The Most Important Challenge Facing Augmented Reality

Editors' Note: To celebrate Presence*'s 25th year of publication, we have invited selected members of the journal's original editorial board and authors of several early articles to contribute essays looking back on the field of virtual reality, from its very earliest days to the current time. This essay comes from Ron Azuma who, starting from an article that defined augmented reality (AR) and stands as one of the most influential MIT Press articles of all time, highlights the opportunities presented by recent technological innovation and the greatest challenge facing AR today.*

1 Background and Future of Augmented Reality

In 1997, *Presence* published my survey of Augmented Reality (Azuma, 1997), which became the most cited reference in the field because it defined what "Augmented Reality" was and what the key characteristics and problems were. That report was recently included in a list of "50 influential articles" selected from the entire history of MIT Press journals (2012). Twenty years later, I'd like to reflect on the progress since then, predict the future impact of Augmented Reality, and identify the single biggest challenge that needs to be solved before AR fulfills its ultimate potential.

When I first starting working in Augmented Reality, the only applications researchers could envision anybody using this technology for were professional applications, such as the maintenance and repair of complex equipment and medical visualization. Why? At that time, building an AR system was a ridiculously expensive effort, requiring researchers to build or customize their own tracking systems, graphics engines, and head-worn displays. The only customers who might afford such contraptions were governments and companies. Since then, versions of AR have become affordable to almost everyone. Precise registration that used to require dedicated and expensive tracking hardware became available first through computer vision tracking of markers, then tracking specific images or environments where a 3D model existed, and finally via SLAM (Simultaneous Localization and Mapping) techniques where no a priori model of the environment exists. These enabled AR applications to be deployed on platforms that users already purchased for other purposes: desktops and laptops with built-in cameras, tablets, and smartphones. Most recently, companies have invested billions of dollars into head-worn AR platforms, such as Microsoft's HoloLens and Magic Leap's display.

In 2016, several affordable and high-performance Virtual Reality head-worn display platforms became available to consumers, but some forecasters predicted that Augmented Reality will eventually become an even larger market than VR (Mims, 2016). Pokémon Go opened people's eyes worldwide to the potential of AR experiences that consumers find compelling.

The initial demand for such AR systems is likely to be in the enterprise, for professional applications such as those I previously mentioned, but consumer applications will eventually drive most of the market. AR displays will enable natural interactions with virtual content that is integrated with the surrounding real world, while the users remain engaged with and aware of the real world. Compact, stylish, and portable wide field-of-view head-worn AR displays have the potential to supplant desktop, laptop, tablet, and even smartphone displays. As more of the real world is instrumented with the Internet of Things, a "physical web" will become established where

Ronald T. Azuma
Principal Engineer and Research Manager
Intel Labs
Santa Clara, CA 95054
ronald.t.azuma@intel.com
http://ronaldazuma.com

Presence, *Vol. 25, No. 3, Summer 2016,* 234–238
doi:10.1162/PRES_a_00264

information is tied to tangible objects and locations, and AR will provide the natural interface to that data.

However, many lingering problems must be overcome before we have viable AR platforms that are suitable for ubiquitous consumer deployment. These problems include:

- Precise tracking across large environments to support pixel-accurate registration, both indoors and outdoors, and in all weather conditions, including at night.
- Wide field-of-view optical see-through near-eye displays in compact form factors that can also selectively block light from the surrounding real world.
- Innovative interfaces to control near-eye AR systems without a keyboard and mouse.
- Semantic understanding of real-world objects in large-scale environments without emplaced infrastructure.

2 AR as a Form of Media

Even beyond the technical problems that currently limit the impact of Augmented Reality lies the ultimate and most important challenge facing AR, which is experiential in nature: *How will we establish Augmented Reality as a new form of media, enabling new types of experiences that differ from established media?* If AR is to become ubiquitous in consumer usage, then we must find answers to this question by developing new types of experiences that are engaging and compelling in different ways than traditional media such as books, movies, and even Virtual Reality.

A key limitation is the last technical challenge I previously mentioned: semantic understanding of the surrounding real world. A dirty secret is that most AR systems and experiences really understand very little about the real world around the user. For example, SLAM-based tracking, which works without an a priori model, can recover a point cloud of the surrounding environment, but it only knows *where* those points are, not *what* they are. Therefore, we have AR systems that can embed virtual 3D objects convincingly, putting them into the *context* of the reality, but those objects are not really *connected* to reality. Here's an example to illustrate the difference. With today's tracking technologies, we can build a fighting game between two virtual characters where they appear to exist in the real environment around the user. While they can appear well integrated into the environment, since all the system semantically understands are the virtual characters themselves, all the value from the experience derives solely from the virtual content. The real world is just a background, and therefore the augmentation isn't particularly meaningful. To connect the virtual characters meaningfully, the system needs to understand the real world, so that one character could pick up a real rock, know the properties of that rock, and enable the character to use it in a fight, or perhaps another character could jump behind a real brick wall and use it for protection, and the system would recognize that the brick wall offers protection whereas a hanging towel offers none.

3 Approaches to Make Compelling AR Media

My core hypothesis is that the key to establishing AR as a new form of media is to make the *combination* of the real and virtual crucial, where virtual content is *connected* to reality in compelling and meaningful ways, and the experience cannot be derived solely from the real content or solely from the virtual content.

Accomplishing this requires technical progress in semantic recognition of the real environment surrounding the viewer. There are two main approaches for tackling this problem. First, we can develop systems that can accurately perform object and scene recognition, based upon what we can sense and databases we can draw from (Salas-Moreno, Newcombe, Strasdat, Kelly, & Davison, 2013). This approach will be aided by depth sensing systems such as Intel's RealSense. Second, we can model the real world beforehand, developing extensive databases of reality and using those to help both render and understand the real world (Arth, Pirchheim, Ventura, Schmalstieg, & Lepetit, 2015; Wither, Tsai, & Azuma, 2011). The latter approach might be faster to implement, exploiting extensive infrastructure and databases already being employed for mapping and autonomous driving applications, but at finer detail.

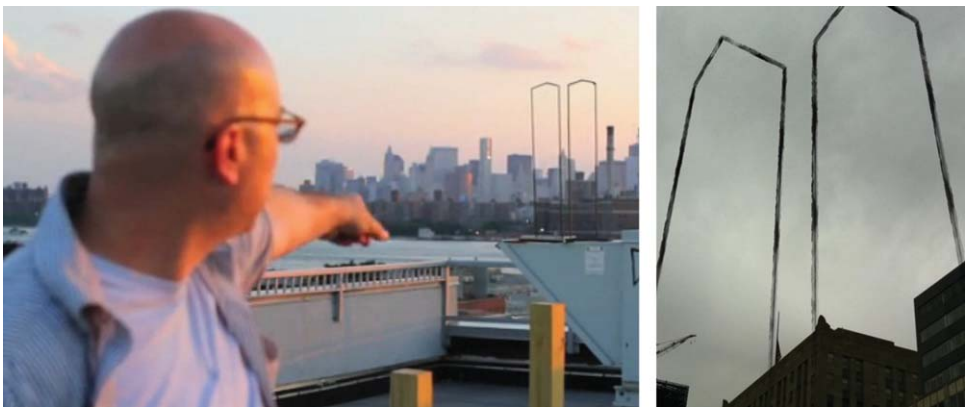

Figure 1. *110 Stories.*

Even after solving the hard problems in semantic understanding, we face an even more difficult design and content problem: How do we make use of this semantic understanding? What design strategies should we employ to build compelling experiences? I speculate that there are at least three approaches, which I call Reinforcing, Reskinning, and Remembering.

In *Reinforcing*, the strategy is to select real locations that are inherently important and powerful by themselves, even without any augmentation, and then to complement that inherent power with augmentations and experiences that are appropriately matched to those locations. The goal is to build a new type of experience that is more powerful than just the virtual content or the real locations by themselves. For example, if a person were interested in learning about the Battle of Gettysburg, he or she could watch a movie or documentary about that event. But since Gettysburg is a real place, he or she can also visit the actual site of that battle. If you know why that battle is important, simply being at the actual site can be an emotional and poignant experience. An AR experience built with the Reinforcing strategy would emplace virtual 3D recreations of the events directly upon the actual site of the battle, attempting to merge the best aspects of both virtual and real.

My favorite example of the Reinforcing strategy is *110 Stories* (August, 2011). If you are in or near Manhattan, this experience renders a depiction of where the Twin Towers should be. The creator, Brian August, decided to render the towers not as photorealistic depictions but instead just as outlines, as if they were drawn against the sky with a grease pencil, as shown in Figure 1. This is an example of how solving this ultimate challenge is as much an artistic and design challenge as it is an engineering challenge. By rendering just the outline, which is technically easier than rendering a photorealistic building, he improved the experience because it matched the message that it tries to convey: Those towers aren't there anymore. And they are supposed to be there.

In the *Reskinning* strategy, we recognize that for the vast majority of real locations there is nothing particularly special about that location, so we instead remake, reinterpret, or redefine it to suit the needs of the story or experience that we are overlaying on top of reality. Vernor Vinge's award-winning science fiction novel *Rainbows End* (Vinge, 2006) imagined a world where perfect AR systems exist and people create persistent virtual worlds called Belief Circles that are overlaid 1:1 upon reality, so that instead of seeing a modern town, one might choose a Belief Circle that would reskin that town as if it existed in the 1800s, with horse-drawn carriages replacing automobiles.

Pokémon Go is clearly the most commercially successful example of the Reskinning strategy. It is the first mass-market hit of any VR or AR experience, with over 100 million installations within the first month of release (Perez, 2016). By redefining real-world locations as vir-

tual gyms and places where you can catch virtual creatures or refill your supply of virtual balls used to catch those creatures, Pokémon Go reinterprets the real world into a landscape where the players have to go to specific locations to make progress in the game. It motivates players to explore the world around them by walking around and to perhaps discover locations and other people that they would otherwise not seek out or interact with.

Finally, the *Remembering* strategy also acknowledges that many real-world locations may be mundane, but some memories that happened at those locations are important, providing a lever that we can exploit to make compelling experiences. This strategy is similar to Reinforcing, but it occurs at a more personal level. Sites such as the Gettysburg battlefield or the World Trade Center have overwhelming historical importance that everyone knows and recognizes, and that limits the types of Reinforcing experiences at such locations to ones appropriate to those contexts. In contrast, Remembering is much more personal, as the same mundane location could evoke different meanings in different people. The same spot on the Berkeley campus might be remembered by one person for her participation in the Free Speech Movement, by another as the location where he first met his future spouse, and by me as the spot where I first heard of the loss of the space shuttle Challenger. Furthermore, experiences based on Remembering might be of interest only to a small group of people: perhaps yourself and some friends and family. However, such personal stories and experiences are no less important than ones created by professional storytellers for a mass audience.

AR experiences that recreate key moments in the life of your family, at the actual locations where those occurred, would be compelling instantiations of the Remembering strategy. I have photos and videos of my wedding, but if I could show an AR recreation of that event at the actual gazebo where it occurred, it could be compelling, almost like taking a time machine to see an important event in my past. Similarly, a hallway in my house may be mundane, but seeing a recreation of my son's first steps at the spot where that actually happened would be a powerful experience. The Holoportation project from Microsoft Research (2016) is a step toward providing the technologies needed to make this vision a reality.

4 Conclusion

How will we know when we have succeeded in the ultimate challenge of establishing AR as a new form of media? My answer is that success comes when this new media form replicates the power of traditional media, such as books and films, to change a person's perspective, behaviors and beliefs. There are numerous examples of books, films, and other traditional media changing the lives of many people. If an AR experience is powerful enough to give a viewer a different perspective about something, whether that perspective is historical, cultural, social, political, or anything other, and it is sufficiently compelling that it makes the viewer change his or her beliefs and behaviors, then we will know we have succeeded.

A project called Three Angry Men (MacIntyre et al., 2003) was an experiment that strove for this type of impact. In this experience, you observe the deliberations of three jury members, where one is African-American and another is a bigot. You sit in one of three chairs around a table. While in that chair, you see the viewpoint of the juror who is represented by that chair and you hear his inner thoughts. At any point, you can stand up, and the experience pauses. Then you can walk to another chair and sit down. At this point, the narrative resumes from the place where it paused, but the entire experience changes. Now you see and hear things from another juror's point of view. The exact words, intonations, and even the appearances of the other jurors change to conform to the prejudices of the juror you are "inhabiting."

I believe the full power of Augmented Reality will be unleashed once the field has established it as a new, viable, and powerful form of media, with unique strengths. It has taken over 20 years to move Augmented Reality from ridiculously expensive, custom systems that existed only in research labs to the threshold of self-contained, head-worn systems that consumers can afford. I hope it will take less than 20 more years for the visions I described here to become reality.

If you are interested in a deeper discussion on this topic, please see my book chapter on this subject (Azuma, 2015).

References

Arth, C., Pirchheim, C., Ventura, J., Schmalstieg, D., & Lepetit, V. (2015). Instant outdoor localization and SLAM initialization from 2.5D maps. *IEEE Transactions on Visualization and Computer Graphics, 21*(11), 1309–1318.

August, B. (2011). *110 stories—What's your story?* Retrieved from http://110stories.com

Azuma, R. T. (1997). A survey of augmented reality. *Presence: Teleoperators and Virtual Environments, 6*(4), 355–385.

Azuma, R. (2015). Location-based mixed and augmented reality storytelling. In W. Barfield (Ed.), *Fundamentals of wearable computers and augmented reality* (2nd Ed., pp. 259–276). Boca Raton: CRC Press.

MacIntyre, B., Bolter, J. D., Vaughn, J., Hannigan, B., Gandy, M., Moreno, E., . . . Voida, S. (2003, March 24–26). Three angry men: An augmented-reality experiment in point-of-view drama. *Proceedings of the 1st International Conference on Technologies for Interactive Digital Storytelling and Entertainment,* 230–236.

Microsoft Research (2016). *Holoportation—Microsoft Research.* Retrieved from https://www.microsoft.com/en-us/research/project/holoportation-3/

Mims, C. (2016, March 14). Here's the most exciting technology you'll ever encounter. *The Wall Street Journal.* Retrieved from http://www.wsj.com/articles/heres-the-most-exciting-technology-youll-ever-encounter-1457928061

MIT Press Journals (2012). *50 influential journal articles.* Retrieved from http://www.mitpressjournals.org/page/50articles

Perez, S. (2016, August 1). Pokémon Go passed 100 million installs over the weekend. Retrieved from https://techcrunch.com/2016/08/01/pokemon-go-passed-100-million-installs-over-the-weekend/

Salas-Moreno, R., Newcombe, R., Strasdat, H., Kelly, P. H. J., & Davison, A. (2013, June 23–28). SLAM++: Simultaneous localisation and mapping at the level of objects. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition,* 1352–1359.

Vinge, V. (2006). *Rainbows end.* New York: Tor Doherty Associates.

Wither, J., Tsai, Y.-T., & Azuma, R. (2011). Indirect augmented reality. *Computers & Graphics, 35*(4), 810–822.