The Development of Mobile Augmented Reality

Lawrence J. Rosenblum, National Science Foundation*

Steven K. Feiner, Columbia University

Simon J. Julier, University College London*

J. Edward Swan II, Mississippi State University*

Mark A. Livingston, Naval Research Laboratory

Abstract: This chapter provides a high-level overview of fifteen years of augmented reality research that was sponsored by the U.S. Office of Naval Research (ONR). The research was conducted at Columbia University and the U.S. Naval Research Laboratory (NRL) between 1991 and 2005 and supported in the later years by a number of university and industrial research laboratories. It laid the groundwork for the development of many commercial mobile augmented reality (AR) applications that are currently available for smartphones and tablets. Furthermore, it helped shape a number of ongoing research activities in mobile AR.

Keywords: augmented reality, mobile computing, usability, situational awareness, user interfaces, human factors, computer graphics

Introduction

In 1991, Feiner, working at Columbia University, received an ONR Young Investigator Award for research on “Automated Generation of Three-Dimensional Virtual Worlds for Task Explanation.” In previous work, his Computer Graphics and User Interfaces Lab had developed IBIS, a rule-based system that generated 3D pictures that explained how to perform maintenance tasks (Seligmann and Feiner, 1989; Seligmann and Feiner, 1991), and an AR window manager that embedded a stationary flat panel display within a surrounding set of 2D windows presented on a home-made, head-tracked, optical see-through display (Feiner and Shamash, 1991). The goal of the new ONR-funded research was to expand this work to generate 3D virtual worlds that would be viewed through head-tracked

---

* Current affiliation; research performed at the Naval Research Laboratory

displays. Beginning in the summer of 1991, Feiner and his PhD students Blair MacIntyre and Dorée Seligmann modified IBIS and combined it with software they developed to render 3D graphics for their head-tracked, optical–see-through, head-worn display. The new system, which they later named KARMA (Knowledge-based Augmented Reality for Maintenance Assistance), interactively designed animated overlaid graphics that explained how to perform simple end-user maintenance for a laser printer (Feiner et al., 1992; Feiner et al., 1993). This was the first of a set of ONR-funded projects their lab created to address indoor AR.

In the course of their work, Feiner had realized that despite the many difficult research issues that still needed to be solved to make indoor AR practical, taking AR outside would be a crucial next step. He had heard about work by Loomis and colleagues (Loomis et al., 1993) using differential GPS and a magnetometer to track a user's head and provide spatial audio cues in an outdoor guidance system for the visually impaired. Inspired by that work, Feiner decided to combine these position and orientation tracking technologies with a see-through head-worn display to create the first example of what his lab called a Mobile AR System (MARS). Starting in 1996, Feiner and his students developed the (barely) wearable system shown in Fig. 1. This system was mounted on an external frame backpack, and was powered by a battery belt (Feiner et al., 1997). A stylus-based handheld computer complemented the head-worn display. The system was connected to the Internet using an experimental wireless network (Ioannidis et al., 1991).

Fig. 1 The Columbia Touring Machine in 1997. *Left*: A user wearing the backpack and operating the hand-held display. *Right*: A view through the head-worn display. (Recorded by a video camera looking through the head-worn display.)

The initial MARS software was developed with colleagues in the Columbia Graduate School of Architecture and conceived of as a campus tour guide, named the "Touring Machine." As the user looked around, they could see Columbia's buildings and other major landmarks overlaid by their names, as shown in Fig. 1, obtained from a database of geocoded landmarks. Using head-orientation to approximate gaze tracking, the object whose name stayed closest to the center of a small circular area at the middle of the head-worn display for a set period of time was automatically selected, causing a customized menu to be presented at the top of the display. The menu could be operated through a touch pad mounted on the back of the hand-held display, allowing the user to manipulate the touchpad easily while holding the hand-held display. This controlled a cursor presented on the head-worn display. One menu item overlaid the selected building with the names of its departments; selecting a department name would cause its webpage to be displayed on the hand-held display. The overlaid menus viewed on the head-worn display were also presented on the hand-held display as custom web pages. A conical cursor at the bottom of the display pointed to the currently selected building.

The software was split into two applications, written using an infrastructure that supported distributed applications (MacIntyre and Feiner, 1996). The tour application on the backpack was responsible for generating graphics and presenting it on the head-worn display. The application running on the hand-held computer was a custom HTTP server in charge of generating custom web pages on the fly and accessing and caching external web pages by means of a proxy component. This custom HTTP server communicated with an unmodified web browser on the hand-held computer and with the tour application.

Program Development

Many important research issues would need to be addressed to make the Touring Machine into more than a research prototype. After Rosenblum's completion of a two-year tour at the ONR European Office (ONREUR) in 1994, he founded and directed the NRL Virtual Reality Laboratory (VRL). Rosenblum had seen the potential of Feiner's research and had included it in talks he gave about the ONR computer science research program in Europe while at ONREUR. In early 1998, Rosenblum suggested that Julier, then a VRL team member, and Feiner put together a proposal to ONR that would explore how mobile AR could be developed to make practical systems for use by the military. This funding was awarded and, for NRL, was supplemented by an NRL Base Program award. The program, called the Battlefield Augmented Reality System (BARS™) (Julier et al., 2000; Livingston et al., 2002), would investigate how multiple mobile AR users on foot could cooperate effectively with one another and with personnel in combat operations centers, who had access to more powerful computing and display facilities. The proposed work would build on the Touring Machine at Columbia and on previous NRL research using the VRL's rear-projected workbench (Rosenblum et al., 1997) and CAVE-like multi-display environment (Rosenberg et al., 2000). Several challenges became apparent: building and maintaining environmental models of a complex and dynamic scene, managing the information relevant to military operations, and interacting with this information. To achieve such a system, the architectures for the software to encapsulate these features had to be developed. Although this also required high-fidelity tracking of multiple mobile users, our primary focus was on the information management and interaction components.

Information Management

Fig. 2 Situated documentary. A 3D model of an historic building, long since demolished, is shown at its former location. (Recorded by a video camera looking through the head-worn display.)

Situated documentaries. In addition to the spatialized text and simple graphics supported by the Touring Machine, it was clear that many AR applications would benefit from the full range of media that could be presented by computer. To explore this idea, Columbia developed *situated documentaries*—narrated hypermedia briefings about local events that used AR to embed media objects at locations with which they were associated. One situated documentary, created by Feiner and his students in collaboration with Columbia colleagues in Journalism, presented the story of the 1968 Columbia Student Strike (Höllerer et al., 1999). Virtual 3D flagpoles located around the Columbia campus were visible through the head-worn display; each flagpole represented part of the story and was attached to a menu that allowed the user to select portions of the story to experience. While still images were presented on the head-worn display, playing video smoothly on the same display as the user looked around was beyond the capabilities of the hardware, so video was shown on the hand-held display. In developing our situated documentaries, we were especially interested in how multimedia AR could improve a user's understanding of their environment. One example (Fig. 2) presented 3D models of historic buildings on the head-worn display, overlaid where they once stood. The user could interact with a timeline presented on the hand-held display to move forward and backward in time, fading buildings up and down in synchrony with a narrated presentation.

Some of the key scientific contributions of the Columbia/NRL research were embodied in our development of a general model for mobile AR user interfaces (Höllerer et al., 2001). Our model comprised three essential phases, software implementations of which were included in our prototypes: information filtering, UI component design, and view management.

Information filtering. The display space for a mobile AR system is limited, and, in order to utilize the technology in a 3D urban environment, it was clear that effective methods were needed to determine what to display. Based in part on the user's spatial relationship to items of interest, algorithms were developed (Julier et al., 2000) to determine the information that is most relevant to the user (Fig. 3).

Fig. 3 The need for information filtering. *Left:* "raw" data, a confusing clutter of many different labels and objects. *Right*: filtered output draws the foreground building for context, the path the user is following, and a potential threat. (Recorded by a video camera looking through the head-worn display.)

UI component design. This phase determines how the selected information should be conveyed, based on the kind of display available, and how accurately the user and objects of interest can be tracked relative to each other. For example, if sufficiently accurate tracking is possible, a representation of an item can be overlaid where it might appear in the user's field of view; however, if the relative location and orientation of the user and object are not known with sufficient accuracy, the item might instead be shown on a map or list.

View management. View management (Bell et al., 2001). refers to the concept of laying out information on the projection plane so that the relationships among objects are as unambiguous as possible, and physical or virtual objects do not obstruct the user's view of more important physical or virtual objects in the scene. Our work on view management introduced an efficient way of allocating and querying space on the viewplane, dynamically accounting for obscuration relationships among objects relative to the user.

Authoring tools. Authoring mobile AR experiences using our early systems was tedious, and relied on coding large portions of the experience in textual programming languages, along with creating databases using conventional tools (Fig. 4). This required that programmers be part of any authoring team. Inspired by multimedia authoring systems (for example, Macromedia Director), AR authoring tools were developed to allow content developers to create richer AR experiences (Julier et al., 1999). A key concept was to combine a 2D media timeline editor, similar to that used in existing multimedia authoring systems, with a 3D spatial editor that allowed authors to graphically position media objects in a representation of the 3D environment (Güven and Feiner, 2004).

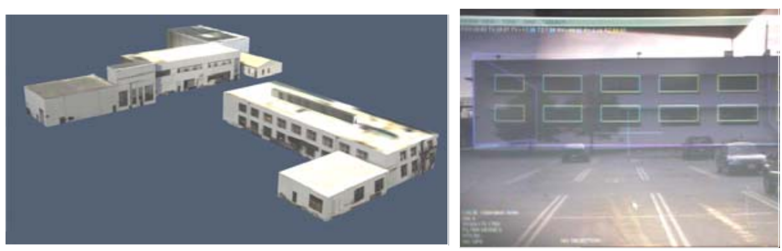

Fig. 4 *Left*: Campus model geared towards visualization (without semantic elements). *Right*: The model shown in AR with a wireframe overlay, recorded by a video camera looking through the head-worn display. Note the misalignment in the top-left corner caused by optical distortion in the head-worn see-through display. This is one of the challenges of mobile AR systems.

Development Iterations

The earlier development of BARS was carried out in two distinct phases. The Phase I mobile system was a high performance (for its time) mobile hardware platform with the software and graphical infrastructure needed to be able to deliver information about a dynamically changing environment to a user with limited interaction capabilities. The initial BARS prototype consisted of a differential kinematic GPS receiver, an orientation tracker, a head-worn display, a wearable computer and a wireless network. The BARS software architecture was implemented in Java and C/C++. The initial user interface had simple graphical representations (wireframe icons) and was enhanced using information filtering. Techniques for precise registration were developed, including algorithms for calibrating the properties of the head-worn display and the tracking system. To mitigate the problem of information overload, a filtering mechanism was developed to identify the subset of information that must be shown to the user. Accurate models of some of the buildings and building features were developed for both NRL and Columbia. The Phase II system integrated the mobile AR system into a multi-system collaborative environment. The BARS system architecture was extended to allow multiple, distributed systems to share and change a common environment. Preliminary implementations of components were completed.

Two systems were developed—one based on consumer grade hardware, the other using embedded computers (Fig. 5). There was a direct tradeoff of capability and weight versus usability. Both systems used Sony Glasstron optical see through head-worn displays, and a loosely integrated tracking solution consisting of a real-time kinematic GPS receiver and an orientation sensor. The first demonstration of BARS was in November 1999. NRL and Columbia demonstrated early versions of some of this joint work at ISWC 2000, showing the new backpack systems (Fig. 5). At SIGGRAPH's Emerging Technologies Pavilion (Feiner et al., 2001), we first demonstrated integration with wide-area tracking in a joint effort with InterSense; Eric Foxlin contributed an early version of the IS-1200 tracker technology and large ceiling-mounted fiducials.

Fig. 5: Experimental mobile AR systems of NRL (left) and Columbia (right) in 2000.

Program Expansion

The preliminary prototypes demonstrated the capabilities and potential of single user AR. One area of shortcoming was in the user interface and information visualization. NRL and Columbia continued their research in these areas to develop new information filtering algorithms and display techniques. They addressed issues such as the "X-ray vision" problem for occlusion (described below). However, other hard problems remained. Additional issues were addressed by a combination of university and industrial research and development (sometimes working individually and sometimes with NRL/Columbia). These topics included 3D urban terrain reconstruction, tracking and registration, usability of mobile AR systems, and display hardware.

ONR Program Expansion

Because the NRL/Columbia BARS system had successfully demonstrated the potential of mobile AR, Andre van Tilborg, then the Director of the Mathematical, Computer, and Information Sciences and Technology Division at ONR, asked Rosenblum, who was working part time for ONR while serving as Director of the Virtual Reality Laboratory at NRL, to assemble a primarily university-based research program to complement the Columbia/NRL research program and assure that the field advanced. We believe this program, combined with the NRL/Columbia effort, was the largest single effort through that time to perform the research necessary to turn mobile AR into a recognized field and that it provided the basis for advances on an international scale.

The program was based upon several options available within ONR and U.S. DoD for funding research and totaled several million dollars annually for approximately five years, although most PIs were funded for differing periods during that time. The majority of the awards were the typical three-year ONR research grants for university projects (similar to those of the National Science Foundation), but also included two industrial awards as well as related research conducted under a DoD Multidisciplinary University Research Initiative (MURI), which was a $1M/year award for five years to researchers at the University of California Berkeley, the Massachusetts Institute of Technology, and the University of California San Francisco. Only a portion of the MURI research, relating to the reconstruction of 3D urban terrain from photographs, applied directly to the ONR mobile AR program. Institutions and lead PIs involved in this program were:

- Tracking and Registration Ulrich Neumann, University of Southern California; Reinhold Behringer; Rockwell)
- Usability of Mobile AR systems (Deborah Hix, Virginia Polytechnic Institute and State University; Blair MacIntyre, Georgia Institute of Technology; Brian Goldiez, University of Central Florida)
- 3D Urban Terrain Reconstruction (Seth Teller, Massachusetts Institute of Technology; Jitendra Malik, University of California at Berkeley; William Ribarsky, Georgia Institute of Technology)
- Retinal Scanning Displays (Tom Furness, University of Washington; Microvision, Inc.)

Also, two separately funded NRL projects funneled results into BARS:

- 3D Multimodal Interaction (NRL and Phil Cohen, Oregon Graduate Institute)

- Interoperable Virtual Reality Systems (NRL)

The remainder of this subsection briefly summarizes a few of these projects.

The Façade project at Berkeley acquired photographs (of a limited area) and developed algorithms to reconstruct the geometry and add texture maps, using human-in-the-loop methods. This research inspired several commercial image-based modeling packages. The Berkeley research went on to solve the difficult inverse global illumination problem: given geometry, light sources, and radiance images, devise fast and accurate algorithms to determine the (diffuse and specular) reflectance properties (although this portion of the research was not directly related to mobile AR).

The 3D urban terrain reconstruction research at MIT made seminal algorithmic advances. Previous methods, including the Berkeley work, relied on human-in-the-loop methods to make point or edge correspondences. Teller developed a sequence of algorithms that could take camera images collected from a mobile robot and reconstruct the urban environment. Algorithms were developed for image registration, model extraction, facade identification, and texture estimation. The two main advances of this research were to provide a method that did not require human intervention and to develop algorithms that allowed for far faster reconstruction than was previously possible. The model extraction algorithm was shown to be $O(N+V)$, where N is the number of images and V is the number of voxels, while previous methods were $O(N*V)$.

One missing component in the development of mobile AR prior to the ONR program was integrating usability engineering into the development of a wearable AR system and into producing AR design guidelines. Virginia Tech, working jointly with NRL, performed a domain analysis (Gabbard et al., 2002) to create a context for usability engineering effort, performed formative user-based evaluations to refine user interface designs, and conducted formal user studies, both to understand user performance and to produce design guidelines. An iterative process was developed, which was essential due to the extremely large state space generated by the hundreds of parameters that arise from the use of visualization and interaction techniques. The team developed a use case for a platoon in an urban setting and tested BARS interaction and visualization prototypes using semi-formal evaluation techniques with domain experts (Hix et al., 2004). Out of these evaluations emerged two *driving problems* for BARS, both of which led to a series of informal and formal evaluations: (1) AR depth perception and the "X-ray vision" problem (i.e., correct geospatial recognition of occluded objects by the user), and (2) text legibility in outdoor settings with rapid and extreme illumination changes. For the text legibility problem, Virginia Tech and NRL designed an active color scheme for text that accounted for the color capabilities of optical see-through AR displays. Appropriate coloring of the text foreground enabled faster reading, but using a filled rectangle to provide a background enabled the fastest user performance (Gabbard et al., 2007).

Tracking the user's head position relative to the real-world scene remains one of the difficult problems in mobile AR. Research at the University of Southern California developed an approach based on 2D line detection and tracking. Features included the use of knowledge that man-made structures were in the scene. The nature of these structures permitted use of larger scale primitives (e.g. windows) that provided more geometrical information for stable matching. This approach proved more robust than the use of point like features. A line-based autocalibration algorithm was also developed.

Because tracking head-motion and aligning the view correctly to the real world is so difficult, methods are needed to convey registration uncertainty. Note that this tends to be task dependent, since placing a label on a building requires quite a different accuracy than identifying a specific window. Joint research by Georgia Tech and NRL resulted in a methodology for portraying uncertainty (MacIntyre et al., 2002). The statistics of 3D tracker errors were projected into 2D registration errors on the display. The errors for each object were then collected together to define an error region. An aggregate view of the errors was then generated using geometric considerations based on computing an inner and outer convex hull and placed over the scene (Fig. 6).

Fig. 6 *Left*: Accurately aligning a marker on a window can be hard to achieve with tracking errors. *Center*: A sufficiently large boundary can be guaranteed to enclose the desired object if tracking error is bounded. *Right*: Text indicators can direct users to the correct point when tracking errors prevent correct registration.

The one disappointing area of the research program was in the attempt to produce the hardware for the AR display. The Sony Glasstron did not have sufficient brightness for the augmented image to be seen in bright sunlight; it was nearly unusable under that condition. Program management felt that the Microvision retinal scanning display, which used a laser to scan an image directly onto the eye, had the potential to overcome the scientific issues involved in producing a display with sufficient resolution and field of view and would produce sufficient luminance to work under conditions ranging from bright sunlight to darkness. While Microvision made advances in their display technology, they did not produce a display that completely met the needs of mobile AR. The University of Washington performed basic research to scan bright images on the retina while also tracking the retinal and head position using the same scanning aperture. The research was theoretically successful, but (at least in the time period of the program) it was not transitioned into a commercial product.

The "X-ray Vision" Problem and the Perception of Depth

Fig. 7: *Left*: one of the concept sketches for how occluded buildings and units might be represented in BARS. *Right*: a photograph taken through our optical see-through display in 2003, with a similar protocol implemented.

Our domain analysis revealed that one challenge of urban operations is maintaining understanding of the location of forces that are hidden by urban infrastructure. This is called the "X-ray vision" problem: Given the ability to see "through" objects with an AR system, how does one determine how to effectively represent the locations of the occluded objects? This led us to develop visualization techniques that could communicate the location of graphical entities with respect to the real environment. Drawing on earlier work at Columbia to represent occluded infrastructure (Feiner and Seligmann, 1992), NRL implemented a range of graphical parameters for hidden objects. NRL and Virginia Tech then conducted a user study to examine which of the numerous possible graphical parameters were most effective. We were the first to study objects at far-field distances of 60–500 meters, identifying visualization parameters (Fig. 7) such as drawing style, opacity settings, and intensity settings that could compensate for the lack of being able to rely on a consistent ground plane and identifying which parameters were most effective (Livingston et al., 2003). NRL began to apply depth perception measurement techniques from perceptual psychology. This led us to adopt a perceptual matching technique (Swan et al., 2006), which we used to study AR depth perception at distances of 5–45 meters in an indoor hallway. Our first experiment with this technique showed that user behavior with real and virtual targets was not significantly different when performing this perceptual matching against real reference objects (Livingston et al., 2005). We later used the technique to study how AR depth perception differs in indoor and outdoor settings (noting an underestimation indoors and overestimation outdoors) and how linear perspective cues could be simulated outdoors to assist users (Livingston et al., 2009). The studies have produced some conflicting data regarding underestimation and overestimation. This remains an active research area, with many parameters being investigated to explain the effects observed in the series of experiments.

Integration of a Component-based System

The software architecture had to support two goals: coordination of all the different types of information required and providing flexibility for the different systems under test. NRL implemented a substantial amount of the system using the Bamboo toolkit (Watson and Zyda, 1998). Bamboo decomposed an application into a set of modules that could be loaded in a hierarchical manner with dependencies between them. Into this framework, NRL researchers could plug in UI components, such as the event manager for display layout, designed and tested at Columbia (Höllerer et al., 2001).

One example of the success of this architecture was the demonstration at the International Symposium on Mixed and Augmented Reality in November 2004. Into the NRL BARS framework (with video to provide a multi-person AR view of Washington, DC) were integrated Columbia’s view management for placing labels and Virginia Tech’s rules for providing color or intensity contrast to ensure label legibility. Another success was a variation on the BARS system to integrate semi-automated forces, providing a realistic training scene for military call-for-fire. This system was demonstrated at Quantico Marine Corps Base in October 2004.

Ongoing Research

The ONR Mathematical, Computer, and Information Sciences and Technology Division program helped to launch major efforts within the U.S. Department of Defense to build usable mobile AR systems for military applications. These programs focused on applications, but recognized the need for fundamental research and enabled continued efforts in the basic research as well as applied research domains. These programs enabled some members of the ONR AR program to continue their work. This section focuses on recent NRL and Columbia research and development.

Two particularly broad efforts, both inspired by the NRL-led work, are the operationally-focused DARPA Urban Leader Tactical Response Awareness and Visualization (ULTRA-Vis) program, and the DoD Future Immersive Training

Environments (FITE) Joint Capability Technology Demonstration; a follow-up ONR program called Next-generation Naval Immersive Training (N2IT) carries on the training research.

NRL participated in both programs, building on its experiences with both the training applications for urban combat skills and the human factors evaluations, which apply to both training and operational contexts. User interface techniques continue to be a critical element of the research (Livingston et al., 2011). NRL in recent years has also continued to study the human factors issues described above. Livingston and Feiner collaborated on exploring AR stereo vergence (Livingston et al., 2006). Livingston and Swan have maintained collaboration on the depth perception and X-ray vision research (Swan et al., 2007; Livingston et al., 2009), as well as other human factors issues. We became interested in using perceptual-motor tasks, which have been widely applied in perceptual psychology, to study AR depth perception (Jones et al., 2008; Singh et al., 2010). Recent work has studied reaching distances, which are important for other AR applications, such as maintenance. At NRL, the original operational context of "X-ray vision" continues to be a topic of interest (Livingston et al., 2011). NRL continues to offer technical support to ONR programs sponsoring research on improving see-through displays and tracking systems appropriate for training facilities.

Columbia was funded through the Air Force Research Laboratory, and later through ONR, to examine the feasibility and appropriate configuration of AR for maintenance of military vehicles (Henderson and Feiner, 2010; Henderson and Feiner, 2011). Feiner and his students have also continued to explore a broad range of research issues in AR. The concept of situated documentaries has led to the study of *situated visualization*, in which information visualizations are integrated with the user's view of the environment to which they relate, with applications to site visits for urban design and urban planning (White and Feiner, 2009). Interacting with a scale model of an environment in AR is a challenge; in some cases, performance can be improved when 3D selection is decomposed into complementary lower dimensional tasks (Benko and Feiner, 2007). Leveraging the ubiquity of handheld devices with built-in cameras and vision-based tracking, Columbia has investigated the advantages of having users take snapshots of an environment and quickly switch between augmenting the live view or one of the snapshots (Sukan and Feiner, 2010).

Predictions for the Future

When mobile AR research began, few people saw the potential applications as having a deep impact in the consumer market. However, if one compares images of our early work to images of tourist guides now available for mobile phones (Fig. 8), it is apparent that our vision of mobile AR has reached the consumer market, even if the application requirements in the military domain have proven more challenging to fulfill. Even though AR is no longer merely a laboratory curiosity, we believe that many challenges remain.

Fig. 8 *Top Left*: An image from the 2002 implementation of the Touring Machine, recorded by a video camera looking through the optical–see-through head-worn display, shows an AR restaurant guide, a civilian example of supporting a user exploring an unknown urban environment (Bell et al., 2002). *Top Right*: An image from Mtrip Travel Guides shows a modern implementation of commercial AR guidance. Image © 2011 Mtrip Travel Guides, http://www.mtrip.com; used by permission. *Bottom*: BARS was envisioned to be able to provide urban cues integrated in 3D. This BARS image shows a compass for orientation and a route for the user to follow in addition to a street label and the location of a hidden hazard. This video capture image is from 2003.

Tracking

There have been many advances in hardware design. Tracking sensors are now readily available. Almost all recent mobile phones contain built-in GPS and inertial measurement (magnetometers, accelerometers and gyroscopes) sensors. However, despite this wide availability of sensing devices and decades of intensive research, tracking remains one of the most significant challenges facing AR. Non-line-of-sight and multi-path means that GPS position solutions can contain errors of between tens and hundreds of meters. Metallic structures can introduce angular errors of 180° in magnetometer readings. As mobile devices improve in power, we are already seeing vision-based algorithms for tracking new environments being applied to consumer AR games. However, many of these systems rely on the assumption that the entire world is static.

Currently, very accurate tracking is available in two cases. The first set of cases consists of niche applications, such as surgical assistance or maintenance, repair, or fabrication of delicate equipment. These can justify the use of expensive, intrusive, and dedicated equipment. The second case can rely on vision-based algorithms to lock virtual cues to specific locations. Vision-based tracking can be used effectively with known planar targets (e.g., the discrete markers of ARToolKit or the clusters of natural features used in the Qualcomm AR SDK). Sophisticated vision algorithms that search for features to track in previously unknown static environments are now being deployed commercially in mobile games. As a result, we believe these cases will continue to be important to AR applications.

In the long-term, we see multiple directions for tracking solutions. First, hybrid systems of sensors have long demonstrated how one type of sensor can compensate for even catastrophic errors in other sensors. As sensors improve, the number of useful combinations and the accuracy increase. Second, as hardware performance increases, more advanced vision-based algorithms become available to mobile hardware. Vision-based systems are moving towards the use of large static structures as tracking landmarks. A more advanced system could recognize specific structures and compute a matching perspective view of virtual objects, without computing metric estimates of position and orientation. A related question is whether absolute 3D spatial models are required in many mixed-reality applications. If an augmentation can be defined relative to recognizable landmarks in the real world, it may be necessary only to have accuracy relative to that landmark. For example, a proposed extension to a building must connect to that building accurately, whether or not the 3D model of the building is accurate relative to some external coordinate system. Third, the robustness of sensors and hybrid systems to well-known disturbances can be improved; this is especially critical in dynamic, uncontrolled outdoor scenarios (e.g., with difficult lighting conditions or moving people and objects). We also believe that the use of robust interfaces, cognizant of the structure of the environment, the ambiguity of information, and the impact of errors can be used to adapt the display to mitigate the effects of tracking errors. Finally, the size, weight, and power requirements of mobile tracking solutions will continue to be reduced.

Form Factor

Many current AR applications are based on hand-held devices such as mobile phones. For many reasons (e.g., ease of being carried or fit into a pocket), the devices cannot become substantially larger. However, this leads to a mismatch—the camera has a wide field-of-view (in some cases, more than 60°), but the angle subtended by a hand-held display is very small (typically 12–16°). As a result, this introduces many user interface challenges. Apart from issues such as fatigue, such displays can monopolize a user's attention, potentially to the exclusion of other things around them. This is clearly unacceptable for dangerous tasks such as disaster relief. Even in tourism applications, a tourist needs to be aware of the environment to navigate effectively. Furthermore, hand-held devices, by definition, also need to be held, which can make many common tasks that could benefit from AR hard to perform.

We believe that if AR is to realize its full potential, hand-held form factors, despite much of the hype they are receiving now, simply are not adequate. Rather, AR systems will need to be based on head-worn displays—eyewear—which must become as ubiquitous as earphones. For that to happen, AR eyewear must be comfortable, good-looking, of sufficient optical quality that they feel like looking through properly fitted eyeglasses, and relatively inexpensive. Many of the other hardware barriers to mobile AR have fallen, thanks to small but powerful sensor-laden smartphones, coupled with affordable high-bandwidth data access, and rapidly improving tracking ability. Consequently, we are now seeing far-sighted consumer electronics companies, both large and small, exploring how to develop appropriate AR eyewear.

Summary

We have been very fortunate to work on mobile AR at a pivotal time in its development. Through the research programs described, we have been able to explore many important issues, and it is good to see that some of the once impractical ideas we investigated are now incorporated in applications running on consumer devices. However, despite its promise, mobile AR has a substantial way to go to realize its full potential. If AR is to become an effective, ubiquitous technology, many fundamental research and development challenges remain to be overcome.

Acknowledgements

The authors wish to thank Yohan Baillot, Reinhold Behringer, Blaine Bell, Dennis Brown, Aaron Bryden, Enylton Coelho, Elliot Cooper-Balis, Deborah Hix, Joseph Gabbard, Brian Goldiez, Tobias Höllerer, Bryan Hurley, Marco

Lanzagorta, Dennis Lin, Blair MacIntyre, Douglas Maxwell, Ulrich Neumann, Gregory Schmidt, Erik Tomlin, Ross Whitaker, Suya You, and Catherine Zanbaka. We appreciate the support we had over this extended time period from ONR. In particular, we thank Andre van Tilborg, Wen Masters, Paul Quinn, and Ralph Wachter. We also thank Randy Shumaker and John McLean for their support for the NRL portion of the research. Opinions expressed in this article are those of the authors and do not represent official positions of the Naval Research Laboratory, the National Science Foundation, or any other institution.

References

Bell B, Feiner S, and Höllerer T (2001). "View Management for Virtual and Augmented Reality." *ACM Symposium on User Interface Software and Technology*, pages 101-110

Bell B, Feiner S, and Höllerer T (2002). "Information at a glance." *IEEE Computer Graphics & Applications* 22(4):6-9

Benko H and Feiner S (2007). "Balloon Selection: A Multi-Finger Technique for Accurate Low-Fatigue 3D Selections." *IEEE Symposium on 3D User Interfaces*, pages 79-86

Feiner S, Bell B, Gagas E, Güven S, Hallaway D, Höllerer T, Lok S, Tinna N, Yamamoto R, Julier S, Baillot Y, Brown D, Lanzagorta M, Butz A, Foxlin E, Harrington M, Naimark L, and Wormell D (2001). "Mobile Augmented Reality Systems," *28th International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 2001), Conference Abstracts and Applications*, page 129

Feiner S, MacIntyre B, and Seligmann D (1992). "Annotating the real work with knowledge-based graphics on a see-through head-mounted display." *Graphics Interface '92*, pages 78-85

Feiner S, MacIntyre B, and Seligmann D (1993). "Knowledge-based augmented reality." *Communications of the ACM*, 36(7):52-62

Feiner S and Shamash A (1991). "Hybrid user interfaces: Breding virtually bigger interfaces for physically smaller computers." *ACM Symposium on User Interface Software and Technology,* pages 9-17

Feiner S, MacIntyre B, Höllerer T, and Webster T (1997). "A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment." *International Symposium on Wearable Computers*, pages 74-81

Feiner, S and Seligmann D (1992). "Cutaways and ghosting: Satisfying visibility constraints in dynamic 3D illustrations." *The Visual Computer*, 8(5–6):292–302.

Gabbard, JL, Swan II JE, Hix D, Lanzagorta M, Livingston MA, Brown D, and Julier SJ (2002). "Usability Engineering: Domain Analysis Activities for Augmented Reality Systems," *Stereoscopic Displays and Virtual Reality Systems IX*, SPIE Vol. 4660, pages 445–457

Gabbard, JL, Swan II JE, Hix D, Si-Jung K, and Fitch G (2007). "Active Text Drawing Styles for Outdoor Augmented Reality: A User-Based Study and Design Implications." *IEEE Virtual Reality*, pages 35–42

Güven S and Feiner S (2004). “A Hypermedia Authoring Tool for Augmented and Virtual Reality.” *The New Review of Hypermedia and Multimedia* 9:89-116

Hix D, Gabbard JL, Swan II JE, Livingston MA, Höllerer, T, Julier SJ, Baillot Y, and Brown D (2004). “A Cost-Effective Usability Evaluation Progression for Novel Interactive Systems”, *Hawaii International Conference on System Sciences (HICSS-37)*

Henderson S and Feiner S (2010). “Opportunistic Tangible User Interfaces for Augmented Reality.” *IEEE Transactions on Visualization and Computer Graphics*, 16(1):4–16

Henderson S and Feiner S (2011). “Exploring the Benefits of Augmented Reality Documentation for Maintenance and Repair.” *IEEE Transactions on Visualization and Computer Graphics* 17(10):1355-1368

Höllerer T, Feiner S, Hallaway D, Bell B, Lanzagorta M, Brown D, Julier S, Baillot Y, and Rosenblum L (2001). “User interface management techniques for collaborative mobile augmented reality,” *Computers and Graphics* 25(5):799-810

Höllerer T, Feiner S, and Pavlik J (1999). “Situated Documentaries: Embedding Multimedia Presentations in the Real World.” *International Symposium on Wearable Computers*, pages 79-86

Ioannidis J, Duchamp D, Maguire Jr GQ (1991). “IP-based Protocols for Mobile Internetworking.” *ACM SIGCOMM*, pages 235-245

Jones JA, Swan II JE, Singh G, Kolstad E, and Ellis SR (2008). “The Effects of Virtual Reality, Augmented Reality, and Motion Parallax on Egocentric Depth Perception.” *Symposium on Applied Perception in Graphics and Visualization*, pages 9–14

Julier S, Lanzagorta M, Baillot Y, Rosenblum L, Feiner S, Höllerer T, Sestito S (2000). “Information Filtering for Mobile Augmented Reality.” *IEEE International Symposium on Augmented Reality*, pages 3-11

Julier S, Baillot Y, Lanzagorta M, Brown D, and Rosenblum L (2000). “BARS: Battlefield Augmented Reality System.” *NATO Symposium on Information Processing Techniques for Military Systems*, pages 9-11

Julier S, Baillot Y, Lanzagorta M, Rosenblum LJ, and Brown D (2001). “Urban Terrain Modelling for Augmented Reality Applications.” In 3D Synthetic Environment Reconstruction, chapter 6, pages 119-136, Kluwer Academic Press

Julier S, Feiner S, and Rosenblum L (1999). “Augmented Reality as an Example of a Demanding Human-Centered System.” *First EC/NSF Advanced Research Workshop*

Livingston MA, Ai Z, Swan II JE, and Smallman HS (2009). “Indoor vs. Outdoor Depth Perception for Mobile Augmented Reality.” *IEEE Virtual Reality* pages 55–61

Livingston MA, Karsch K, Ai Z, and Gibson GO (2011). “User Interface Design for Military AR Applications,” *Virtual Reality* 15():175-184, Springer UK

Livingston MA, Lederer A, Ellis SR, White SM, and Feiner SK (2006). “Vertical Vergence Calibration for Augmented Reality Displays.” *IEEE Virtual Reality* (Poster Session)

Livingston MA, Rosenblum LJ, Julier SJ, Brown DG, Baillot Y, Swan II JE, Gabbard JL, and Hix D (2002) "An Augmented Reality System for Military Operations in Urban Terrain." In *Interservice/Industry Training, Simulation, and Education Conference*, page 89

Livingston MA, Swan II JE, Gabbard, JL, Höllerer TH, Hix D, Julier SJ, Baillot Y, and Brown DG (2003). "Resolving Multiple Occluded Layers in Augmented Reality", *2nd International Symposium on Mixed and Augmented Reality*, pages 56–65

Loomis JM, Klatzky RL, Golledge RG, Cicinelli JG, Pellegrino JW and Fry PA (1993). "Nonvisual navigation by blind and sighted: Assessment of path integration ability." *Journal of Experimental Psychology, General* 122(1):73-91

MacIntyre B, Coelho EM, and Julier SJ (2002). "Estimating and Adapting to Registration Errors in Augmented Reality Systems." *IEEE Virtual Reality*, pages 73-80

MacIntyre BM and Feiner S (1996). "Future Multimedia User Interfaces." *Multimedia Systems* 4(5):250-268

Rosenberg R., M. Lanzagorta, E. Kuo, R. King and L. Rosenblum (2000). "Immersive Scientific Visualization," *NRL Review*, 137-139

Rosenblum, L., Durbin J, Doyle R, and Tate D (1997). "Situational Awareness Using the VR Responsive Workbench." *IEEE Computer Graphics and Applications* 17(4):12-13

Seligmann D and Feiner S (1989). "Specifying composite illustrations with communicative goals." *ACM Symposium on User Interface Software and Technology*, pages 1-9

Seligmann D and Feiner S (1991). "Automated generation of intent-based 3D illustrations." *Computer Graphics* 25(4):123-132

Singh G, Swan II JE, Jones JA, and Ellis SR (2010). "Depth Judgment Measures and Occluding Surfaces in Near-Field Augmented Reality." *Symposium on Applied Perception in Graphics and Visualization*, pages 149–156

Sukan M and Feiner S (2010). "SnapAR: Storing Snapshots for Quick Viewpoint Switching in Hand-held Augmented Reality." *IEEE International Symposium on Mixed and Augmented Reality*, pages 273–274

Swan II JE, Jones JA, Kolstad E, Livingston MA, and Smallman HS (2007). "Egocentric Depth Judgments in Optical, See-Through Augmented Reality", *IEEE Transactions on Visualization and Computer Graphics* 13(3):429–442

Swan II JE, Livingston MA, Smallman HS, Brown DG, Baillot Y, Gabbard JL, and Hix D (2006). "A Perceptual Matching Technique for Depth Judgments in Optical, See-Through Augmented Reality", *IEEE Virtual Reality*, pages 19–26

Watsen K and Zyda M (1998). "Bamboo—A Portable System for Dynamically Extensible, Networked, Real-Time, Virtual Environments." *Virtual Reality Annual International Symposium*, pages 252-259

White S and Feiner S (2009). "SiteLens: Situated Visualization Techniques for Urban Site Visits." *ACM SIGCHI Conference on-Human Factors in Computing Systems*, pages 1117-1120