Real World Teleconferencing

Mark Billinghurst[a], Adrian Cheok[b], Hirokazu Kato[c], Simon Prince[b]

[a]HIT Lab New Zealand
[b]National University of Singapore
[c]Hiroshima City University

> *If, as it is said to be not unlikely in the near future, the principle of sight is applied to the telephone as well as that of sound, earth will be in truth a paradise, and distance will lose its enchantment by being abolished altogether.*
>
> Arthur Strand, 1898

Twenty years after the invention of the telephone, Strand foresaw how technology could be used to enable communication with anyone, anywhere and anytime. The current growing use of mobile phones and spread of wireless service are making this vision a reality. In many ways cell phones are just smaller, portable versions of a hundred year old invention. Following the same trend, desktop video conferencing will soon be available on mobile devices. However the new form factor creates both opportunities and challenges. Video conferencing on a handheld device with small screen and limited input may be a less than satisfying experience.

There are many research groups addressing this problem. In our work we have been exploring how augmented reality technology can create fundamentally new forms of remote collaboration for mobile devices. Augmented Reality (AR) involves the overlay of virtual graphics and audio on reality. Typically, the user views the world through a handheld or head mounted display (HMD) that is either see-through or overlays graphics on video of the surrounding environment. Unlike other computer interfaces that draw users away from the real world and onto the screen, AR interfaces enhance the real world experience. Doctors can see ultrasound information on the patient, while scientists could view virtual mathematical models superimposed on the space between them.

AR Remote Collaboration

Research on AR interfaces for remote conferencing has largely been in the area of sharing remote workspaces. A collaborator can look from someone else's viewpoint at the real world and place virtual annotations there to help them with the task at hand [Kuzuoka 92]. Unlike this work, in our research we are focusing on the communication aspects that the technology affords. We want users to be able to see their distant friends in their real environment with them.

Our early work used a wearable computer, internet telephony and a see-through head mounted display to create a wearable conferencing space [Billinghurst 98]. In the WearCom interface, when a remote collaborator called, the user would see a static virtual picture of the person superimposed over the real world and hear spatialized audio coming from their virtual location. Several users could connect at once, enabling the user to be surrounded by virtual images of their friends, and hearing their speech surrounding them. In our evaluation studies, users found that they could easily separate out simultaneous speakers because of the spatial visual and audio cues that the AR interface provided.

However the wireless bandwidth and CPU power was not sufficient to provide a live virtual video view of the remote collaborators, so in many ways this interface was little better than adding spatial audio to a mobile phone. More recently we have developed interfaces that do overlay virtual video of remote participants on the real world. In our AR Conferencing work [Billinghurst 99][Kato 01] a user wears a video see-through head mounted display connected to a computer. When they look at a real name card they can see a live-sized virtual video image of a remote collaborator (see figure 1). A black tracking pattern printed on the card is used by our ARToolKit computer vision software[1] to track the users viewpoint in real time and precisely overlay virtual video on the card. The cards are physical representations of remote participants. Users can arrange multiple cards about them in space to create a virtual meeting space and the cards are small enough to be easily carried, ensuring portability.

Figure 1: AR Conferencing – live virtual video on the real world

There are a number of other significant differences between this and traditional video conferencing. The remote user can appear as a life-sized image, and a potentially arbitrary number of remote users can be viewed at once. Since the virtual video windows can be placed about the user in space, spatial cues can be restored to the collaboration. Finally, the remote user's image is entirely virtual, so a real camera could be placed at the user's eye point, allowing support for natural gaze cues.

Perhaps the greatest difference is that the user is no longer tied to the desktop and can conference from any location. This means that the remote collaborators can become part of any real-world surroundings, potentially increasing the sense of social presence. In our evaluation studies, users felt that their remote collaborator was more present in the real world with them than in more traditional video and audio conferencing interfaces. With several remote collaborators (figure 1) users found it easy to know which participants were speaking and to whom

However in this application the live video was texture mapped onto a flat polygon, so the remote users appeared flat. This made it difficult to convey some non-verbal cues such as pointing and body motion. Our most recent work overcomes this limitation. The 3D-Live interface [Prince 2002], being

[1] Available for free download from http://www.hitl.washington.edu/artoolkit/

developed at the National University of Singapore, uses multiple video cameras to capture a remote participant and then generates a virtual image of that person from the viewpoint of the AR user. In the HMD, the user can see the remote participant as if they were really standing in front of them (see figure 2). They can move around them (or turn them in their hands) to see from all sides. In this way the non-verbal communication cues are easily transmitted.

Figure 2. 3D Live interface

The 3D live system has two main components (figure 3); video capture and viewpoint generation hardware and software, and an AR conferencing client. The capture hardware and software is a commercial system manufactured by Zaxel[2]. Fifteen video cameras are placed surrounding the remote collaborator. The video from these cameras is sent to five powerful computers that segment out the person from the background and then pass on images to a viewpoint server machine. On this computer, the Zaxel software implements a real-time novel view generation algorithm, based on shape-from-silhouette information. When told which viewpoint the remote person should be seen from, the software generates a synthetic view from that viewpoint, using pixels from the live video streams. The software can run at 30 frames per second and returns an image of up to 640x480 pixel resolution.

[2] See www.zaxel.com

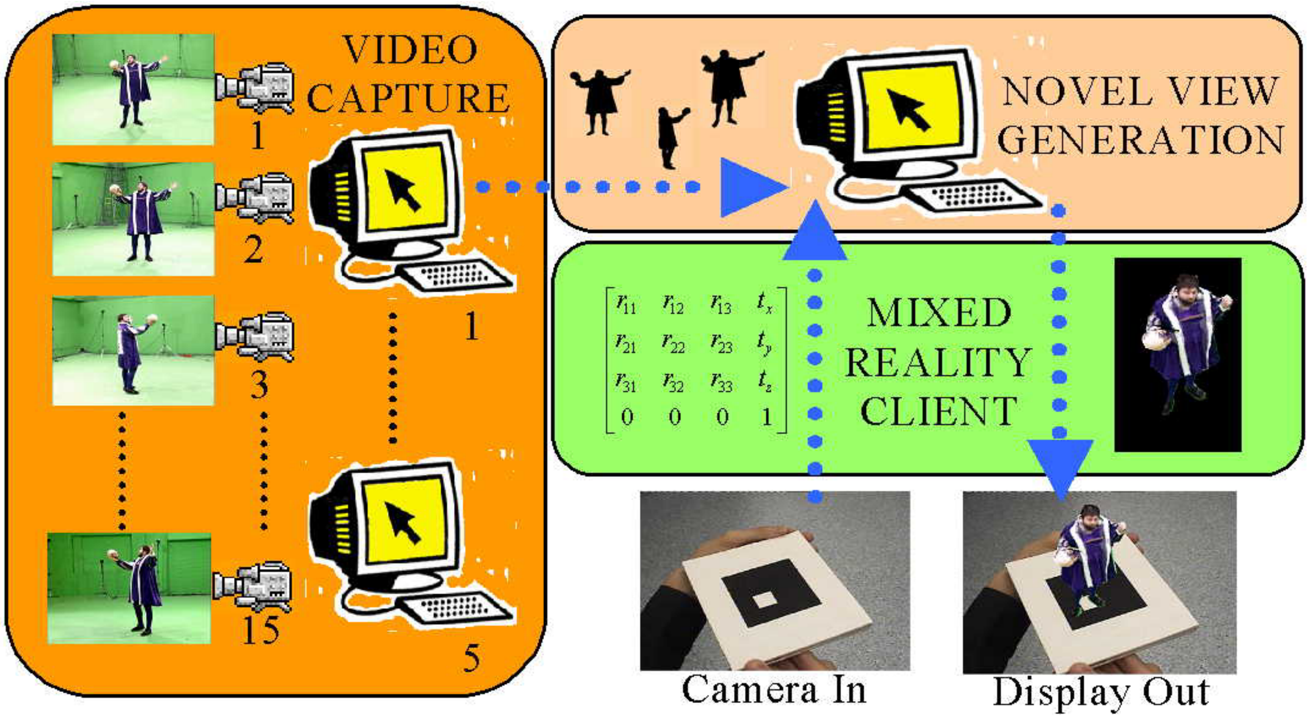

Figure 3: 3D Live system specification

In the AR interface, as the user looks at the tracking marker, the ARToolKit library is used to calculate their viewpoint which is sent to the Zaxel system. This returns an image of the remote user from that position which is then overlaid on the real world. With high enough network bandwidth the views of the remote user can be updated at more than 20 frames per second, creating an extremely compelling sense of presence. Although the capture system required extensive CPU power and capture hardware, the AR client can run on a single (potentially mobile) computer with an AR display attached. Furthermore, a new capture system that only requires six cameras and low computing power is currently under development.

Conclusions

We have described on-going work in exploring how AR technology can be used to explore interfaces for the mobile communication devices of the future. Our systems have evolved from static images, to live virtual video textured onto a flat polygon to an image-based rendering technique that allows a live view of the remote collaborator to be generated from any viewpoint. Each of these systems provides a high degree of realism and remote presence.

However these systems are highly experimental, and in many ways are like the early video conferencing systems of 30 years ago. Rigorous user studies need to be conducted to evaluate the effect on communication of using such systems, and to determine how to use AR technology to provide the best collaborative experience.

References

[Billinghurst 98] Billinghurst, M., Bowskill, J., Morphett, J. (1998) WearCom: A Wearable Communications Space. British Telecom Technical Journal, October 1998.

[Billinghurst 99] Billinghurst, M., Kato, H., (1999) Real World Teleconferencing. In Proceedings of the conference on Human Factors in Computing Systems (CHI 99). May 15th-20th, Pittsburgh, USA, 1999.

[Kato 01] Kato, H., Billinghurst, M., Morinaga, K., Tachibana, K. (2001) The Effect of Spatial Cues in Augmented Reality Video Conferencing. In Proceedings of the 9th International Conference on Human-Computer Interaction (HCI International 2001), August 5-10th, New Orleans, LA, USA, 2001.

[Kuzuoka 92] Hideaki Kuzuoka. *Spatial workspace collaboration: A SharedView video support system for remote collaboration capability*. In Proceedings of CHI'92, pp. 533--540, 1992.

[Prince 02] Prince, S., Williamson, T., Cheok, A., Farbiz, F., Billinghurst, M., Kato, H. 3-D Live: Real-Time Interaction for Mixed Reality. In Proceedings of the ACM Conference on Computer Supported Collaborative Work (CSCW 2002), Nov. 16-20th, New Orleans, Louisiana, 2002