ConnectBoard: Enabling Genuine Eye Contact and Accurate Gaze in Remote Collaboration

Kar-Han Tan; Ian N. Robinson; Bruce Culbertson; John Apostolopoulos

HP Laboratories
HPL-2011-83

Keyword(s):

Eye contact; gaze awareness; immersive experiences; natural interaction; remote collaboration

Abstract:

Conventional telepresence systems allow remote users to see one another and interact with shared media, but users cannot make eye contact, and gaze awareness with respect to shared media and documents is lost. In this paper, we describe a remote collaboration system based on a see-through display to create an experience where local and remote users are seemingly separated only by a vertical sheet of glass. Users can see each other and media displayed on the shared surface. Face detectors are applied on the local and remote video streams to introduce an offset in the video display so as to bring the local user's face, the local camera, and the remote user's face image into collinearity. This ensures that, when the local user looks at the remote user's image, the camera behind the see-through display captures an image with the "Mona Lisa effect," where the eyes of an image appears to follow the viewer. Experiments show that, for one-on-one meetings, our system is capable of capturing and delivering realistic, genuine eye contact as well as accurate gaze awareness with respect to shared media.

External Posting Date: June 21, 2011 [Fulltext] Approved for External Publication
Internal Posting Date: June 21, 2011 [Fulltext]

ConnectBoard: Enabling Genuine Eye Contact and Accurate Gaze in Remote Collaboration

Kar-Han Tan, *Senior Member, IEEE*, Ian N. Robinson, *Member, IEEE*, Bruce Culbertson, *Member, IEEE*, and John Apostolopoulos, *Fellow, IEEE*

***Abstract*—Conventional telepresence systems allow remote users to see one another and interact with shared media, but users cannot make eye contact, and gaze awareness with respect to shared media and documents is lost. In this paper, we describe a remote collaboration system based on a see-through display to create an experience where local and remote users are seemingly separated only by a vertical sheet of glass. Users can see each other and media displayed on the shared surface. Face detectors are applied on the local and remote video streams to introduce an offset in the video display so as to bring the local user's face, the local camera, and the remote user's face image into collinearity. This ensures that, when the local user looks at the remote user's image, the camera behind the see-through display captures an image with the "Mona Lisa effect," where the eyes of an image appears to follow the viewer. Experiments show that, for one-on-one meetings, our system is capable of capturing and delivering realistic, genuine eye contact as well as accurate gaze awareness with respect to shared media.**

***Index Terms*—Eye contact, gaze awareness, immersive experiences, natural interaction, remote collaboration.**

I. Introduction

THE best remote collaboration systems in existence today strive to create the illusion that the remote and local meeting participants are in the same room. Using high-quality audio visual capture and rendering as well as low-latency compression and streaming, these systems are able to deliver high-fidelity imagery and sound across the globe without noticeable delay. Carefully designed rooms with large displays can present the remote users at the same size as they would appear if they were in the same room. To a large extent, all of these pieces work together to successfully recreate a realistic meeting experience.

While these high-end systems come close to reproducing the experience of colocated meetings, there are still technological barriers to be overcome before remote meetings can be as natural and effective as physical face-to-face meetings, particularly in cases where participants interact closely and/or with shared content. In this paper, we address the problems given here.

- **Eye Contact**. One of the most important aspects of person-to-person social interaction, eye contact is still not fully supported in video conferencing systems on the market today. We would like to allow a user to be able to make eye contact when they look at the remote user's image. This is not possible with today's displays and camera systems, where the camera is typically placed above the display and the user can only look at the display or the camera, but not both. This discrepancy is exacerbated in collaborative setups where the user is closer to the system.
- **Gaze Awareness**. The accurate communication of eye gaze is also crucial in collaboration tasks. Conveying the focus of the remote user's attention (or lack thereof) with regard to the shared content (e.g., "are they looking where I'm pointing?") is an important part of establishing the intersubjectivity required for effective communication. In today's systems, typically the shared documents are displayed separately from the user screens, and gaze direction is rarely conveyed correctly.

These problems are particularly important in designing collaboration systems to support small one-on-one meetings, where users are typically interacting at much closer distances and often work with shared content. In this paper, we present our solution based on recent developments in see-through screen-based collaboration systems. Specifically, we build upon the ConnectBoard system [1] to enhance it to provide further improvements in the key attributes of eye contact and gaze awareness. We show that in the case of a one-on-one meeting, we can deliver both genuine eye contact as well as correct gaze awareness.

II. Previous Work

The authors of [2] recently used a light field display in combination with a real-time 3-D capture system to deliver one-to-many eye contact. However, the light field display was only big enough to display one user's head, and the capture system is unable to recreate fine features like hair. [3] uses a virtual environment and this allows perhaps the most flexibility, but the system uses 3-D polygon avatars that are not photorealistic. The work in [4] also places users in a virtual environment but uses texture-mapped models to create more natural-looking avatars that can convey richer facial expressions. [5] uses 3-D graphics to modify images to improve facial expressions and gaze awareness. In [6], the work exploits experimental evidence on the asymmetry in the perception of eye contact to design a system that places a camera above a display since sensitivity to gaze in the downward direction is lower than in other directions. However, the visual angle between the camera and the eyes rendered on the display has to be less than 5°, which places a strong constraint on the size of the rendered image as well as viewing distances.

Manuscript received November 02, 2010; revised February 28, 2011; accepted March 07, 2011. Date of publication March 22, 2011; date of current version May 18, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Oscar Au.

The authors are with Hewlett-Packard Laboratories, Palo Alto, CA 94304 USA (e-mail: Kar-han.tan@hp.com; Ian.robinson@hp.com; Bruce.culbertson@hp.com; japos@hpl.hp.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2011.2130516

A. See-Through Displays

In order to achieve eye contact at close range to a large display, the camera needs to be behind the screen, shooting through the display. There have been many previous attempts at see-through displays in the past. The Teleprompter [7], which is widely used today by television newscasters and public speakers, is an early example. While Teleprompters are one-way communication devices, similar devices like Gazecam [8] and the EuroPARC Reciprocal Video Tunnel [9] were used in teleconferencing systems. These systems allow users to look at the remote user's image while looking into the camera at the same time. Using a half-silvered mirror, which is typically angled at 45° from the display, results in a large footprint. Stray reflections off the mirror can also create distracting views, for example, of the ceiling or floor. Creating eye contact using this method thus typically results in deep enclosures that limit the range of usable viewing angles.

A creative variation of the half-silvered mirror technique was used in ClearBoard [10], where a polarizing film was sandwiched between a projection screen and a half mirror. The system has a "drafting table" design where the work and display surface is placed at a 45° angle, and a camera captures the mirror image of the user from above. Images from the display are blocked by another polarizing filter on the camera to ensure that only the mirrored image is captured. However, the drafting table design produces an unnatural view of the remote user, who would appear to be leaning backwards while working on the shared surface, even though he/she is sitting upright.

Another way to create a see-through display is to use switchable liquid crystal diffusers, which is a technique demonstrated by Shiwa at NTT [11]. Such a diffuser can switch quickly between two states: transparent and diffusing. In the transparent state, synchronized cameras can capture images of the user. In the diffusing state, synchronized projectors can render images of the remote user. This technique was also used in *blue-c* [12]. More recently, SecondLight [13] used a switching diffuser to allow projection onto objects above the screen, enabling viewing of overlay visualizations. The switching diffuser technique allows smaller footprints and wider viewing angles. The key technical limitation as reported in [10] and [14] is that currently available diffusers may not switch sufficiently quickly, especially in larger screens, resulting in flickering images. It is possible to overdrive the screens to achieve higher switching frequencies, but transition times between the two states are still significant enough to reduce the actual duty cycles of both the synchronized projector and camera, resulting in dim or noisy images.

Instead of half-mirrors or switching diffusers, TouchLight [15] uses a screen that diffuses only light incident from prespecified angles and allows light to pass through otherwise. This gives a transparent screen which can display images if the projector is placed at the right location. It does not require special synchronized cameras and projectors, thus offering greater freedom for designers. However, as the diffuser bounces light from the projector back into the camera as well, this *backscattered* portion of the displayed content gets superimposed on the image of the user captured by the camera.

The HoloPort [14] systems use a semi-transparent holographic projection screen, together with temporal multiplexing, alternating projected images with image capture to separate the light from the projected images from that captured through the projection screen by the camera.

Fig. 1. ConnectBoard in action, linked to a second prototype in a remote location. Note that the image of the remote user is horizontally flipped so that drawings made on the screen appear correct to both users.

III. Our Solution

As we have seen, there have been several attempts at designing systems which simultaneously support eye contact and nonverbal communications. Ideally, we would like to create a ClearBoard-like experience, except with a vertical surface that captures a frontal view of the user. We would also like to be able to build our system from off-the-shelf components, without requiring synchronization between projectors and cameras, so that they can operate at their respective optimal frame rates and exposure settings.

To meet these requirements, we employ a directionally selective diffuser similar to that used in HoloPort. Unlike the HoloPort system, we choose not to use active synchronized projector cameras. Instead, we use *wavelength multiplexing* in the visible light spectrum. The idea is that, if the projector outputs light in spectral ranges that do not overlap those observable by the camera, the camera will not sense the backscattered light emitted from the projector. A wavelength multiplexed projector–camera pair can be built with the use of interference filters originally designed for viewing stereoscopic 3-D movies [16]. We call our system *ConnectBoard* [1] (see Figs. 1 and 2).

The "see-through screen" arrangement of ConnectBoard allows for capturing an image of the remote user from a viewpoint that corresponds to that of the local user, as if they were colocated either side of a glass screen (except for the mirroring necessary to keep the content intelligible). Part of such a system is illustrated in Fig. 3. See-through screen systems recreate a whiteboard-like collaboration experience, in which the users are not constrained to stand at the center of the screen, but instead are free to move about the area in front of the screen and use its whole surface for interaction. This causes a problem with communicating eye contact, as illustrated in Fig. 4. In this paper, we present a method for better communicating eye contact in a see-through screen collaboration system, irrespective of user position.

A. Enabling Eye Contact

To capture a view of a remote user that communicates eye contact, that user must be looking directly into the camera [17]. The image presented to the local user is then one in which the

Fig. 2. Optical backscatter removal. (a) Display–camera setup. (b) Backscatter clutters user image. (c) Backscatter removed.

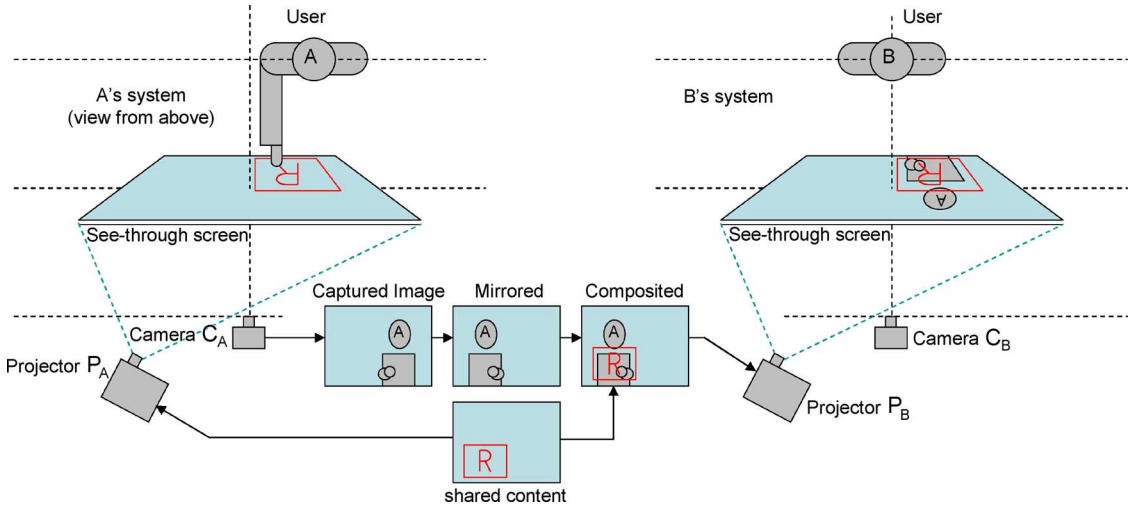


Fig. 3. Part of a "see-through screen" system. User A creates some content R on the screen. Camera C_A captures A's image through the screen (but not the projected content). This image is mirrored left to right and combined with the shared content for display for user B (right). Not shown is the reverse path where B's image is captured by camera C_B, mirrored and displayed with the shared content on A's screen.

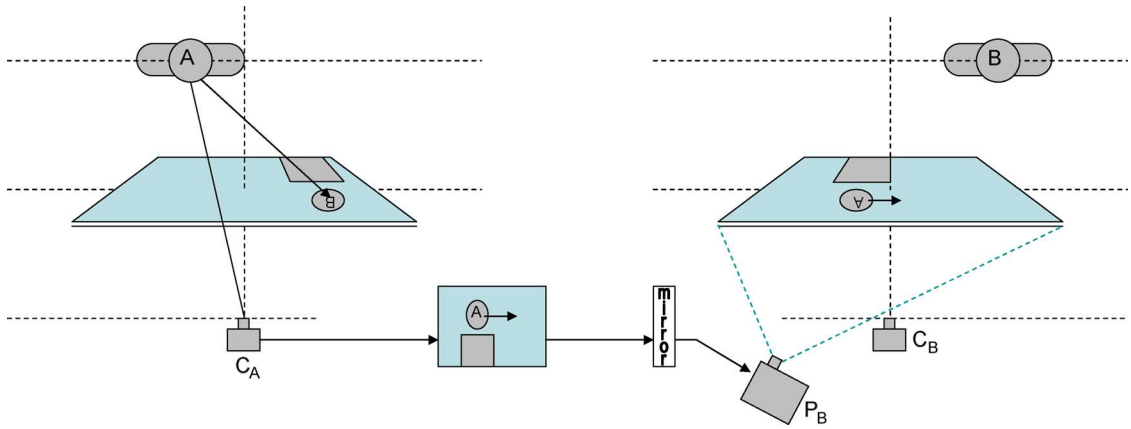


Fig. 4. Gaze problem. On the left, user A is looking directly at user B's image. However, from camera C_A's viewpoint, A appears to be looking to their left. When this image is relayed to user B (right), B does not get the sense that A is looking directly at B. The reverse situation is similar, so neither party feels that eye contact is being established.

remote user appears to look straight out of the screen. It has been noted that the eyes of a static portrait appears to follow the observer [18]. By virtue of this "Mona Lisa effect," the image will appear to the local user to be looking at them, irrespective of their relative positions on and in front of the screen.

In order to have the local user look into the camera when attempting to make eye contact with the remote user, the image of the remote user is shifted so that it lies on the path between the local user and camera. On a see-through display, this means that we can achieve realistic eye contact by ensuring that the camera, the eyes of the local user, and the image of the remote user's eyes are collinear, as shown in Fig. 5. The amount of shift is equal to the difference between the position of the local user's face captured in the local camera, and the position of the remote

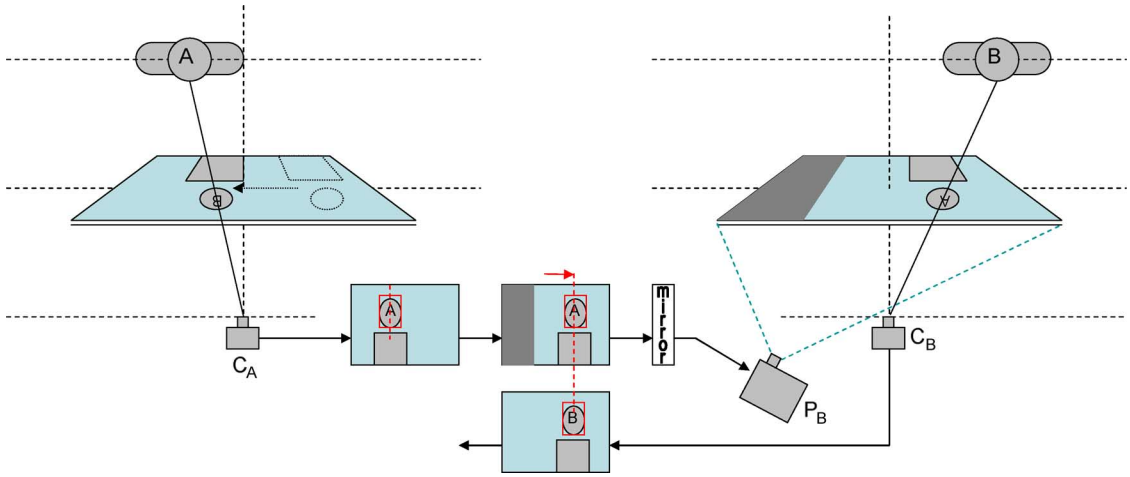


Fig. 5. Fixing the gaze problem. On the left, the image of user B is shifted so that it is displayed on a line between user A's head and the camera C_A. In this way, C_A captures an image of A looking directly into the camera. The mechanism for achieving this is shown for the display of A's image to B: face detectors are run on the images captured by cameras C_A and C_B to determine the positions of A's and B's heads, respectively (the red bounding boxes). The location of B's head indicates the desired position of user A's image (so that it will lie on the line B-C_B). The image from camera C_A is then shifted (red arrow) so that the detected face of A lines up with this desired position.

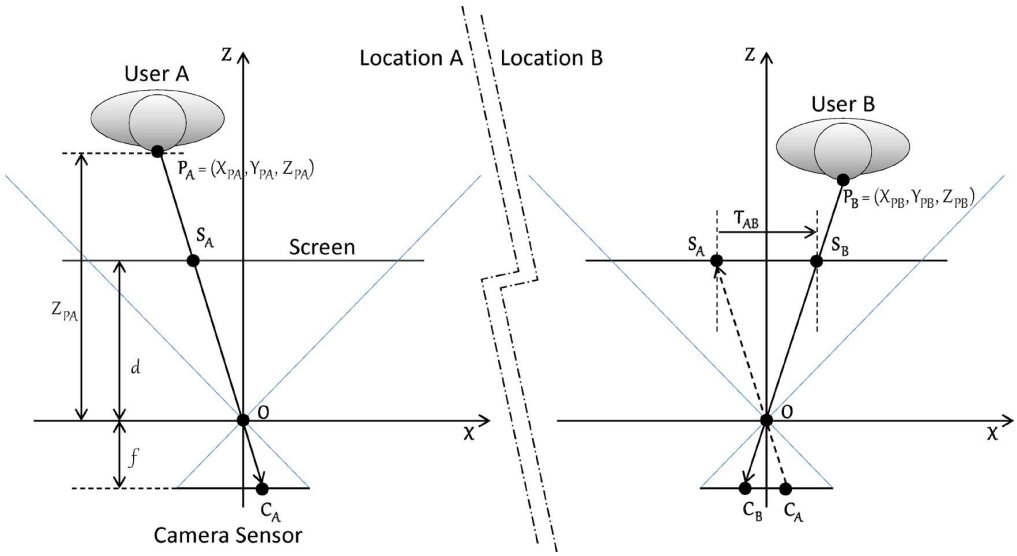


Fig. 6. Geometric analysis of shifting operation.

user's face captured in the remote camera. By symmetry, this means that the amount of shift on the two connected systems are equal and opposite.

B. Computing the Shift Vector

Consider Fig. 6, which shows the imaging geometry in a pair of identical ConnectBoards in geographically separated locations A and B. Two users A and B are collaborating remotely via their respective ConnectBoards. We denote the spatial locations of the user's faces as $P_A = (X_{PA}, Y_{PA}, Z_{PA})$ and $P_B = (X_{PB}, Y_{PB}, Z_{PB})$, respectively. A camera behind each screen is at location O, which we use as the origin of our coordinate frame, with the Z axis pointed out through the screen. To simplify the analysis, we assume that the camera optical axis is perpendicular to the screen, and the camera sensor plane is parallel to the screen as well. Using a pinhole projection model, we can draw a ray from P_A through O, which intersects the sensor at point C_A. The ray also intersects the screen at the point S_A, which we can write in terms of P_A as follows:

$$S_A = \left(d\frac{X_{PA}}{Z_{PA}}, d\frac{Y_{PA}}{Z_{PA}} \right) \tag{1}$$

Similarly, we have

$$C_A = \left(f\frac{-X_{PA}}{Z_{PA}}, f\frac{-Y_{PA}}{Z_{PA}} \right) \tag{2}$$

where d is the distance from O to the screen, and f is the distance from O to the sensor, both along the Z-axis.

The image of user A captured by the camera is then streamed to location B. Without any shifting, it will be displayed such that the image of user A appears at the same location S_A. Note that, since the two ConnectBoards are identical, we reuse the same coordinate frame. Now, by a similar analysis, if we shoot a ray from user B's location P_B through O to the sensor that it

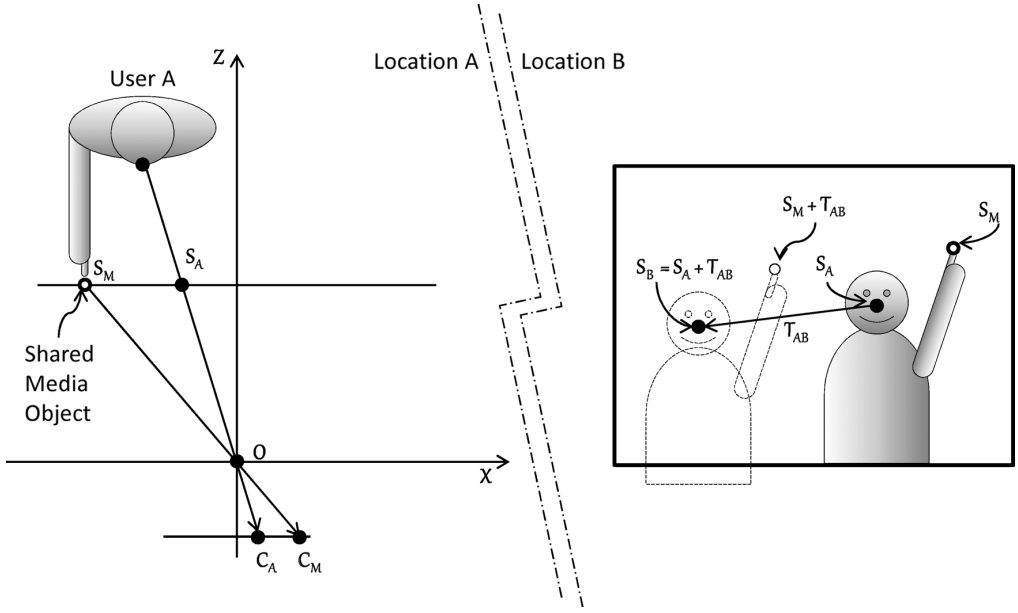


Fig. 7. Geometric analysis of shared media shifting operation.

intersects at the point C_B, the ray also intersects the screen at the point S_B, which can be written as

$$S_B = \left(d\frac{X_{PB}}{Z_{PB}}, d\frac{Y_{PB}}{Z_{PB}} \right). \tag{3}$$

In order to enforce collinearity between user B at P_B, the camera at O, and the image of user A, we therefore need to translate the image of user A by the vector T_{AB} as

$$\begin{aligned} T_{AB} &= S_B - S_A \\ &= \left(d\frac{X_{PB}}{Z_{PB}}, d\frac{Y_{PB}}{Z_{PB}} \right) - \left(d\frac{X_{PA}}{Z_{PA}}, d\frac{Y_{PA}}{Z_{PA}} \right). \end{aligned} \tag{4}$$

Recall that user A is projected onto the image sensor at C_A. Using a face detector, we can *directly estimate* the location (X_{FA}, Y_{FA}) of the user's face on the image. Therefore, $C_A = (X_{FA}, Y_{FA})$, and we can write

$$\begin{aligned} X_{FA} &= f\frac{-X_{PA}}{Z_{PA}} \\ -\frac{1}{f}X_{FA} &= \frac{X_{PA}}{Z_{PA}} && (5) \\ -\frac{1}{f}Y_{FA} &= \frac{Y_{PA}}{Z_{PA}} && (6) \\ -\frac{1}{f}X_{FB} &= \frac{X_{PB}}{Z_{PB}} && (7) \\ -\frac{1}{f}Y_{FB} &= \frac{Y_{PB}}{Z_{PB}}. && (8) \end{aligned}$$

Now, we can rewrite T_{AB} in terms of face detector estimates

$$\begin{aligned} T_{AB} &= \left(-\frac{d}{f}X_{FB}, -\frac{d}{f}Y_{FB} \right) - \left(-\frac{d}{f}X_{FA}, -\frac{d}{f}Y_{FA} \right) \\ &= -\frac{d}{f}(X_{FB} - X_{FA}, Y_{FB} - Y_{FA}). \end{aligned} \tag{9}$$

Two aspects are worth noting on this expression for T_{AB}.

- First, $-(d/f)$ is a constant that can be estimated from a one-time calibration of the camera-display system.
- Second, only 2-D face detector estimates $(X_{FA}, Y_{FA}), (X_{FB}, Y_{FB})$ are required, and the depth Z_{PA}, Z_{PB} are not.

Again by symmetry, the amount of translation required at location A is equal and opposite:

$$\begin{aligned} T_{BA} &= S_A - S_B \\ &= -\frac{d}{f}(X_{FA} - X_{FB}, Y_{FA} - Y_{FB}) \end{aligned} \tag{10}$$

C. Gaze and Shared Media

An important property of see-through screen collaboration systems is that they enable accurate communication of user gaze and gestures with respect to the shared media. For example, user A can point at a part of the shared content and look to see whether user B is following the gesture, or concentrating on some other part. Thus, when the image of the user is shifted, the content needs to be shifted accordingly. Since the user and shared media are at difference distances from the camera, it would seem that the user image and shared media images need to be shifted differently. However, since the user image is ultimately projected onto the same plane as shared content, the required amount of shift is identical for both user and content images.

This is illustrated in Fig. 7. In the figure, user A is pointing at a shared media object being displayed at position S_M on the screen. If the camera were to be able to observe the shared media object, it would project in the same way as the user onto the camera sensor. When this image is reproduced at location B without any shifting, user A's image is at S_A, and similarly S_M would be the location of the shared media object. In order for the images of the shared media and the user to be correctly aligned on the screen at location B, the on-screen displacement between S_A and S_M has to be preserved. Therefore, if user A's image

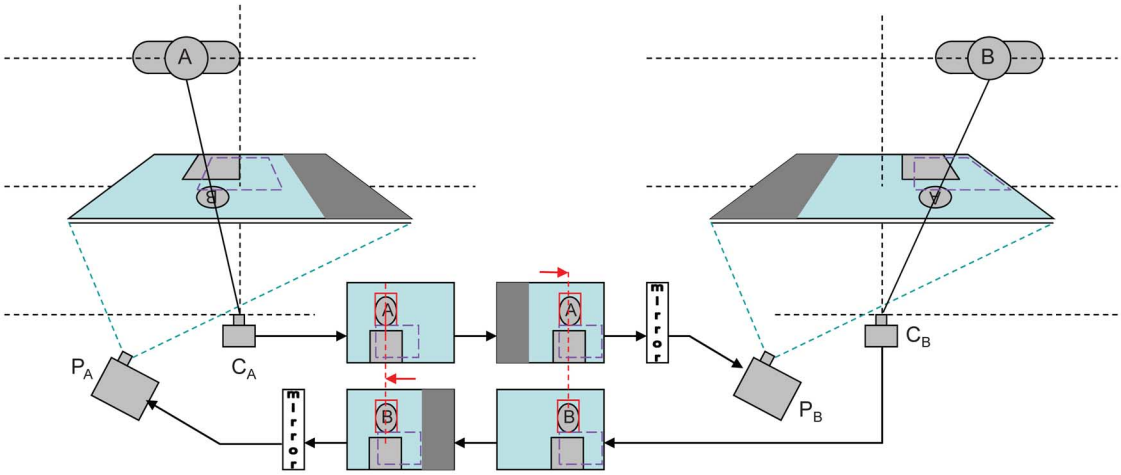


Fig. 8. Shifting shared content. Here, the location of the shared content (purple dashed rectangle) is shown in both video paths (note that the camera does not "see" the projected image; the content is digitally composited, as shown in Fig. 1). Content is shifted by the same amount as the user image, thus preserving the user's gaze angles with respect to the content. Note that the magnitude of the shift is the same in both paths, only the direction is different. Hence, the content locations are consistent between the two screens. Also note that the blank area of the screen resulting from the image shift serves to indicate to each user the screen area that cannot be used for shared content.

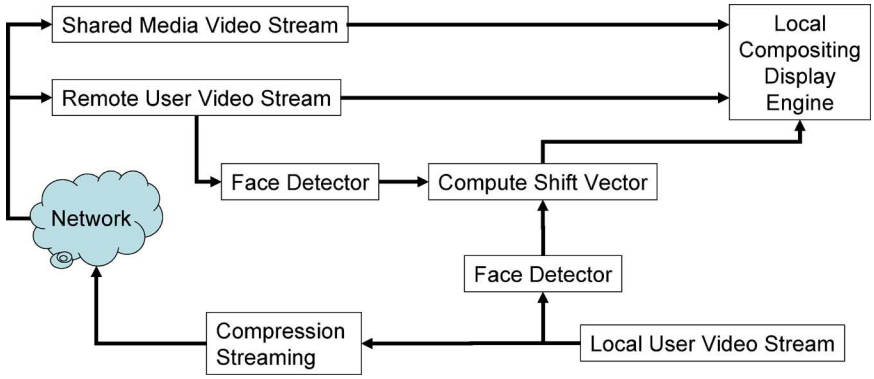


Fig. 9. Real-time processing pipeline.

needs to be shifted by vector T_{AB} then in order to preserve the on-screen displacement between S_A and S_M, the shared media image needs to be shifted by the same vector T_{AB}.

Recall that the magnitude of the shift is the same in both paths, only the direction is different. Hence, the content shift going from A to B is undone by the shift from B to A, so that the content locations remain consistent.

D. Impact on User Experience

Shifting the images of the remote user and shared content in this way supports eye contact and preserves gaze and gesture directions with respect to the shared media. However there are two main consequences that could impact the user experience. First shifting the screen image introduces a blank area, as shown in Figs. 5 and 8. This actually has the following useful properties: it serves to indicate to each user the extent of the screen area that can be used for shared content (note also that the non-usable area is always farthest from the user), and its dynamic coupling to the users' shifts in position tends to induce users to move in such a way as to maximize the collaboration area.

The second consequence of our described method is that, if a user moves to one side or the other to get closer to some part of the shared media, the shared media will move in the same direction, away from the user. However, the shared media is shifted to a lesser extent than the user moves (roughly half if the camera distance behind the screen mimics the user distance in front of the screen). It may be useful to include a method that allows the user to easily scroll the shared content so that they can bring the desired area of the content to them, rather than moving to it.

E. Implementation and Experiments

For our experiments, we use the ConnectBoard see-through display solution because the system can be built with off the shelf components, requiring no custom electronics to synchronize the projectors and cameras. An additional benefit is that the projectors and cameras can operate at their respective optimal frame rates and exposure settings since temporal synchronization is not necessary.

We implemented the face alignment algorithm using a version of the Viola-Jones face detector [19] optimized for video processing based on staggered sampling [20], where a coarse sampling grid is shifted between video frames so that all points on a fine grid is sampled over a number of frames. The algorithm is integrated into a media streaming and compression framework [21], and Fig. 9 is an illustration of our data flow processing pipeline.

Fig. 10. Enabling eye contact. As the local user moves from the left side of the display to the right, the image of the remote user is automatically shifted so that the camera, local user's face, and remote user's face is always collinear. This way, when the local user looks at the face of the remote user, he/she is always also looking into the camera, which captures the "eye contact" view. The black bars in the screen in the top and bottom images serve to indicate the limits of the shared area between the two users. This subtly encourages local and remote users to position themselves such that the shared area is as large as possible, as is the case in the middle image.

Our prototype runs in real time, and we were able to test the eye-contact enhancement algorithms, as shown in Fig. 10. In our experiments, users reported that they were able to make eye contact with remote users, and they were also able to correctly infer the gaze direction with respect to shared media, as shown in Fig. 11.

IV. Discussion and Future Work

We have presented a novel solution for delivering genuine eye contact and accurate gaze awareness in small one-to-one meetings. As the system uses high-quality low-latency audio and video, the experience created is natural and realistic. The ability to make eye contact and use nonverbal communications like gaze and gestures effectively enables richer and more intimate interactions than existing systems. The system design also supports a number of important features for remote collaboration, such as concurrent access to shared media and having an integrated display of users and shared media [22].

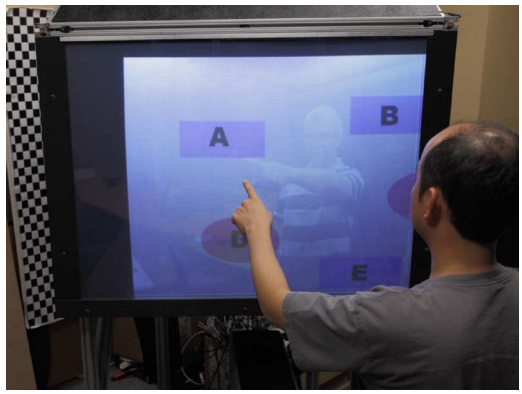


Fig. 11. Ensuring gaze awareness. As the remote user's image is shifted, the shared media (shown here digital composited) layer is also shifted correspondingly to ensure a correct rendering of the remote user's gaze direction with respect to shared content. At the same time, the users can make eye contact if they look at each other in the eye.

For future work, a more detailed study of gaze perception on a see-through display setup would help shed light on possible enhancements that may make the collaboration experience even more natural and engaging. In particular, it would be useful to understand the effect of our shifting operations to the perceived "mental image" [17], [23], [24] and understand the limits of the "Mona Lisa effect" [25], [26] to induce the perception that one is making eye contact. We are also looking at "pseudo-3-D" effects [27] that may enhance the user experience.

Acknowledgment

The authors would like to thank the anonymous reviewers for their valuable feedback and the members of the MIX Lab, Hewlett-Lackard Labs, for their continuous help.

References

[1] K.-H. Tan, I. Robinson, R. Samadani, B. Lee, D. Gelb, A. Vorbau, B. Culbertson, and J. Apostolopoulos, "Connectboard: A remote collaboration system that supports gaze-aware interaction and sharing," in *Proc. IEEE Workshop Multimedia Signal Process.*, 2009, pp. 1–6.

[2] A. Jones, M. Lang, G. Fyffe, X. Yu, J. Busch, I. McDowall, M. Bolas, and P. Debevec, "Achieving eye contact in a one-to-many 3D video teleconferencing system," *Proc. ACM SIGGRAPH*, vol. 28, no. 3, pp. 1–8, 2009.

[3] D. Roberts, R. Wolff, J. Rae, A. Steed, R. Aspin, M. McIntyre, A. Pena, O. Oyekoya, and W. Steptoe, "Communicating eye-gaze across a distance: Comparing an eye-gaze enabled immersive collaborative virtual environment, aligned video conferencing, and being together," in *Proc. IEEE Virtual Reality Conf.*, 2009, pp. 135–142.

[4] H. H. Baker, N. Bhatti, D. Tanguay, I. Sobel, D. Gelb, M. E. Goss, B. W. Culbertson, and T. Malzbender, "Understanding performance in coliseum, an immersive videoconferencing system," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 1, no. 2, pp. 190–210, 2005.

[5] J. Gemmell, K. Toyama, C. L. Zitnick, T. Kang, and S. Seitz, "Gaze awareness for video-conferencing: A software approach," *IEEE Multimedia*, vol. 7, no. 1, pp. 26–35, Jan. 2000.

[6] M. Chen, "Leveraging the asymmetric sensitivity of eye contact for videoconference," in *Proc. ACM SIGCHI Conf. Human Factors Computing Syst.*, 2002, pp. 49–56.

[7] J. Oppenheimer, "Prompting Apparatus," U.S. Patent 2 883 902, Apr. 28, 1959.

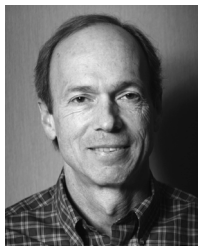
[8] S. R. Acker and S. R. Levitt, "Designing videoconference facilities for improved eye contact," *J. Broadcasting Electron. Media*, vol. 31, no. 2, pp. 181–191, 1987.
[9] B. Buxton and T. Moran, "Europarc's integrated interactive intermedia facility (iiif): Early experiences," in *Proc. IFIP WG 8.4 Conf. Multi-User Interfaces Applic.*, Amsterdam, The Netherlands, 1990, pp. 11–34.
[10] H. Ishii and M. Kobayashi, "Clearboard: A seamless medium for shared drawing and conversation with eye contact," in *Proc. ACM SIGCHI Conf. Human Factors Comput. Syst.*, 1992, pp. 525–532.
[11] S. Shiwa and M. Ishibashi, "A large-screen visual telecommunication device enabling eye contact," *SID Dig.*, vol. 22, pp. 327–328, 1991.
[12] M. Gross, S. Würmlin, M. Naef, E. Lamboray, C. Spagno, A. Kunz, E. Koller-Meier, T. Svoboda, L. Van Gool, S. Lang, K. Strehlke, A. V. Moere, and O. Staadt, "Blue-c: A spatially immersive display and 3d video portal for telepresence," in *Proc. ACM SIGGRAPH*, 2003, pp. 819–827.
[13] S. Izadi, S. Hodges, S. Taylor, D. Rosenfeld, N. Villar, A. Butler, and J. Westhues, "Going beyond the display: A surface technology with an electronically switchable diffuser," in *Proc. 21st Annu. ACM Symp. User Interface Software Technol. (UIST)*, 2008, pp. 269–278.
[14] M. Kuechler and A. Kunz, "Holoport—A device for simultaneous video and data conferencing featuring gaze awareness," in *Proc. IEEE Conf. Virtual Reality*, 2006, pp. 81–88.
[15] A. Wilson, "Touchlight: An imaging touch screen and display for gesture-based interaction," in *Proc. Int. Conf. Multimodal Interfaces (ICMI)*, 2004, pp. 69–76.
[16] H. Jorke and M. Fritz, "A new stereoscopic visualization tool by wavelength multiplex imaging," in *Proc. Electron. Displays*, Sep. 2003, pp. 407–410.
[17] J. J. Koenderink, A. J. van Doorn, A. M. L. Kappers, and J. T. Todd, "Pointing out of the picture," *Perception*, vol. 33, pp. 513–530, 2004.
[18] W. H. Wollaston, "On the apparent direction of eyes in a portrait," *Philos. Trans. Roy. Soc. London*, vol. 114, pp. 247–256, May 1824.
[19] P. Viola and M. Jones, "Robust real-time object detection," Compaq CRL Tech. Rep. CRL-2001-01, Feb. 2001 [Online]. Available: http://www.hpl.hp.com/techreports/Compaq-DEC/CRL-2001-1.html
[20] D. Greig, "Video object detection speedup using staggered sampling," in *Proc. IEEE Workshop Applic. Comput. Vis.*, 2009, pp. 1–7.
[21] D. Tanguay, D. Gelb, and H. H. Baker, "Nizza: A framework for developing real-time streaming multimedia applications," HP Labs Tech. Rep. HPL-2004-132, 2004 [Online]. Available: http://library.hp.com/techpubs/2004/HPL-2004-132.html
[22] J. C. Tang and S. Minneman, "Videowhiteboard: Video shadows to support remote collaboration," in *Proc. SIGCHI Conf. Human Factors Computing Syst.: Reaching Through Technol.*, 1991, pp. 315–322, ser. CHI '91.
[23] J. J. Koenderink, A. J. van Doorn, A. M. L. Kappers, and J. T. Todd, "Ambiguity and the 'mental eye' in pictorial relief," *Perception*, vol. 30, pp. 431–448, 2001.
[24] J. T. Todd, A. H. J. Oomes, J. J. Koenderink, and A. M. L. Kappers, "On the affine structure of perceptual space," *Psycholog. Sci.*, vol. 12, no. 3, pp. 191–196, May 2001.
[25] D. Todorović, "Geometrical basis of perception of gaze direction," *Vis. Res.*, vol. 46, pp. 3549–3562, 2006.
[26] S. Anstis, J. Mayhew, and T. Morley, "The perception of where a face or television 'portrait' is looking," *Amer. J. Psychol.*, vol. 82, no. 4, pp. 474–489, 1969.
[27] C. Zhang, Z. Yin, and D. Florencio, "Improving depth perception with motion parallax and its application in teleconferencing," in *Proc. IEEE Workshop Multimedia Signal Process.*, 2009, pp. 1–6.

Kar-Han Tan (SM'11) received the B.Sc. degree from the National University of Singapore, Singapore, the M.S. degree from the University of California, Los Angeles, and the Ph.D. in computer science from the University of Illinois at Urbana-Champaign, Urbana, IL.

While with the University of Illinois at Urbana-Champaign, he was a Beckman Graduate Fellow. He is currently a Senior Research Scientist with the Mobile and Immersive Experience Lab (MIXL), Hewlett-Packard Laboratories, Palo Alto, CA, where he is working on 3-D capture and display technologies as well as next-generation remote collaboration systems. He contributes actively to the research community and has received best paper awards for his work. Prior to joining Hewlett-Packard, he was Manager of Algorithms Group, EPSON R&D, where he led the invention of View Projection, a technique that enables one-touch setup of light displays on arbitrary surfaces. He coinvented multi-flash imaging at Mitsubishi Electric Research Lab (MERL) and the Virtual Structures algorithm at the University of California, Los Angeles, which is widely recognized today as one of the fundamental techniques for mobile robot formation control.

Ian N. Robinson (M'09) received the B.Sc. degree in physics from the University of Sussex, Sussex, U.K.

He is currently a Senior Research Scientist with the Mobile and Immersive Experiences Lab, Hewlett- Packard Laboratories, Palo Alto, CA. He worked with the research labs at GEC, Fairchild, and Schlumberger before joining Hewlett-Packard Labs in 1988. His previous work ranged from parallel architectures for machine vision, artificial intelligence, and cluster computing to new user interface devices and experiences. His recent work has focused on systems to support remote collaboration.



Bruce Culbertson (M'83) received the M.A. degree in mathematics from the University of California, San Diego, and the M.S. degree in computer and information science from Dartmouth College, Hanover, NH.

He joined Hewlett-Packard in 1983 and has worked at Hewlett-Packard Laboratories, Palo Alto, CA, since 1984. He manages researchers with expertise in computer vision and graphics who have applied those disciplines primarily to multimedia communications for the last decade. Prior to that, he worked on voice and data networks, computer design, computer-aided design of digital logic, reconfigurable computers, and defect-tolerant computer design. He has served on the program committees of IEEE CVPR and IEEE ICCV. His research interests are in computer vision and multimedia.

John Apostolopoulos (F'08) received the B.S., M.S., and Ph.D. degrees from the Masachusetts Institute of Technology (MIT), Cambridge.

He is a Distinguished Technologist and the Director of the Mobile and Immersive Experience Lab (MIXL) with Hewlett-Packard Laboratories, Palo Alto, CA. He also teaches and conducts joint research at Stanford University, Stanford, CA, where he is a Consulting Associate Professor of Electrical Engineering, and he is a frequent Visiting Lecturer with MIT. His research interests include immersive communication and improving the reliability, fidelity, scalability, and security of multimedia communications over wired and wireless packet networks.

Dr. Apostolopoulos was the recipient of Emmy Award Certificate for his contributions to the U.S. Digital TV standard. He was named "one of the world's top 100 young (under 35) innovators in science and technology" (TR100) by *MIT Technology Review* in 2003 and was the recipient of a number of Best Paper awards.