A Hand-Held, Self-Contained Simulated Transparent Display

Conference Paper · September 2016

DOI: 10.1109/ISMAR-Adjunct.2016.0049

Daniel Andersen
Purdue University

Chengyuan Lin
Purdue University

Maria Eugenia Cabrera
Purdue University

Aditya Shanghavi
Purdue University

A Hand-Held, Self-Contained Simulated Transparent Display

Daniel Andersen* Voicu Popescu Chengyuan Lin Maria Eugenia Cabrera Aditya Shanghavi
Juan Wachs

Purdue University

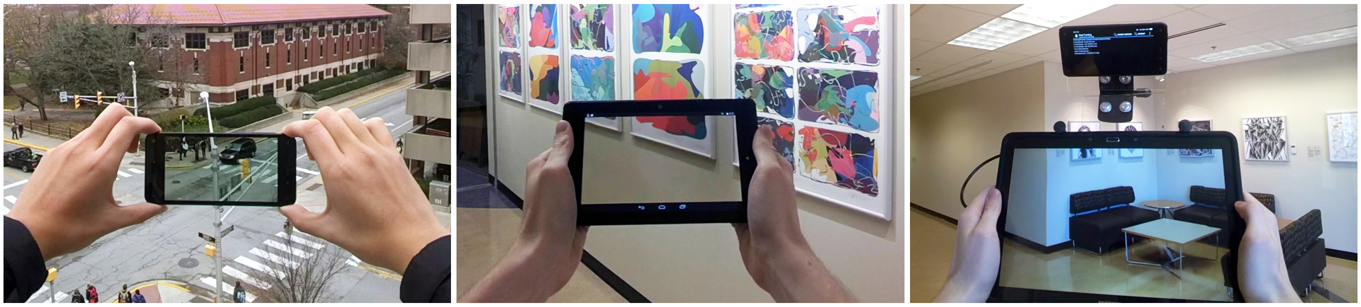

Figure 1: Actual first-person photographs of our three transparent display prototypes: P1 is compact and it adapts to the user's viewpoint, but the transparency effect is only accurate for scenes that are far away (left); P2 is compact and it enables accurate transparency even for nearby scenes, but it assumes a user viewpoint that is fixed with respect to the display (middle); P3 supports both nearby scenes and a changing user viewpoint by means of depth sensing and head tracking accessories (right).

Abstract

Hand-held transparent displays are important infrastructure for augmented reality applications. Truly transparent displays are not yet feasible in hand-held form, and a promising alternative is to simulate transparency by displaying the image the user would see if the display were not there. Previous simulated transparent displays have important limitations, such as being tethered to auxiliary workstations, requiring the user to wear obtrusive head-tracking devices, or lacking the depth acquisition support that is needed for an accurate transparency effect for close-range scenes.

We describe a general simulated transparent display and three prototype implementations (P1, P2, and P3), which take advantage of emerging mobile devices and accessories. P1 uses an off-the-shelf smartphone with built-in head-tracking support; P1 is compact and suitable for outdoor scenes, providing an accurate transparency effect for scene distances greater than 6m. P2 uses a tablet with a built-in depth camera; P2 is compact and suitable for short-distance indoor scenes, but the user has to hold the display in a fixed position. P3 uses a conventional tablet enhanced with on-board depth acquisition and head tracking accessories; P3 compensates for user head motion and provides accurate transparency even for close-range scenes. The prototypes are hand-held and self-contained, without the need of auxiliary workstations for computation.

Keywords: Simulated transparent smartphone, simulated transparent tablet, infrastructure for augmented reality applications.

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities

1 Introduction

The advent of powerful smartphone and tablet hand-held devices has opened the door to augmented reality (AR) applications that overlay information onto the real world without the need of eyewear. For example, an airplane maintenance technician holding a tablet in front of an open engine bay can receive guidance through graphical annotations. A new student on a university campus can use their smartphone to receive navigational cues overlaid onto the view of the campus. A tablet placed over an operating field can provide guidance to an overseas military trauma surgeon from an expert thousands of miles away.

Hand-held devices have high-resolution video cameras and displays that can capture and show to the user a high-quality live video feed of the real world scene. Augmenting this video feed with textual and graphical annotations provides the user with information about the scene that is easy to parse, because each piece of information is attached to the real world element that it describes. However, this augmented video feed is not adapted to the user viewpoint, which results in visual discontinuity and redundancy between the parts of the scene viewed directly and the parts viewed on the display. This places an additional cognitive load on the user who has to map the information given, from the context of the display, to the context of the scene observed directly. The user has to switch back and forth between the display and the direct view of the scene to translate the information received on the display to the real world. Furthermore, relying on the device-perspective view of the scene shown on the display can lead to incorrect depth interpretation and an inability to properly estimate distances [9].

This problem, known as the *dual-view perceptual issue*, has been the subject of extensive investigation to confirm its existence and to determine the effects of multiple perspectives on users of AR applications. It has been found that when users are presented with a user-perspective view of a scene, they have significantly improved spatial perception [20]. A recent study found that when users encounter the dual-view problem, their use of visual information from the part of the scene that is viewed directly (i.e. not viewed through the display) is significantly reduced [21]. This behavior was also

*e-mail:andersed@purdue.edu

seen in a user study in which users were timed while using a map navigation application with either a joystick, a dynamic peephole, or a device-perspective magic-lens navigation interface. Subjects who used the device-perspective magic-lens interface, despite being provided with additional visual context outside of the borders of the display, did not perform significantly faster at map navigation than users who lacked the additional visual context. It was concluded that this was because of the high cognitive cost of switching between multiple layers of visual presentation [12].

What is needed is a transparent display that lets the user see the real world as if they are looking through a window. With such a transparent display there is a perfect alignment between the parts of the scene viewed directly and the parts viewed on the display. The transparent display enables integrating the AR annotations seamlessly into the field of view of the user.

One approach is to develop truly transparent tablets and smartphones. Large OLEDs (i.e. 55”) with transparency levels of 40% have been developed [14], but letting more light pass through is a substantial technological challenge. Moreover, porting the transparent display technology to the self-contained form of a tablet or smartphone requires drastic miniaturization to hide the opaque components of the hand-held device such as the battery and CPU.

Another approach is to *simulate* the transparency of the display by reprojecting the video feed acquired by the device to the user’s viewpoint. This way, the image on screen appears aligned with the real-world view outside the screen’s borders, making the display appear transparent.

In this paper we describe a hand-held self-contained simulated transparent display, with three prototype implementations. Simulating a transparent display requires acquiring the color and geometry of the parts of the scene viewed through the display, tracking the user’s head position, and rendering the color and geometry data from the user’s viewpoint. Color acquisition, 3D rendering, and display are solved problems since modern tablets and smartphones have capable back-facing cameras, GPUs, and LCDs. Geometry acquisition and user head tracking are capabilities that are beginning to appear in hand-held devices.

Our first prototype, *P1*, is based on a smartphone with built-in head tracking support. P1 is compact, it adapts to the user’s viewpoint, and it is suitable for outdoor scenes (Fig. 1, left). However, P1 does not have geometry acquisition capability, so an accurate transparency effect relies on the assumption that the scene is far away (i.e. beyond 6m).

Our second prototype, *P2*, is based on a tablet with built-in depth acquisition capability. P2 is compact, and it can provide an accurate transparency effect even when the scene is nearby (Fig. 1, middle). However, P2 does not have user tracking capability, so the user has to hold the display at a fixed position relative to their head. Moreover, P2 acquires depth with an active approach which restricts its use to indoor scenes.

Our third prototype, *P3*, is based on a large, conventional tablet with an attached depth camera and an attached head tracker. P3 is a complete transparent display system that supports nearby geometry as well as user head motion (Fig. 1, right). Like for P2, active depth acquisition restricts P3’s use to indoor scenes.The head tracking and geometry acquisition attachments make P3 less compact and less maneuverable than the other two prototypes. We foresee that future tablets and smartphones will integrate both depth acquisition and head tracking capabilities, which will provide compact platforms for the complete transparent display pipeline demonstrated by P3.

For all three prototypes all acquisition and computation is performed on board the transparent display system, without the aid of any auxiliary workstation, which is an important prerequisite for practical deployment in the context of mobile AR applications. We also refer the reader to the accompanying video. The first person illustrations of the transparent display effect shown in this paper and in the video were captured by having the user wear a head mounted camera (i.e. Google Glass). Wearing the camera was only necessary in order to capture the first-person illustrative footage, and it is not needed during normal use of our transparent displays.

2 Prior Work

The ARScope is an early simulated transparent display system [22]. The user holds an opaque surface like they would hold a magnifying glass. The surface is made to appear transparent to the user by projecting on it an image that approximates what the user would see in the absence of the surface, using a projector mounted on the user’s head. The system works by acquiring the scene with two cameras, one mounted on the user’s head, and one attached to the hand-held surface. The system computes a homography between the two acquired images based on matching features, the homography is used to warp the hand-held camera’s image to the user’s viewpoint, and the warped image is projected onto the hand-held surface using the head-mounted projector. This early system demonstrates simulated transparency, but it suffers from limitations such as the reliance on encumbering head-mounted cameras and projectors, the reliance on a nearby workstation, user tethering, and severely simplifying assumptions about scene geometry.

The passive surface requiring external projection to make it seem transparent was later replaced with an LCD that can display the image needed to simulate transparency. We inventory prior LCD-based simulated transparent display systems according to how they track the user’s head, to how they acquire the scene geometry, and to whether or not they are tethered to a nearby workstation to which they off-load computation.

Some systems track the user’s head with a head-mounted sensor [3, 13]. The encumbrance of a head-mounted device is avoided by tracking the user with a camera attached to the display [8, 2, 15, 16, 17, 18, 19, 23, 10, 7].

Several systems assume that the scene is planar, which only requires registering the position and orientation of the display with respect to the scene plane. Registration is performed by using a manually-set depth [23, 10], by using markers placed in the scene [8, 17, 19, 7, 13], or by using features detected in the scene image [15, 16]. Other systems do not assume scene planarity and they acquire scene geometry actively using on-board depth cameras [3, 18], or passively from the scene video frames [2].

Recent systems attempt to break free from tethering the display to a nearby workstation. Such self-contained systems take advantage of the general-purpose and graphics computing capabilities of modern tablets and smartphones to perform all computation onboard [19, 23, 10, 7, 13].

Our prototypes advance the state of the art in simulated transparent displays as follows. P3 is the first untethered transparent display system that is complete, i.e. that performs unobtrusive head tracking, and that acquires scene geometry; all prior complete transparent displays are tethered [18, 2]. P2 is the first transparent display system that uses integrated active depth acquisition; prior systems that use active depth acquisition are tethered *and* they rely on an attached depth camera [3, 18]. P1 is the first untethered transparent display system that uses integrated multi-camera head position tracking; the user’s head position is triangulated using two cameras which improves z-tracking accuracy compared to prior systems that use a single camera [7].

3 Simulated Transparent Display

To simulate a transparent display using a conventional LCD, one has to display the image that the user would see in the absence of the display. The part of the scene obstructed by the LCD has to be captured with a camera. Placing the camera at the user’s viewpoint is not beneficial because the camera’s view would also be obstructed by the LCD, in addition to the disadvantage of the user having to

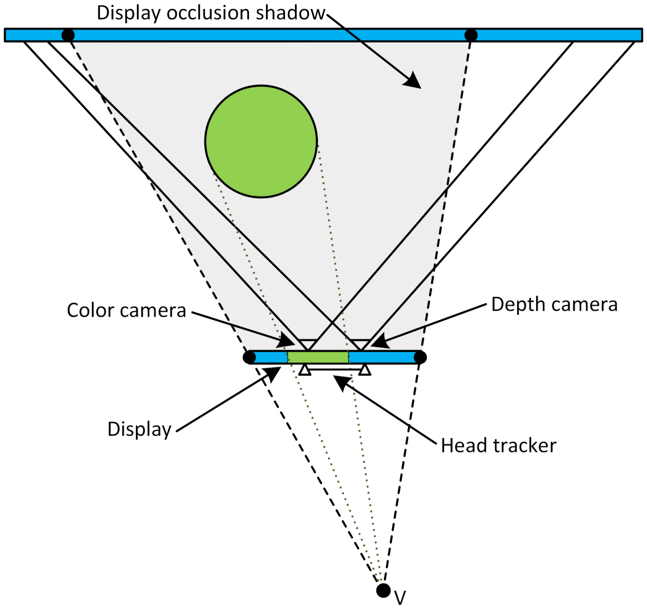

Figure 2: Overview of simulated transparent display.

wear the camera. Consequently, the camera has to be placed at a different viewpoint, beyond the LCD, such that the scene is captured without occlusions. The frame captured by the camera has to be reprojected to the user's viewpoint, which requires knowledge of scene geometry. In Fig. 2, the parts of the scene in the display occlusion shadow are acquired with a color camera and a depth camera. The user's viewpoint is acquired with a tracker that triangulates the position of the user's head. The color and depth data is rendered from the user's viewpoint to simulate transparency.

We implement the simulated transparent display as follows. Depth is acquired with a real-time on-board depth camera. The current depth frame is triangulated. For depth cameras that acquire a sparse set of depth values we use a 2-D Delaunay triangulation in the depth image plane. For depth cameras that acquires a dense set of depth values we build a complete regular depth map by filling in holes using a pull-push approach [6], and we triangulate using the implicit connectivity of the regular depth map. Color is acquired with an on-board color camera. The depth and color cameras are fixed with respect to each other and their relative position and orientation is pre-calibrated using a checkerboard pattern visible to both cameras. The color frame is used to projectively texture map the triangle mesh. The user's head position is acquired with an on-board tracker. The textured mesh is rendered from the user viewpoint on the mobile device's GPU to simulate transparency.

4 Implementation and Results

We pursue the implementation of the simulated transparent display pipeline in a compact form factor without wires or the need of auxiliary workstations. Tablet and smartphone platforms now have high resolution video cameras and display. At the time of this writing there is no tablet or smartphone that combines all capabilities needed for a complete implementation of the simulated transparent display pipeline. Emerging platforms provide either user head tracking or depth acquisition. We have implemented three prototypes. P1 takes advantage of a smartphone with integrated head tracking capability. P2 takes advantage of a tablet with integrated depth acquisition. P3 implements the simulated transparent display pipeline using a conventional tablet enhanced with on-board accessories for head tracking and depth acquisition.

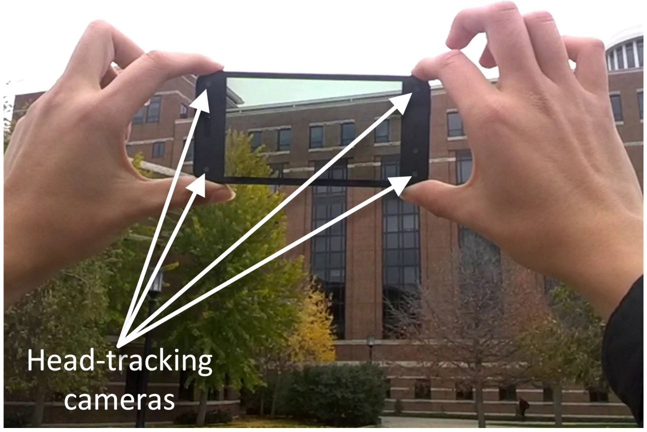

Figure 3: Prototype P1.

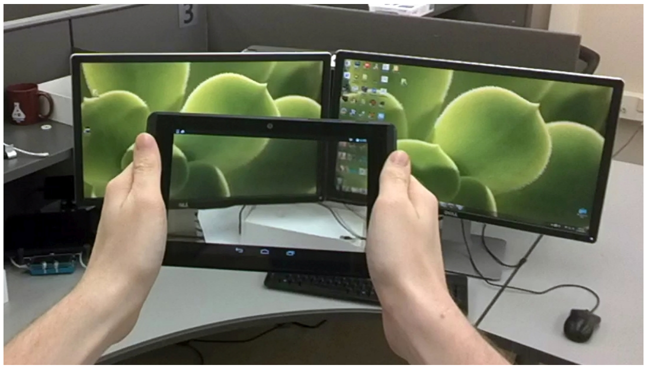

Figure 4: Prototype P2.

4.1 Prototype P1

P1 leverages Amazon's Fire Phone [1], a 4.7-inch smartphone with four front-facing cameras dedicated to tracking the user's head (Fig. 3). The device has four cameras to increase the chance that at least two of them have a good view of the user's head, free of finger occlusion. The head position is triangulated from the frames of the two cameras that provide the best view of the user's head. Compared to tracking the user's head with a single camera, triangulation has the advantage of better z tracking accuracy. The Fire Phone API provides a tracking frame rate of up to 100Hz. The Fire Phone does not have depth acquisition capability. We compute the transparency effect for P1 under the assumption that the scene is infinitely far away, an assumption that is reasonable for outdoor scenes. As discussed in Section 4.4, the transparency effect error is below 5% when the scene is farther than 6m. As can be seen in Fig. 1, left, in Fig. 3, and in the accompanying video, P1 is well suited for outdoor scenes. P1 is very compact and portable, which readily supports driving and walking navigation assistance applications. Although active depth acquisition outdoors is not yet practical for smartphone-like devices, the benefit from depth acquisition would be small since the scene is typically away from the user.

4.2 Prototype P2

P2 leverages Google's Project Tango Tablet [5], a 7-inch tablet with integrated depth from structured light acquisition (Fig. 4). The Project Tango tablet acquires approximately 10,000 depth points per frame at 3Hz. The depth data is too sparse to attempt to re-

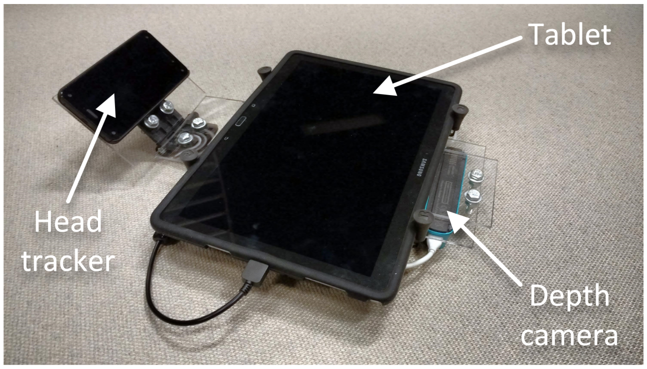

Figure 5: The hardware components of P3.

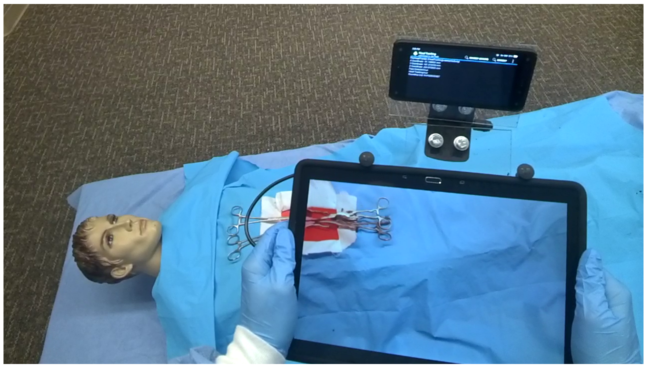

Figure 6: Prototype P3.

construct a complete regular depth map and we rely on Delaunay triangulation instead. The Project Tango tablet does not provide user head tracking, so the user has to hold P2 at a fixed position and orientation with respect to their head. We pre-calibrate the natural position and orientation for the tablet, for each user. A user walking around with the transparent display tends to keep it roughly in the same spot. If the tablet moves out of position, the user can intuitively reposition to restore the transparency effect by aligning the displayed image to the surrounding scene.

The Project Tango tablet does have a front-facing camera (i.e. facing the user), which could be used to track the user's head. However, the API does not allow turning on the front-facing camera while the back-facing camera is in use, presumably in order to accommodate for a hardware limitation. This problem was also noted in previous work [23, 10]. Furthermore, tracking the user head with a single camera would provide poor accuracy in z. P2 is most useful for close-range indoor scenes, and where compactness is more important than a large display size.

4.3 Prototype P3

P3 is built around a conventional tablet (Fig. 5). We use Samsung's 12.2-inch Galaxy Tab Pro tablet for its large screen size, but any tablet could be swapped in. Depth acquisition is provided by a Structure Sensor [11], which is a compact structured light depth camera accessory connected to the tablet via a USB interface. The Structure Sensor acquires 160 x 120 resolution depth frames at 30Hz. Head tracking is provided by a Fire Phone [1] rigidly attached to the tablet. The head position is sent to the tablet via a Bluetooth wireless connection at 100Hz. The depth camera and the head tracker are registered in the tablet's coordinate system using a calibration procedure that shows a black and white checkerboard to the the depth camera and the tablet's back-facing camera, and then to the front-facing cameras of both the tablet and the Fire Phone. P3 is most useful in indoor environments, and for applications where the user moves with respect to the display (Fig. 6). A possible use is in a surgical telementoring scenario where the tablet is suspended into the field of view of a trainee surgeon and enhances the operating field with graphical annotations from a remote mentor.

4.4 Quality of transparent display effect

Perfect transparency requires displaying exactly what the user would see if the display were not there. We analyze the quality of the transparency achieved by our prototypes both theoretically and empirically. We define the transparency error ε at a point p on the simulated transparent display as

$$\varepsilon = \left\| p - p^0 \right\| / d \tag{1}$$

The numerator is the distance on screen between the actual position p and the correct position p^0 of the scene 3D point imaged at p, and d is the display's diagonal length.

Because P1 assumes that scene geometry is infinitely far away from the display, the transparency error is only low when the scene geometry is far from the display. The maximum transparency error for P1 that results from the infinite scene depth assumption is

$$\varepsilon = \frac{z_{vd}}{2 z_{ds}} \tag{2}$$

where z_{vd} is the distance in z between the viewpoint and the display (we assume z_{vd} = 0.5m), and z_{ds} is the actual distance in z between the display and the scene. The transparency error is less than 5% when the display-to-scene distance is greater than 6m.

P2 and P3 use active, structured-light depth acquisition, which is not always accurate. Moreover, missing depth data is approximated by triangulation (P2) or by pull-push (P3). Given a depth acquisition error Δ_{depth}, the maximum transparency error for P2 and P3 as a function of depth acquisition error is

$$\varepsilon = \frac{1}{2} \left| \frac{z_{vd} + z_{ds}}{z_{vd} + z_{ds} + \Delta_{depth}} - 1 \right| \tag{3}$$

In a typical use case for P2 and P3, the scene is assumed to be 1m away behind the display ($z_{ds} \approx$ 1m), and the user viewpoint is assumed to be 0.5m away in front of the display ($z_{vd} \approx$ 0.5m). Typical depth acquisition errors for P2 and P3 are in the 10mm range, which corresponds to a low transparency error of 3%.

Error in head position tracking increases transparency error as the display renders the scene imagery for an invalid viewpoint. The amount of transparency error differs depending on whether the tracking error is in a direction parallel to the display (in x or y), or in a direction normal to the display (in z). Given a head tracking error in x Δ_x, the maximum transparency error is

$$\varepsilon = \frac{z_{ds} \Delta_x}{d(z_{vd} z_{ds})} \tag{4}$$

Given a head tracking error in z Δ_z, the maximum transparency error is

$$\varepsilon = \frac{z_{ds} \Delta_z}{2 z_{vd} (z_{vd} + \Delta_z + z_{ds})} \tag{5}$$

Here we assume an infinite display-to-scene distance for P1, a display-to-scene distance of 1m for P3, and a viewpoint-to-display distance of 0.5m for both P1 and P3. Head tracking in P1 and P3 is typically accurate to less than 10mm in x and 30mm in z. Within these bounds, the maximum transparency error from head tracking in x is 8.4% for P1 and 2.2% for P3, and the maximum transparency error from head tracking in z is 3.0% for P1 and 2.0% for P3.

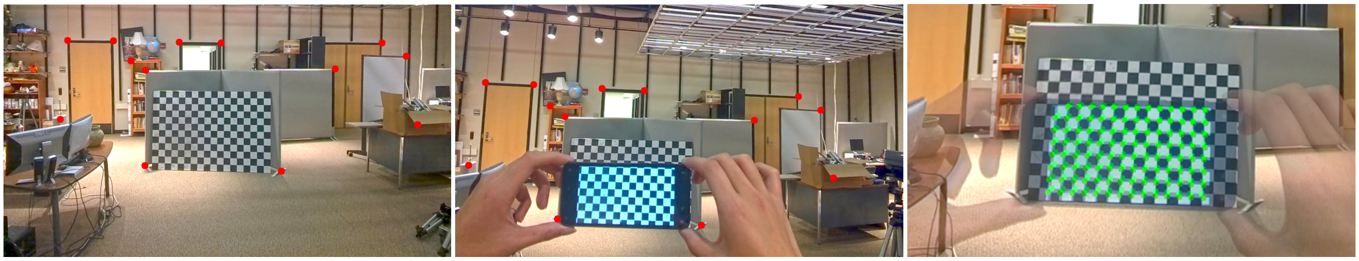

Figure 7: Empirical transparency error measurement. Left: Reference image of the scene taken by Google Glass. Middle: Image taken by Google Glass while using the transparent display. The red dots illustrate manually selected salient features in the region outside of the transparent display, which are used to align the two images. Right: Overlay image where the actual transparency error is measured using manually selected correspondences (green dots) in the region covered by the transparent display.

Table 1: Empirical transparency errors for our three simulated transparent display prototypes, measured as the mean transparency error across all tracked features in the scene image.

Prototype	Scene	Transparency error ε [%]
P1	Fig. 3	1.2
	Fig. 1, left	1.7
P2	Fig. 1, middle	2.3
	Fig. 4	5.4
P3	Fig. 6	3.1
	Fig. 1, right	1.6

All first person images shown in this paper and accompanying video were taken by having the user wear the Google Glass head mounted camera [4]. In addition to their illustrative purpose, we also use these first person images to estimate the transparency error empirically, as shown in Figure 7. First, the user acquires a scene image I_1 using the Google Glass camera (Fig. 7, left). Next, the user acquires a second scene image I_2 while holding up the simulated transparent display, which has been calibrated to generate a transparent effect for the viewpoint of the Google Class camera (Fig. 7, middle). Since the user is likely to tilt their head slightly as they acquire the two images, I_1 and I_2 have to be first aligned using the region outside the transparent display. We align the two images by computing a homography between I_1 and I_2 using manually selected corresponding salient features in the region outside the display. The homography is used to compute an overlaid image I_3 (Fig. 7, right). The transparency error is then computed by measuring the distance between manually selected corresponding features in I_3 that are within the transparent display region. Table 1 gives actual transparency error values for our prototypes, measured as the mean transparency error of all the feature points corresponding between images I_1 and I_2. These empirical results show that our prototypes achieve a good transparency effect. The small error values indicate that the actual head tracking errors are smaller than the upper bounds used in the theoretical analysis above.

4.5 Frame rate and latency

As objects in the scene move and as the user's head moves with respect to the display, the transparent effect must be recomputed to match the current configuration. There is a delay between when the scene or head position changes and when the transparent effect is reestablished. The latency is due to delays accumulated in the color acquisition, depth acquisition, depth hole filling, triangulation, head tracking, head tracking communication, and rendering. Color is acquired by the on-board color camera at 30 Hz for P1, P2, and P3. Depth is acquired at 3Hz and 30Hz for P2 and P3. Depth hole filling takes 6ms for P3, on average. Delaunay triangulation takes 56ms for P2, on average. The Fire Phone tracks the user's head at 100Hz. Bluetooth head tracking data communication latency is 30ms. Rendering takes 3ms (P1), 14ms (P2), and 43ms (P3).

The average latency for our three simulated transparent display prototypes is 120ms (P1), 144ms (P2), and 172ms (P3). The latency was measured using the Google Glass first person video feed by counting the number of frames it takes to the transparency effect to converge after a change in the scene or in the user's head position occurs. We have also measured the latency of displaying a video frame as it is acquired, without any processing. For the Fire Phone, the Project Tango tablet, and the Samsung Galaxy Tab Pro devices that underlie our three prototypes, this acquire-and-display latency is 114ms, 100ms, and 125ms. Consequently, most latency in our prototypes comes from the latency with which the devices display the images they acquire. This indicates that, in addition to integrating depth sensing and user head tracking capabilities into next generation tablets and smartphones, portable device manufacturers should pursue improving support for AR applications by also reducing the acquire-and-display latency of their devices.

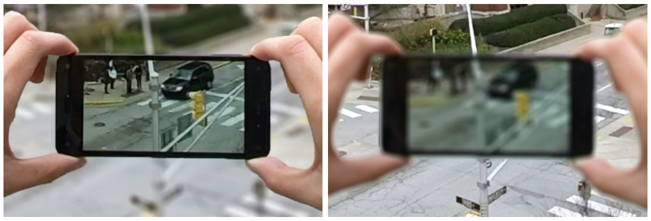

Figure 8: Illustration of the user's perceived disparity between focusing on the display (left) and focusing on the scene (right).

4.6 Limitations

P1 does not acquire scene geometry, so an accurate transparent effect requires the scene to be far away. P2 does not acquire the user's head position so the display has to be kept in a fixed position and orientation with respect to the user. P3 is not as compact since the tablet is enhanced with accessories for depth acquisition and user head tracking. All displays exhibit latency. The depth data is not always complete and filling in holes is only approximate.

For correct transparency, both the color and depth of all scene points that are occluded by the display should be captured. When the user viewpoint is too close to the display, the resulting occlusion frustum has a field of view that is greater than the fields of view of the color and depth cameras. Moreover, the user viewpoint can expose scene points that are not captured due to occlusions. In this work we did not disconnect scene geometry frame mesh at discontinuities, which results in stretching artifacts that we found preferable to the tearing artifacts that results when the frame mesh is disconnected. One possible solution is to accumulate a complete geometric model over several frames as the display moves.

Our transparent displays cater only to a single viewpoint (e.g. one eye or the midpoint between the eyes). The lack of disparity between the images shown to each eye gives the user an unwanted perception of diplopia or double vision. This is a common issue in AR systems, where the user's eyes converge at the depth of either the display or the background scene [21]. Future implementations of transparent displays could overcome this limitation by using autostereoscopic displays and rendering the scene for each eye.

When the scene is far from the display and the display is close to the user, the user cannot simultaneously focus on the scene and the display. Either the display appears in focus while the scene appears blurry, or vice versa. The first person images used in our analysis were taken using a camera with a small aperture, which does not show this limitation, and so we have created an artist rendition through image post-processing to illustrate this effect (Fig. 8).

In principle, our approach does not require the display to be held in a fronto-parallel view with respect to the user viewpoint, given that the tracked eye position is accurate. However, in P1 and P3, the Fire Phone's head tracker only provides the user's head position (i.e. a point between the eyes), not the actual eye positions. We apply a horizontal transformation to the tracked head position to estimate the user's eye position. In these cases, the estimated eye position will remain valid so long as the display is only rotated about the display's x axis (i.e. tilted up or down with respect to the user), which is typical for smartphone and tablet use.

5 Conclusions and future work

We have demonstrated the feasibility of a simulated transparent display that is completely self-contained and untethered. The user does not have to wear any sensors and the scene does not have to be enhanced with markers. We have developed three prototypes that take advantage of emerging mobile platforms. We believe that depth acquisition and user tracking will be commonplace in upcoming mobile devices, in support of powerful AR applications. In addition to improving our simulated transparent displays to alleviate the limitations discussed above, future work also includes using the transparent displays in actual AR applications such as car and pedestrian navigation assistance and surgical telementoring.

We have provided a quantitative analysis of transparency error in a simulated transparent display, and we have defined a metric that allows comparison between current and future simulated transparent displays. Additional future work will be to conduct human perceptual studies to determine the threshold of transparency error that is noticeable to the human eye and acceptable to users when performing specific tasks using AR applications.

Acknowledgements

We thank the STAR research group at Purdue University and at the Indiana University School of Medicine, and the computer graphics group at the computer science department of Purdue University for their feedback regarding our work.

This work was supported by the Office of the Assistant Secretary of Defense for Health Affairs under Award No. W81XWH-14-1-0042. Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the Department of Defense.

References

[1] Amazon.com. Amazon fire phone. http://www.amazon.com/firephone, 2014.

[2] D. Baricevic, T. Hollerer, P. Sen, and M. Turk. User-perspective augmented reality magic lens from gradients. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology*, pages 87–96. ACM, 2014.

[3] D. Baricevic, C. Lee, M. Turk, T. Hollerer, and D. Bowman. A hand-held ar magic lens with user-perspective rendering. In *Mixed and Augmented Reality (ISMAR), 2012 IEEE International Symposium on*, pages 197–206. IEEE, 2012.

[4] Google. Glass press. https://sites.google.com/site/glasscomms/, 2014.

[5] Google. Project tango. https://www.google.com/atap/project-tango/, 2015.

[6] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen. The lumigraph. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 43–54. ACM, 1996.

[7] J. Grubert, H. Seichter, and D. Schmalstieg. Towards user perspective augmented reality for public displays. In *Proc. ISMAR*, 2014.

[8] A. Hill, J. Schiefer, J. Wilson, B. Davidson, M. Gandy, and B. MacIntyre. Virtual transparency: Introducing parallax view into video see-through ar. In *Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium on*, pages 239–240. IEEE, 2011.

[9] E. Kruijff, J. E. Swan II, and S. Feiner. Perceptual issues in augmented reality revisited. In *ISMAR*, volume 9, pages 3–12, 2010.

[10] Y. Matsuda, F. Shibata, A. Kimura, and H. Tamura. Poster: Creating a user-specific perspective view for mobile mixed reality systems on smartphones. In *3D User Interfaces (3DUI), 2013 IEEE Symposium on*, pages 157–158. IEEE, 2013.

[11] Occipital. Structure sensor - 3d scanning, augmented reality, and more for mobile devices. http://structure.io/, 2015.

[12] M. Rohs, J. Schöning, M. Raubal, G. Essl, and A. Krüger. Map navigation with mobile devices: Virtual versus physical movement with and without visual context. In *Proceedings of the 9th International Conference on Multimodal Interfaces*, ICMI '07, pages 146–153, New York, NY, USA, 2007. ACM.

[13] A. Samini and K. L. Palmerius. A perspective geometry approach to user-perspective rendering in hand-held video see-through augmented reality. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology*, pages 207–208. ACM, 2014.

[14] Samsung. Samsung display introduces first mirror and transparent oled display panels. *Business Wire*, June 2015.

[15] M. Tomioka, S. Ikeda, and K. Sato. Approximated user-perspective rendering in tablet-based augmented reality. In *Mixed and Augmented Reality (ISMAR), 2013 IEEE International Symposium on*, pages 21–28. IEEE, 2013.

[16] M. Tomioka, S. Ikeda, and K. Sato. Pseudo-transparent tablet based on 3d feature tracking. In *Proceedings of the 5th Augmented Human International Conference*, page 52. ACM, 2014.

[17] H. Uchida and T. Komuro. Geometrically consistent mobile ar for 3d interaction. In *Proceedings of the 4th Augmented Human International Conference*, pages 229–230. ACM, 2013.

[18] Y. Unuma, T. Niikura, and T. Komuro. See-through mobile ar system for natural 3d interaction. In *Proceedings of the companion publication of the 19th international conference on Intelligent User Interfaces*, pages 17–20. ACM, 2014.

[19] K. Čopič Pucihar, P. Coulton, and J. Alexander. Creating a stereoscopic magic-lens to improve depth perception in handheld augmented reality. In *Proceedings of the 15th International Conference on Human-computer Interaction with Mobile Devices and Services*, MobileHCI '13, pages 448–451, New York, NY, USA, 2013. ACM.

[20] K. Čopič Pucihar, P. Coulton, and J. Alexander. Evaluating dual-view perceptual issues in handheld augmented reality: Device vs. user perspective rendering. In *Proceedings of the 15th ACM on International Conference on Multimodal Interaction*, ICMI '13, pages 381–388, New York, NY, USA, 2013. ACM.

[21] K. Čopič Pucihar, P. Coulton, and J. Alexander. The use of surrounding visual context in handheld ar: Device vs. user perspective rendering. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 197–206, New York, NY, USA, 2014. ACM.

[22] T. Yoshida, S. Kuroki, H. Nii, N. Kawakami, and S. Tachi. Arscope. In *ACM SIGGRAPH 2008 new tech demos*, page 4. ACM, 2008.

[23] E. Zhang, H. Saito, and F. de Sorbier. From smartphone to virtual window. In *Multimedia and Expo Workshops (ICMEW), 2013 IEEE International Conference on*, pages 1–6. IEEE, 2013.