the scene geometry could be arbitrarily manipulated in real time [7], as needed in our technique. Then, an ultra-lightweight clear plastic panel with minimal electronic components, such as IR LED markers and a capacitive touch-sensitive film, can serve the same role as the tracked tablet. We envision 4 applications of our technique in that future, in illustrations: an AR camera, AR web browsing, AR IoT and robotics programming, and AR in-situ 3D sketching (Figures 7, 8, 9, and 10, respectively).

CONCLUSION

Transparency is an interesting material property in that it allows light, but not matter, to pass through. This makes a transparent plane prop an ideal remote haptic proxy that lets users see and also directly touch on anything that appears through its surface, in the HMD-based immersive AR.

Our motivation in this study was to take advantage of the metaphor of transparency in exploring the possibility of something comparable to direct input in space, for distant objects. We wished to get as close to "touching a distant object" as possible while also avoiding the issues associated with binocular parallax. We thus proposed an interaction technique, SView'n'MTouch, consisting of stereoscopic viewing, monoscopic touching, and animated flattening for smooth transitions.

REFERENCES

[1] R. Arora, R. Habib Kazi, T. Grossman, G. Fitzmaurice, & K. Singh. 2018. SymbiosisSketch: combining 2D & 3D sketching for designing detailed 3D objects in situ. In *Proc. CHI '18*, Paper 185, 15 pages.

[2] D. Baričević, C. Lee, M. Turk, T. Höllerer, & D. A. Bowman. 2012. A hand-held AR magic lens with user-perspective rendering. In *Proc. ISMAR '12*, 197-206.

[3] V. Heun, J. Hobin, & P. Maes. 2013. Reality editor: programming smarter objects. In *Proc. UbiComp '13*, 307-310.

[4] S. Kasahara, R. Niiyama, V. Heun, & H. Ishii. 2013. exTouch: spatially-aware embodied manipulation of actuated objects mediated by augmented reality. In *Proc. TEI '13*, 223-228.

[5] J. H. Lee & S. H. Bae. 2013. Binocular cursor: enabling selection on transparent displays troubled by binocular parallax. In *Proc. CHI '13*, 3169-3172.

[6] J. H. Lee, S. G. An, Y. Kim, & S. H. Bae. 2018. Projective windows: bringing windows in space to the fingertip. In *Proc. CHI '18*, Paper 218, 8 pages.

[7] D. Lindlbauer & A. D. Wilson. 2018. Remixed reality: manipulating space and time in augmented reality. In *Proc. CHI '18*, Paper 129, 13 pages.

[8] J. S. Pierce, A. S. Forsberg, M. J. Conway, S. Hong, R. C. Zeleznik, & M. R. Mine. 1997. Image plane interaction techniques in 3D immersive environments. In *Proc. I3D '97*, 39-ff.

[9] D. Schmalstieg, L. M. Encarnação, & Z. Szalavári. 1999. Using transparent props for interaction with the virtual table. In *Proc. I3D '99*, 147-153.