Concept and Partial Prototype Video: Ubiquitous Video Communication with the Perception of Eye Contact

Emmanuel Munguia Tapia Stephen S. Intille John Rebula Steve Stoddard
Massachusetts Institute of Technology
1 Cambridge Center, 4FL
Cambridge, MA 02142 USA
emunguia | intille @mit.edu

ABSTRACT
This concept and partial prototype video introduces a strategy for creating a video conferencing system for future ubiquitous computing environments that can guarantee two remote conversants the ability to establish eye contact. Unlike prior work, eye contact can be achieved even as people move about their respective environments engaging in everyday tasks.

1. INTRODUCTION
Imagine this scenario: *You are working in your future kitchen. You turn on your video conferencing system and make a call to a friend in a distant city. The face of your friend appears just above your work surface on your kitchen cabinet door. When you look at your friend, he perceives excellent eye contact with you. No matter how you move about your space, when the conversation warrants it you can look at a nearby flat surface, see the face of your friend, and initiate eye contact. Establishing eye contact does not restrict you or your friend to standing or sitting in a particular position, chair, or space.* In this video we show how such a system could be achieved.

2. VIDEO CONFERENCING AND EYE CONTACT
Eye gaze conveys important signals about trust and attention during interpersonal communications [1]. A camera that is placed above or to the side of a display of a videoconferencing system forces the person looking at the display (the looker) to look away from the camera's optical path. The displacement angle, θ, between the optical path of the looker's eyes and the camera will result in a perception of lack of eye contact if θ is sufficiently large.

Chen [4] enumerated the three options previously investigated to minimize or eliminate θ. The first is selectively warping the video so that it appears to be captured from the viewpoint in front of the looker's eyes. This strategy can create unnatural faces and enforce eye contact when it should not be perceived. The second approach to minimize θ is to merge the optical path of the camera and the display, which typically requires two conversants sit in predetermined positions in front of their displays with their heads roughly centered in the respective camera's field of view. The third approach to minimize θ is to simply mount the camera sufficiently close to the display so that the display and camera have nearly the same optical path. This can be effective with small, PDA-sized displays [3] but fails with larger (e.g. desktop) size displays. This video shows a fourth option that combines the positive properties of each of these approaches: manipulating the video (via warping and object tracking) to guarantee that the optical path of the looker's eyes and the observer's face are closely aligned by ensuring that the camera is always positioned *in* the display directly between the observer's eyes.

3. A SYSTEM FOR UBIQUITOUS EYE CONTACT
We introduce a design for a video conferencing system that can be installed in existing environments in a space-efficient manner and that permits ubiquitous video communications with the perception of eye contact. Wireless phones freed people to move about their homes and accomplish cognitively simple tasks as they talk; our proposed video conferencing system would permit this type of casual interaction while maintaining the option of effortlessly establishing eye contact when needed. We achieve this by simultaneously exploiting several technologies: perspective warping of imagery, computer vision head tracking, and miniature (i.e., pinhole) video cameras embedded within walls and cabinetry.

3.1 Concept
Figure 1 illustrates the full concept as it could be implemented in a future home with a high-bandwidth connection. Participant A is located in a kitchen where the walls of the kitchen cabinet doors and the refrigerator have been built with embedded pinhole video cameras centered in each door. Participant B is located in the living room of a remote home where pinhole cameras have been embedded at two different heights, one appropriate for sitting and one appropriate for standing.[1] Both environments have an Everywhere Display Projector (EDP) [6] located at the ceiling of one corner of the

[1]These camera units could eventually be installed by simply drilling a small hole in an existing wall or using component-based architectural wall and furniture systems under development in our group that are designed for such sensor integration [5].

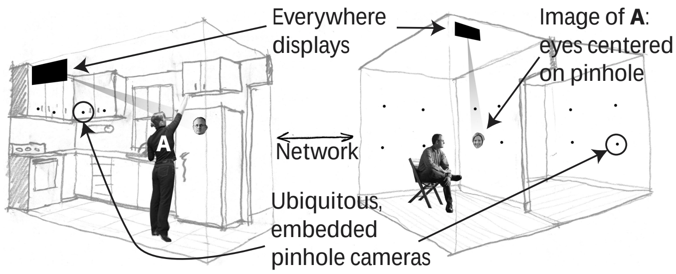

Figure 1: The concept for video conferencing anywhere with sensors embedded in architectural components. Cabinetry with integrated sensing being designed by our group would make such a scenario possible.

room. Each EDP uses a small computer-controlled mirror and perspective warping and can project an undistorted image onto nearly any planar surface in the environment.

The system in person A's space would use face-finding algorithms (e.g. [7]) to detect where person A is located and grabbing the person's image from the closest camera. (2) The system in person B's space would similarly detect where person B's face is located. (3) Based on these locations, the systems would place person A's image in person B's space so that it is near person B. (4) The EDP is used to place the image. The 1.5mm pinhole is barely noticeable on the image of the face. Simultaneously, person B's image appears on the pinhole embedded in the refrigerator appliance in person A's kitchen. As both people move about their respective spaces, the image of the other person will automatically move to the most visually convenient camera position. The system preserves both the ability to make eye contact and the ability to intentionally not make eye contact because the remote participant's eyes are always moved so they are centered directly over a pinhole camera.

3.2 Partial Implementation

We have begun to implement and test the components of this concept. We have two video conferencing stations located in non-adjacent rooms in our laboratory. Each station consists of a flat, white foam core board with a pinhole color camera embedded behind a 1.5mm hole centered horizontally on the foam board. The camera is adjusted so that the optical axis is perpendicular to the board's surface.

Head tracking is performed using the Continuously Adaptive Mean Shift algorithm [2]. The head pixels are translated so that the remote viewer sees the head displayed with the eyes centered over the pinhole camera. A low-pass filter on the face position is used to eliminate jitter motion that would otherwise be visible due to camera noise and sudden flesh changes as people quickly move their heads. The prototype is run using a standard office speaker phone to provide audio. Other mechanisms could be used to improve head tracking (e.g. [7]).

A portable video projector placed to the side of the board uses the same perspective transformation algorithms as the Everywhere Display but permits testing with higher-resolution images than can currently be achieved in the our living room prototype environment with Everywhere Display. When an image of a remote conversant's face is projected with the pinhole centered between the conversant's eyes, the 1.5mm hole is distinguishable but not distracting.

The video also shows our prototype living room with Everywhere Display. Cameras can be embedded directly into surfaces in the space. One is used to create a "magic mirror" effect when used in combination with an Everywhere Display Projector (EDP) [6]. An image from the pinhole camera (flipped) is projected onto the wall centered on the camera itself. The effect is that the user appears to be looking directly at oneself. People who are shown this setup often have a difficult time determining the position of the camera when they are not told about the 1.4mm pinhole in the drywall in advance; only after asking them to "point where the image must be coming from" and pointing out the tiny speck on the wall do visitors typically understand how the effect is achieved.

Eventually we hope to test one half of the full video conferencing system illustrated in this concept video in our prototype living room environment. Resolution limitations of our current Everywhere Display unfortunately prevent an ideal implementation of the concept described. Our group is developing cabinetry components that have integrated sensors networks built in and could easily accommodate the integrated pinhole cameras.

REFERENCES

[1] M. Argyle and M. Cook. *Gaze and Mutual Gaze*. 1976.

[2] G.R. Bradski. Computer vision face tracking for use in a perceptual user interface. *Intel Technology Journal*, 1998.

[3] W. Buxton, A. Sellen, and M. Sheasby. Interfaces for Multiparty Videoconferences. In K. Finn, A. Sellen, and S. Wilbur, editors, *Video-Mediated Communication*, pages 385–400. 1997.

[4] M. Chen. Leveraging the asymmetric sensitivity of eye contact for videoconference. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 49–56. New York, NY, 2002.

[5] K. Larson. Places of Living: Integrated Components for Mass Customization. Changing Places Technical Report, Massachusetts Institute of Technology, March 2002.

[6] C. Pinhanez. The Everywhere Displays Projector: A device to create ubiquitous graphical interfaces. In G.D. Abowd, B. Brumitt, and S.A.N. Shafer, editors, *Proceedings of the Conference on Ubiquitous Computing*, LNCS 2201, pages 315–331. Berlin Heidelberg, 2001.

[7] P. Viola and M. Jones. Rapid Object Detection Using a Boosted Cascade of Simple Features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume I, pages 511–518. 2001.