

[BY MARK BILLINGHURST AND HIROKAZU KATO]

COLLABORATIVE AUGMENTED REALITY

Blending reality and virtuality, these interfaces let users see each other, along with virtual objects, allowing communication behaviors much more like face-to-face than like screen-based collaboration.



Great science fiction often foreshadows great technical advances in communication technology. In Stanley Kubrick's and Arthur C. Clarke's film *2001: A Space Odyssey*, Dr. Haywood Floyd calls home using a video-phone, one of the earliest film appearances of videoconferencing. Little more than a decade later, in George Lucas's original *Star Wars*, remote collaboration is accomplished using life-size virtual images superimposed on the real world. Today, desktop videoconferencing is widely available, while life-size virtual teleconferencing remains far-off fiction.

Ironically, in 1965, just three years before *2001* was released, computer graphics pioneer Ivan Sutherland developed a display technology that made it possible to overlay virtual images on the real world. Attaching two head-worn miniature cathode ray tubes to a mechanical tracker, he created a see-through head-mounted display (HMD); users could see a simple virtual wire-frame cube superimposed on the world, creating the first augmented reality, or AR, interface. The term augmented reality is often used to refer to interfaces in which 2D and 3D computer graphics are superimposed on real objects. The graphics are typically viewed through opti-

(top) Augmented reality coupled with autofabrication technology is used to create a collaborative environment called Tangible Interfaces for Molecular Biology, a joint project of the Molecular Graphics Laboratory at The Scripps Research Institute, the HITLab at the University of Washington, and the Geometric Design and Computation Group at the University of Utah. Photo shows a mockup of an AR interaction between two proteins, calculating electrostatic interaction (Arthur Olson, The Scripps Research Institute, La Jolla, CA).

(left) Mixed reality conference room at the Communications Research Laboratory, Koganei, Japan; two scientists discuss a terrain model of Japanese mountain Yakedake, assisted by Shimadzu monocular head-mounted displays and 3rdTech wide-area optical 3D sensors (Kiyoshi Kiyokawa, Communications Research laboratory).

cal see-through head-mounted or handheld displays or combined with live video in video see-through mode. In practically all AR interfaces, 3D virtual objects appear to exist in the same space as real objects and the surrounding natural environment. Indeed, a long-term goal of AR researchers is to make it possible for the real and virtual worlds to blend together seamlessly, so they are indistinguishable from one another.

Researchers have developed single-user AR interfaces enabling people to interact with the real world in ways never before possible. For example, doctors can see virtual ultrasound images overlaid on a patient's body, giving them the equivalent of X-ray vision during a needle biopsy, while soldiers on the battlefield have personal head-up displays overlaying targeting information on the world around them. The technology is also appearing in commercial applications; for example, head-up displays in some aircraft cockpits overlay instrument readings on the view out the cockpit windows, and car maker General Motors is producing Cadillacs with an augmented night vision system overlaying infrared imagery on the road ahead. (For an exhaustive review of AR technology and applications, see [1].)

Although single-user AR applications show great promise, especially for industrial and medical applications, researchers have begun to explore how AR can be used to enhance face-to-face and remote collaboration as well. Here, we describe several prototype collaborative AR interfaces, the lessons learned from them, and the challenges that must be addressed before they are widely implemented.

Collaborative Technology

In natural face-to-face collaboration, people use speech, gesture, gaze, and nonverbal cues to attempt

to communicate. In many cases, the surrounding physical world and objects also play an important role, particularly in design and spatial collaboration tasks. Real objects support collaboration through their appearance, physical affordances, such as size and weight, use as semantic representations, and ability to create reference frames for communication. In

contrast, most computer interfaces for collocated collaboration create an artificial separation between the real world and the shared digital task space. People looking at a projection screen or crowded around a desktop monitor are often less able to refer to real objects or use natural communication behaviors. Observations of the use of large shared displays have found that simultaneous interaction rarely occurs due to the lack of software support and input devices for co-present collaboration [10].

AR technology promises to enhance such face-to-face communication. AR interfaces blend the physical and virtual worlds so real objects can interact with 3D digital content and improve users'

shared understanding. Tangible interaction methods [6] can be combined with AR display techniques to develop interfaces in which physical objects and interactions are as important as the virtual imagery. Such interfaces naturally support face-to-face collaboration; for example, in a collaborative urban design application users might sit at a real table and see virtual building models appear in the middle of the table. If these buildings were then attached to physical objects, the users could pick up and place the buildings on a real street map.

Technology for remote collaboration also involves limitations. For example, it is difficult for current technology to provide remote participants with the same experience they would have if they were in a

Figure 1. A collaborative AR interface.



TANGIBLE interaction methods can be combined with AR display techniques to develop interfaces in which physical objects and interactions are as important as the virtual imagery.

Figure 2. Live virtual video avatars for remote collaborative AR interfaces.



co-located meeting. Audio-only interfaces remove the visual cues vital for conversational turn taking, leading to increased interruptions and overlap, difficulty disambiguating between speakers and determining another's willingness to interact. With conventional videoconferencing, subtle user movements or gestures cannot be captured, there are few spatial cues among participants, the number of participants is limited by monitor resolution, and participants cannot readily make eye contact. Speakers also cannot know when people are paying attention to them or when it might be permissible to hold side conversations.

Researchers have begun exploring how desktop and immersive collaborative virtual environments might provide spatial cues to support group interaction. These interfaces restore some of the spatial cues common in face-to-face conversation but require users to enter a virtual world separate from their physical environment. In addition, it is difficult to transmit the equivalent fidelity of nonverbal communication cues present in a face-to-face meeting and so create the same sense of presence. The Office of the Future, developed at the University of North Carolina, is perhaps closest to the goal of perfect remote telepresence [11]. Multiple cameras are used to capture and reconstruct a virtual geometric model and live video avatar. Although the work is impressive, it shares the common limitations of projection-based interfaces in that the display is not portable, and virtual objects appear only relative to the projection surface.

AR technologies can also provide spatial audio and visual cues to overlay a person's real environment and support remote collaboration. In this way, the remote

participants are added to the users' real world rather than separating them from it. AR technology again provides a seamless blending of reality and virtuality.

For Face-to-Face Collaboration

AR can be used to enhance a shared physical workspace, creating an interface for intuitive 3D computer-supported collaborative work (CSCW). Figure 1 shows a typical AR interface for face-to-face collaboration; users are seated across a table wearing HMDs with cameras attached. In their displays, they see live video of the real world with graphics superimposed. Using AR software, they can view a 3D virtual building superimposed over the table, while seeing each other at the same time.

One of the earliest interfaces to demonstrate the use of AR for face-to-face collaboration was the Studierstube project at Vienna University of Technology, Austria [12]. First reported in 1996, the project used see-through HMDs to allow users to collaboratively view 3D virtual models superimposed on the real world. Users found the interface intuitive and conducive to collaboration, because unlike other interfaces, the groupware support can be kept simple and left mostly to social protocols. Similarly, the Japanese Mixed Reality Systems Laboratory's AR2 Hockey interface allows two users to play a game of air hockey using a virtual puck [9]. Users compete with each other, interacting with the virtual content at the same time. The Studierstube researchers identified five key attributes of collaborative AR environments:

Virtuality. Objects that don't exist in the real world can be viewed and examined.

Augmentation. Real objects can be augmented with virtual annotations.

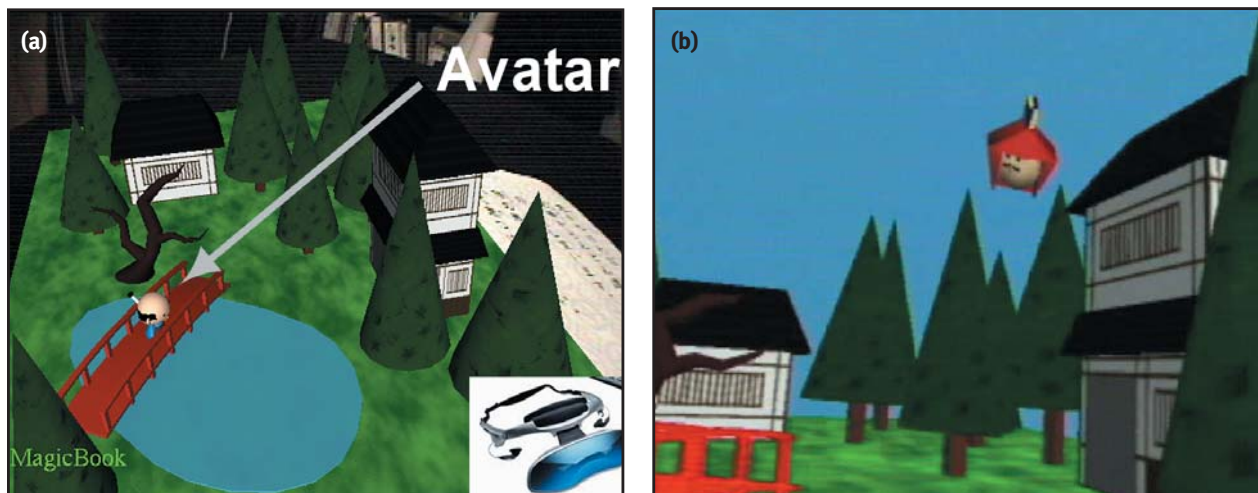
Cooperation. Multiple users can see each other and cooperate in natural ways.

Independence. Individual users control their own independent viewpoints.

condition in which to work together. The performance improvement was largely the result of the improved perception of nonverbal cues, as supported by the collaborative AR interface.

Similarly, collaborative AR interfaces can produce communication behaviors that are more like unmedi-

Figure 3. Multiscale collaboration in the MagicBook interface; (a) immersed user in an AR scene; (b) AR user as seen by the immersed user.



USERS felt they could pick up and move objects as easily in the AR condition as in the face-to-face condition and significantly easier than in the projection condition.

Individuality. Displayed data can appear in different form for individual viewers depending on their personal needs and interests.

Perhaps most important is the seamless nature of collaborative AR interfaces. Users see each other at the same time they see virtual objects in their midst. Unlike some other CSCW technologies, co-located AR interfaces do not separate the communication space from the task space, allowing users to interact with virtual content by using familiar real objects. Seamlessness is a key characteristic of successful CSCW interfaces [5].

The value of seamlessness has been confirmed by several user studies in recent years comparing collaborative AR interfaces to other technologies. One study [8] measured performance on a pointing task in an AR interface to the same task in an immersive VR setting. Subjects performed significantly faster in the AR interface while feeling this was the most natural

ated face-to-face collaboration than screen-based collaboration. In a recent experiment, we compared communication behaviors used to complete a two-person logic puzzle [4], testing three conditions:

- Face-to-face collaboration with real objects;
- Co-located AR collaboration with virtual objects; and
- Co-located projection-screen-based collaboration with virtual objects.

The virtual objects were exact copies of the real objects; in the AR case, they were attached to real objects, making it possible to use tangible manipulation techniques. We videotaped user behavior and analyzed transcriptions of the speech and gestures the test subjects used to solve the puzzles. Although they did not find the face-to-face and AR conditions to be the same, they exhibited speech and gesture behaviors more alike in these conditions than in the

screen-based condition. Users felt they could pick up and move objects as easily in the AR condition as in the face-to-face condition and significantly easier than in the projection condition.

Although these results are encouraging for showing that communication behaviors with an AR interface are similar to natural face-to-face communication, they are just the beginning of a deeper exploration. Our first user studies focused on object-centered collaboration where users actively engage in working together. However, previous work on teleconferencing found that negotiation and conversational tasks are even more sensitive to differences between communication media. Future experiments need to include object manipulation as only a small part of the task at hand.

For Remote Collaboration

AR technology can also be used to support remote collaboration. In an AR conferencing interface we developed in 1998, a user wore a lightweight HMD (with camera) and could see a virtual image of a remote collaborator attached to a physical card as a life-size, live virtual video window [2]. We used computer-vision techniques to track black squares on the card, ensuring the virtual video appeared precisely aligned with the real object. The overall effect was that the remote collaborator sitting at a desktop computer appeared projected into the local user's real workspace (see Figure 2).

A number of other significant factors differentiate this type of conferencing interface from traditional desktop videoconferencing. Users could arrange the cards on any surface to create a virtual spatial conferencing space; the cards were also small enough to be carried easily, ensuring portability. This interface frees users from the desktop, potentially enabling conferencing from any location; thus, remote collaborators become part of any real-world surroundings, increasing the sense of social presence. Remote users can appear as life-size images; a potentially arbitrary number of them can participate and be seen simultaneously. Since the virtual video windows can be placed about the user in space, spatial cues can be restored to the collaboration. Finally, because remote-user images are entirely virtual, a real camera can be placed at user eye level, allowing support for natural gaze cues.

In a 1999 user study we conducted comparing AR conferencing to traditional audio- and videoconferencing, subjects reported a significantly stronger sense of presence for their remote counterparts in the AR conferencing condition, and that it was easier to perceive one another's nonverbal communication cues [2]. More recently, we developed an interface sup-

porting multiple remote users, applying alpha mapping techniques to extract video of users from the background, as in Figure 2, right. In our related study [7], test subjects felt the AR condition provided significantly more co-presence and improved their personal understanding of the conversational relationships between participants.

Multiscale Collaboration

AR techniques can also be used to support multiscale collaboration, where users collaboratively view a data set from different viewpoints. We explored this in our MagicBook work [3], using a real book as an AR interface object. Several readers look at the same book and share the story. If they then pick up handheld AR displays, they see the virtual models superimposed over the book pages from their own viewpoints. Since they can see each other and the real world simultaneously (as virtual models), they can readily use normal face-to-face communication cues. Individual users of the MagicBook interface have their own independent view of the content; any number of people can view and interact with a virtual model as easily as they interact with a real object.

The interface also supports collaboration on multiple scales. Users can "fly into" AR scenes, experiencing them as immersive virtual environments. Multiple users can be immersed in the same virtual scene, seeing each other represented as virtual characters. More interesting, one or more users can be immersed in the virtual world while others are viewing its content as an AR scene. Here, the AR user sees exocentric views of miniature figures of the immersed users (see Figure 3a); in the immersive world, users viewing the AR scene appear as large virtual heads looking down from the sky (see Figure 3b). Thus a group of collaborators can share both egocentric and exocentric views of the data set, leading to greater understanding of the virtual content.

Research Challenges

A number of challenges must be overcome before AR technology is widely used for collaboration. The display problem is paramount. Gaze provides an important nonverbal cue in normal face-to-face and remote collaboration, yet current-generation HMDs cover the user's eyes. In our AR conferencing interface in Figure 2, while AR users see the eyes of their remote collaborators (at desktop computers), the desktop users see the AR users with their eyes hidden. And in face-to-face collaborative AR interfaces, each user's eyes are covered. We are addressing this problem by developing handheld displays that are

less encumbering than HMDs. For example, a flat panel liquid crystal display can be used as a window into an AR environment, allowing users to see AR content at the same time they see each other's facial expressions. Alternatively, manufacturers of commercial HMDs, such as MicroOptical, are developing displays virtually indistinguishable from a normal pair of glasses (see www.microopticalcorp.com).

A second problem with current displays is that viewing the world through them is not the same as seeing it with the naked eye. Current HMDs have limitations in their field of view, resolution, and color depth. Optical see-through HMDs allow users to view the world normally, but it is difficult to build see-through displays with a wide field of view. Conversely, with video see-through displays the world is seen through a camera lens, thus introducing some of the same problems as associated with traditional videoconferencing.

Another research challenge is the problem of tracking and registration. In order for virtual models to be overlaid precisely on the real world, users' viewpoints need to be tracked. In our applications we use computer-vision-based tracking techniques appropriate for video see-through AR. However, these methods work only when physical tracking markers are in view and may introduce more system delays compared to other magnetic, ultrasonic, and inertial-tracking technologies. Research needs to address which tracking technologies are most appropriate for collaborative AR interfaces. Hybrid tracking approaches combining several techniques (such as vision-based and inertial) are likely to represent a particularly fruitful direction.

Finally, more formal user studies are needed to evaluate the effect on collaboration of AR technologies and explore intuitive interaction techniques. The field of collaborative AR is in many ways at the same point videoconferencing was 25 years ago. It is technically possible to develop collaborative AR systems; still unknown is how such systems can best be used to enhance face-to-face and remote communication.

Conclusion

AR techniques can be used to explore different types of interfaces for face-to-face and remote collaboration thanks to the following properties:

- The ability to enhance reality;
- Seamless interaction between real and virtual environments;
- The presence of spatial cues for face-to-face and remote collaboration; and

- The ability to support multiscale collaboration.

We have described several examples of the interfaces that can be produced from taking advantage of these characteristics. Despite early promising results, a lot of research work needs to be done before collaborative AR interfaces are as well-understood as traditional telecommunication technology. Better display and input devices are needed. Rigorous user studies must be conducted on a variety of tasks and interface types. Hybrid interfaces integrating AR technology with other collaborative technologies need further exploration. When these areas have been addressed, George Lucas's vision of teleconferencing will be tantalizingly close. **G**

REFERENCES

1. Azuma, R., Bailiot, Y., Behringer, R., Feiner, S., Julier, S., and MacIntyre, B. Recent advances in augmented reality. *IEEE Comput. Graph. Appl.* 21, 6 (Nov./Dec. 2001), 34–47.
2. Billinghurst, M., and Kato, H. Out and about: Real-world teleconferencing. *British Telecom Tech. J.* 18, 1 (Jan. 2000), 80–82.
3. Billinghurst, M., Kato, H., and Poupyrev, I. The MagicBook: A transitional AR interface. *Comput. Graph.* 25 (2001), 745–753.
4. Billinghurst, M., Kato, H., Kiyokawa, K., Belcher, D., and Poupyrev, I. Experiments with face-to-face collaborative AR interfaces. *Virtual Reality J.* 4, 2 (2002).
5. Ishii, H., Kobayashi, M., and Arita, K. Iterative design of seamless collaboration media. *Commun. ACM* 37, 8 (Aug. 1994), 83–97.
6. Ishii, H., and Ullmer, B. Tangible bits: Towards seamless interfaces between people, bits, and atoms. In *Proceedings of CHI'97* (Atlanta, Mar. 22–27). ACM Press, New York, 1997, 234–241.
7. Kato, H., Billinghurst, M., Morinaga, K., and Tachibana, K. The effect of spatial cues in augmented reality video conferencing. In *Proceedings of the 9th International Conference on Human-Computer Interaction* (New Orleans, Aug. 5–10). Lawrence Erlbaum Associates, NJ, 2001, 478–481.
8. Kiyokawa, K., Takemura, H., and Yokoya, N. SeamlessDesign for 3D object creation. *IEEE MultiMedia* 7, 1 (Jan. 2000), 22–33.
9. Ohshima, T., Satoh, K., Yamamoto, H., and Tamura, H. AR2 hockey: A case study of collaborative augmented reality. In *Proceedings of the IEEE Virtual Reality Annual International Symposium (VRAIS'98)* (Atlanta, Mar. 14–18). IEEE Computer Society Press, Los Alamitos, CA, 1998, 268–275.
10. Pedersen, E., McCall, K., Moran, T., and Halasz, F. Tivoli: An electronic whiteboard for informal workgroup meetings. In *Proceedings of Human Factors in Computing Systems (InterCHI'93)* (Amsterdam, The Netherlands, Apr.). ACM Press, New York, 1993, 391–398.
11. Raskar, R., Welch, G., Cutts, M., Lake, A., Stesin, L., and Fuchs, H. The Office of the Future: A unified approach to image-based modeling and spatially immersive displays. In *Proceedings of SIGGRAPH'98* (Orlando, FL, July 19–24). ACM Press, New York, 1998, 179–188.
12. Schmalstieg, D., Fuhrmann, A., Szalavri, Z., and Gervautz, M. Studierstube: An environment for collaboration in augmented reality. In *Proceedings of the Collaborative Virtual Environments (CVE'96) Workshop* (Nottingham, Great Britain, Sept. 19–20); www.studierstube.org.

MARK BILLINGHURST (grof@hitl.washington.edu) is a research scientist in the Human Interface Technology Laboratory at the University of Washington, Seattle.

HIROKAZU KATO (kato@sys.im.hiroshima-cu.ac.jp) is an associate professor in the Department of Information Machines and Interfaces at Hiroshima City University, Japan.
