GAZE-2: An Attentive Video Conferencing System

Roel Vertegaal **Ivo Weevers** **Changuk Sohn**

Human Media Lab, CISC
Queen's University
Kingston, ON K7L 3N6
Canada
{roel, ivo, changuk}@cs.queensu.ca

ABSTRACT
GAZE-2 is an attentive video conferencing system that conveys whom users are talking to by measuring whom a user looks at and then rotating his video image towards that person in a 3D meeting room. Attentive Videotunnels ensure a parallax-free image by automatically broadcasting the feed from the camera closest to where the user looks. The system allows attentive compression by reducing resolution of video and audio feeds from users that are not being looked at.

KEYWORDS: Attentive Interfaces, Video Conferencing, Gaze, Eye Tracking, CSCW.

INTRODUCTION
With the recent resurge of interest in video conferencing as a means to conduct business meetings without travel comes a renewed interest in the usability of this technology in supporting group conversations. Several studies have reported problems with multiparty turn taking using video mediated systems. Isaacs and Tang [2] performed a usability study of a group of five participants using a desktop video conferencing system. They found that during video conferencing users addressed each other by name, and started explicitly requesting individuals to take turns. In face-to-face interaction, they found users used their eye gaze to indicate whom they were addressing or to suggest a next speaker. When more than one person started speaking simultaneously, the next speaker was typically determined by the gaze of the previous speaker. Similarly, O'Connaill et al. [4] found that in video conferencing more formal techniques were used to achieve speaker switching than in face-to-face interaction. They too attribute this to the absence of gaze-related speaker-switching cues. Similarly, Vertegaal et al. [9] showed that when eye gaze is *not* conveyed during triadic video mediated conversations, subjects take 25% less turns. 88% of subjects indicated they had trouble perceiving whom their partners were talking to when eye gaze was not conveyed. There are at least two reasons why most video mediated systems do not convey eye gaze signals of users. Firstly, cameras are typically positioned on top of a monitor that displays the participant's eyes, yielding a parallax between those eyes and the camera. Although the nature of the parallax depends on the distance to the screen and the type of camera lens used, typically, when a user looks at the eyes of a partner this is not perceived as eye contact by that partner [8, 9].

CHI 2002, April 20-25, 2002, Minneapolis, Minnesota, USA.
ACM 1-58113-454-1/02/0004.

Secondly, most systems broadcast images of a user from a single camera unit to all other users. Users are therefore not capable of establishing eye contact with a single person. In this paper, we present GAZE-2, an attentive video conferencing system that conveys eye contact using a combination of novel attentive interface technologies. GAZE-2 adds motion video capabilities to the original GAZE Groupware System [8]. The system measures whom each participant looks at using a built-in low-cost eye tracker based on IBM's Blue Eyes technology [3]. Using the eye tracker, the system automatically selects one of three camera units placed inside each user's videotunnel for parallax-free transmission of the user's frontal video image. The video images of users are represented as 2D planes suspended in a 3D virtual meeting room, and automatically orient themselves towards the attended user (see Figure 1). First, we will provide an overview of previous systems that mediated gaze direction. We will then discuss the design of our attentive videotunnel technology, and of our attentive 3D meeting room. Finally, we will discuss how we achieve a reduced network load using attentive compression. Although we are still in the process of evaluating the above technologies, preliminary studies suggest that the above combination of technologies provides improved support for turn management in mediated group conversations.

PREVIOUS WORK
Over the years a number of solutions have been developed to convey eye gaze during multiparty video conferencing. Hydra [7] simulated a four-way round-table meeting by placing boxes containing a camera, a small monitor and speaker on a table in places that would otherwise be held by each remote participant. Each person is therefore presented with his own view of each remote participant. However, Hydra still suffered from a considerable parallax [7, 8]. Also, since each user required a unique video feed from each other user, the system did not allow multicasting. In the MAJIC system [5], cameras were placed behind the eyes of users projected on a semi-transparent life-size screen. Although MAJIC exhibited a greater tolerance to parallax due to the large distance of users to the camera units, when users would move on the screen, the cameras would need to be repositioned to achieve a parallax-free image. In general, the kind of videotunnel technology developed by Rosenthal [6] only works when users keep their heads still. When the eyes of a person projected on the screen are no longer aligned with his camera, eye contact is no longer conveyed.

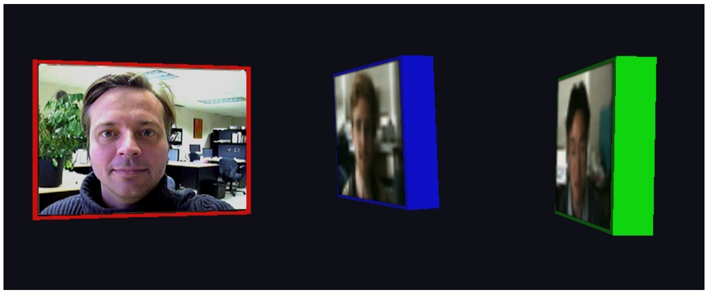

Figure 1. *GAZE-2 session with 4 users. Everyone is currently looking at the left person, who's image is broadcast in a higher resolution. Frontal images are rotated to convey gaze direction.*

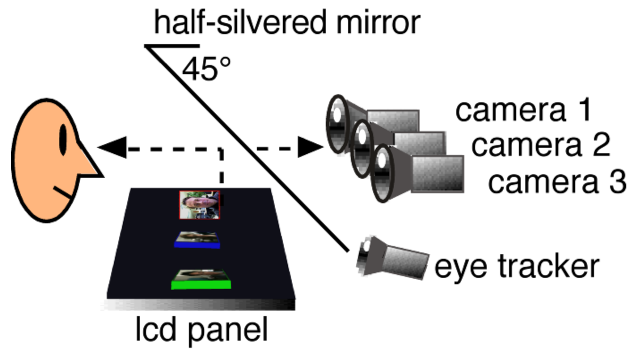

Figure 2. *Attentive Videotunnel. To prevent horizontal parallax, an eyetracker mounted in a video tunnel selects the camera closest to the gaze position of the user for broadcast.*

ATTENTIVE VIDEOTUNNELS

By positioning a half-silvered mirror at a 45 degree angle to the screen, videotunnels allow placement of camera units behind a projection of a virtual screen in front of the user (see Figure 2). In our Attentive Videotunnel design, several cameras are placed behind the screen, such that one of these cameras will always provide a parallax-free frontal image of the user. An eye tracker mounted below the cameras measures which camera the user is looking at, and selects this camera for multicasting to all users.

ATTENTIVE MEETING ROOMS

The frontal, parallax-free image of each user is projected on a 2D plane suspended in a 3D virtual meeting room at locations that would otherwise be held in a round-table meeting (see Figure 1). The eye tracking system measures not only which camera is closest to the user's gaze position, but also which participant the user looks at. To allow users to establish eye contact with the correct participant, each user's image is automatically rotated towards the participant that user is looking at. Such rotation provides a surprisingly powerful affordance of head orientation by the corresponding user. Firstly, like head orientation, the projected surface of a face shrinks with rotation. Secondly, since interlocutors typically establish eye contact with the person they listen or speak to, the limited resolution of peripheral vision strengthens the illusion of head orientation by unattended individuals. Like the Hydra system, GAZE-2 does allow the presentation of images from different camera angles to different users. We chose not to do this to allow the use of a single camera feed for multicasting.

NETWORKS OF ATTENTION

Network bandwidth use is a key aspect of the usability of video conferencing systems. In internet-based video conferencing systems, unavailability of network resources leads to poor image quality through high compression rates, low frame rates, or decreased resolution. Multicasting alleviates network problems by allowing a single video stream to be sent to all users simultaneously, occupying only a single unit of network bandwidth. When GAZE-2 is used by 4 users in multicast mode, only 4 video streams need to be sent over the network. A system that uses images from multiple camera angles to convey eye gaze would require 12 individual streams. However, when multicasting is unavailable, the use of a single camera feed yields little or no bandwidth gains, as images need to be sent individually anyway. To increase network efficiency in such cases, our system allows the employment of attentive compression techniques [1]. A considerable amount of bandwidth is wasted during any multiparty video conferencing session. This is because each user is capable of looking only at a single person at a time. Given that most users typically focus on the same individual, it makes sense to dynamically focus network resources on that individual. Since the system knows where each user is looking, it sends high quality video images to only those participants that look at you. As exemplified by Figure 1, this need not be noticeable by the human eye. In Figure 1, only the left image is sent in hi-res. Try focusing on that image, with your head at a distance of 10 inches. When you now try observing the other users without moving your eye, you notice the limited resolution of your peripheral vision. We employ a similar technique for the attentive compression of audio streams.

CONCLUSIONS

We presented GAZE-2, an attentive video conferencing system that conveys whom users are talking to by measuring whom they look at, and then rotating their live video images towards that person in a 3D virtual meeting room. Attentive Videotunnels ensure a parallax-free image at all times by automatically broadcasting the camera closest to where the user is looking. The system allows the use of attentive compression of video and audio feeds, rendering only those users that are being looked at in high resolution.

REFERENCES

1. Duchowski, A. & McCormick, B. Gaze-Contingent Video Resolution Degradation. In *Human Vision and Electronic Imaging III*. SPIE, 1998.
2. Isaacs, E. and Tang, J. What video can and can't do for collaboration. In *Proceedings of ACM Multimedia '93.* Anaheim: ACM, 1993.
3. Morimoto, C. et al. Pupil Detection and Tracking Using Multiple Light Sources. Image and Vision Computing, vol 18, 2000.
4. O'Connaill, B., Whittaker, S., and Wilbur, S. Conversations Over Video Conferences: An Evaluation of the Spoken Aspects of Video-Mediated Communication. *Human Computer Interaction* 8, 1993.
5. Okada, K. et al. Multiparty Videoconferencing at Virtual Social Distance: MAJIC Design. In *Proceedings of CSCW '94.* Chapel Hill, 1994.
6. Rosenthal, A.H. Two-way television communication unit. US Pat. 1947.
7. Sellen, A.J. Remote conversations: the effects of mediating talk with technology. *Human Computer Interaction* 10(4), 1995.
8. Vertegaal, R. The GAZE Groupware System. In *Proceedings of CHI'99.* Pittsburg, PA: ACM, 1999.
9. Vertegaal, R. et al. Effects of Gaze on Multiparty Mediated Communication. In *Proceedings of Graphics Interface 2000.* Montreal, 2000.