Making Augmented Reality a Reality

Ronald T. Azuma
Intel Labs, 2200 Mission College Blvd., Santa Clara, CA 95054
ronald.azuma@gmail.com

Abstract: Augmented Reality has attracted interest for its potential as a platform for new compelling usages. This paper provides an overview of technical challenges in imaging and optics encountered in near-eye optical see-through AR display systems.
OCIS codes: (000.4930) Other topics of general interest; (000.2170) Equipment and techniques

1. Motivation

Augmented Reality (AR) is an immersive experience that superimposes virtual 3D objects upon a user's direct view of the surrounding real environment, generating the illusion that those virtual objects exist in that space. While Virtual Reality (VR) completely replaces the user's view of the real world, AR supplements it [1]. In the long term, AR potentially has a much larger market than VR, because it improves the user's understanding of and interaction with the real world. AR connects users to the people, locations and objects around them, rather than cutting them off from the surrounding environment. AR is the most likely route by which wearable systems replace smartphones, because of its potential to provide a large visual display in a compact, head-worn form factor. Ultimately, head-worn AR displays might replace all other display form factors: monitors, laptop screens, phones and tablets.

However, fulfilling this vision requires numerous advancements, both experiential and technical. We must develop new types of compelling experiences while also solving key problems in imaging, optics and displays.

2. New forms of media

AR will succeed in the consumer market sooner if it can establish new forms of media that users find compelling. The vision of head-worn AR displays that replace smartphones requires many technical advancements that likely will take several years to achieve. The success of Pokémon Go in 2016 proved that an AR experience can achieve mass market success even without head-worn AR display systems [2]. My hypothesis is that AR will enable new forms of media by making meaningful connections between the virtual content and the surrounding real environment, generating experiences where the power comes from that connection and not solely from the virtual or solely from the real world. Specifically, I proposed three strategies which I call *Reinforcing*, *Reskinning* and *Remembering*. For more details, see my book chapter [3], but I also briefly summarize them here.

Reinforcing is the strategy of leveraging a real location that is inherently meaningful and supplementing it with virtual content appropriate to that location. My favorite example of this is 110 Stories. For users in or near Manhattan, it renders an outline of where the Twin Towers should be, and invites users to write a few sentences explaining what that image means to them. Though the augmentation is simple, just a few lines sketched against an empty sky, the resulting combination is poignant and powerful [4].

Reskinning is a strategy that embellishes mundane real locations with compelling virtual content that a user already knows and loves, changing the real world to fit the needs of the virtual content and story. Pokémon Go is unquestionably the most successful example of this strategy, motivating millions to explore cities and towns. Users discover that monuments, malls and other points of interest were recast into supply depots and places to engage in duels with the Pokémon they collected. At CES 2014, Intel generated an AR experience of a giant flying whale from the fantasy series Leviathan, making it appear to fly over the heads of a live audience of 2,500 [5].

Remembering is an approach recognizing that while most of the real world may be mundane, the personal memories tied to specific locations are not. Three Angry Men is an experimental experience from Georgia Tech that enabled a user wearing an AR display to experience the memories of three jurors [6]. By sitting in one of three chairs at a table, the perceived the deliberations from the point of view of the juror assigned to that chair. The user hears that juror's inner thoughts and hears and sees what he experiences. At any point, the user can stand up, pausing the story, and move to another chair. The experience then changes to conform to the perceptions and biases of the juror assigned to that chair. Three Angry Men brings memories to life through a vivid, interactive experience.

Figure 1: Left: 110 Stories (courtesy Brian August), Center: Leviathan (Source: Intel), Right: Three Angry Men (Source: Georgia Tech)

3. Challenges in imaging

Connecting AR experiences in a meaningful way to the real world means that AR systems must semantically understand the surrounding environment. Most existing AR systems have very little semantic understanding of the real world, which limits the usefulness and richness of such systems. While AR tracking systems have advanced to the point of being able to use computer vision to recover 6D pose estimates of the viewer's position and orientation, enabling accurate 3D augmentations in environments that were not previously modelled or scanned, such systems generally only recover the geometry and not the meaning of the objects in the surrounding environment. Overcoming this limitation requires developing *smart imaging* systems that go beyond simply capturing arrays of pixels by also attaching semantic classifications to groups of pixels.

Depth sensing is a key capability to help make smart imaging feasible. Through a variety of sensing technologies, RGBD imagery adds a depth channel to the usual RGB images. Google Tango and Intel RealSense™ are two examples of nascent platforms that provide RGBD sensing. Processing RGBD data with a variety of computer vision and AI techniques to perform semantic recognition is an active area of research.

4. Challenges in optics and displays

For most of the history of AR systems, we had to build AR applications on equipment that was designed and built for other reasons, such as PC's with webcams, smartphones and tablets. Exploiting the investments that made such platforms available means that some types of AR experiences are available to everyone, but walking around while holding up a phone or tablet display in front of your face is hardly an ideal usage scenario. More recent investments, including the Meta 2 display, Microsoft's HoloLens, and DAQRI's Smart Helmet, are generating emerging platforms of head-worn optical see-through displays specifically designed for AR usages.

Optical see-through displays use optical combiners to add virtual content over a user's direct view of the surrounding environment. Since AR usages may occur outdoors and in everyday, unprepared environments, future AR display platforms focus on optical see-through for two reasons. First, the direct view of the real world avoids distorting or reducing the user's view of the real environment. Second, if power is lost, an optical display does not blind the user, unlike a video see-through approach [1].

Providing idealized head-worn optical see-through displays that can become replacements for smartphones requires overcoming numerous and formidable challenges. Some of the major challenges are:

Form factor: Ideally, head-worn displays would resemble sunglasses and eyewear that are already socially acceptable today. Shrinking and powering displays while fitting into this form factor is a daunting challenge. In particular, attractive eyewear often incorporates curved lenses. Many displays, particularly waveguide-based displays, have flat form factors.

Collaboration: An underrated issue is the interaction of other people with a person wearing a head-worn display. The display should support high transmission of light from the surrounding environment to the wearer's eyes, so that he or she can see the environment. But high transmission of light through the combiner is also important so that other people in the environment can clearly see the wearer's eyes. Many optical combiners route some of the displayed light away from the wearer's eyes, out toward other people in the environment. This light emission can obscure the wearer's eyes. A head-worn display that looks like dark sunglasses or emits bright light outwards is intimidating and an impediment to social collaboration.

Brightness and contrast: Optical see-through displays worn outdoors must be visible even in bright, sunny conditions. Most display technologies are too dim to be visible in daylight conditions. Another problem is that optical see-through displays typically can only add light to the direct view of the real world; they can't selectively block light coming from the real world. This lack of pixel-accurate occlusion reduces the contrast and makes the virtual augmentations appear like transparent ghosts. Research systems have demonstrated optical see-through displays that can selectively occlude the real world, but implementing those in a compact optical path is difficult [7].

Resolution: With today's displays, we can either provide a small field-of-view (FOV) with adequate spatial resolution, or a wide FOV with low resolution. In head-worn displays, we measure spatial resolution in pixels per degree (PPD) rather than pixels per inch (PPI). The required resolution is dictated by the resolution of the fovea. To appear sharp, the spatial resolution should be 60 PPD. Achieving a "retinal display" would require 120 PPD. Providing such resolution across a wide FOV requires microdisplays with thousands of PPI, much higher than what is available in today's tablet and phone displays.

Wide field-of-view: Unlike VR displays, most optical see-through head-worn displays provide narrow fields-of-view. Many provide 20-30 degrees, and 50-60 degrees is a large FOV for this type of display. Ideally, a head-worn AR system would be able to augment a much wider FOV, comparable to what people see through prescription eyewear. With conventional optics, providing a large FOV requires bulky optics, making such a display less attractive in terms of form factor. A company called Innovega is developing an approach where special contact lenses enable the viewer to focus on a display in the form factor of normal eyeglasses [8]. The Pinlight display from UNC Chapel Hill suggests that computational display approaches using unconventional optics may enable compact near-eye displays resembling normal eyewear [9].

Eyestrain: Almost all head-worn displays have a fixed focal distance. If the viewer always fixates on objects at that distance, then this is not a concern. For example, a near-eye display for joggers can have a distant fixed focal distance. However, if the application requires a user to look at a range of distances, near and far and in between, then a fixed focal distance causes eyestrain, due to a conflict between accommodation (the eye focusing at a particular distance) and convergence (two eyes rotating inward to fixate at a particular distance). Optics that dynamically change the focal distance may alleviate this problem, but in the long run we seek displays that support correct accommodation depth cues, via multi-focal plane [10], light-field [11], or even holographic approaches.

Monocular and stereo displays can have their own specific issues. Monocular displays cause binocular rivalry issues and should be assigned to the user's dominant eye, while stereo displays must be matched to the user's interpupillary distance. Providing optical correction for users who wear prescription eyewear remains a significant challenge. Ultimately, custom near-eye displays may be designed and prescribed, just as eyeglasses are today.

For a more detailed discussion of head-worn display designs and technical issues, please see [12].

5. References

[1] Ronald T. Azuma, "A Survey of Augmented Reality," Presence **6**(4), 355-285 (1997).

[2] Nick Wingfield and M. Isaac, "Pokémon Go Brings Augmented Reality to a Mass Audience," New York Times (July 11, 2016).

[3] Ronald Azuma, "Location-Based Mixed and Augmented Reality Storytelling" in *Fundamentals of Wearable Computers and Augmented Reality, 2nd Edition*, Woodrow Barfield, ed. (CRC Press, 2015).

[4] Brian August, "110 Stories – What's your story?" http://110stories.com

[5] Ronald Azuma, "Leviathan at CES 2014," http://ronaldazuma.com/Leviathan_at_CES2014.html

[6] Blair MacIntyre, J. Bolter, J.D. Vaughn, et. al., "Three angry men: An augmented-reality experiment in point-of-view drama," in Proc. International Conference on Technologies for Interactive Digital Storytelling and Entertainment (TIDSE 2003), pp. 230-236.

[7] Kiyoshi Kiyokawa, Y. Kurata and H. Ohno, "An Optical See-Through Display for Mutual Occlusion of Real and Virtual Environments," in International Symposium on Augmented Reality, (IEEE, Los Alamitos, CA, 2000), pp. 60-67.

[8] Innovega, http://www.emacula.io

[9] Andrew Maimone, D. Lanman, K. Rathinavel, et. al., "Pinlight Displays: Wide Field of View Augmented Reality Eyeglasses using Defocused Point Light Sources," ACM Transactions on Graphics **33**(4), article 89 (2014).

[10] Rahul Narain, R. Albert, A. Bulbul, et. al., "Optimal presentation of imagery with focus cues on multi-plane displays," ACM Transactions on Graphics **34**(4), article 59, (2015).

[11] Douglas Lanman and D. Luebke, "Near-eye light field displays," ACM Transactions on Graphics **32**(6), article 220 (2013).

[12] Bernard Kress, "Optics for Smart Glasses, Smart Eyewear, Augmented Reality and Virtual Reality Headsets" in *Fundamentals of Wearable Computers and Augmented Reality, 2nd Edition*, Woodrow Barfield, ed. (CRC Press, 2015).