Enabling Genuine Eye Contact and Accurate Gaze in Remote Collaboration

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

HP Laboratories
HPL-2010-96

Keyword(s):

eye contact, gaze awareness, remote collaboration, natural interaction, immersive experiences

Abstract:

Conventional telepresence systems allow remote users to see one another and interact with shared media and documents, 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 on the local and remote video streams are used 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 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: August 21, 2010 [Fulltext] Approved for External Publication
Internal Posting Date: August 21, 2010 [Fulltext]
Presented at IEEE International Conference on Multimedia and Expo, July 19, 2010

ENABLING GENUINE EYE CONTACT AND ACCURATE GAZE IN REMOTE COLLABORATION

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

Multimedia Communications and Networking Lab
Hewlett-Packard Labs
Palo Alto, California, USA
Email: karhan.tan@hp.com

ABSTRACT

Conventional telepresence systems allow remote users to see one another and interact with shared media and documents, 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 on the local and remote video streams are used 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 our system is capable of capturing and delivering realistic, genuine eye contact as well as accurate gaze awareness with respect to shared media.

Keywords— eye contact, gaze awareness, remote collaboration, natural interaction, immersive experiences

1. 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 these pieces work together to successfully recreate a realistic meeting experience.

While these high end systems come close to reproducing the experience of co-located 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 following problems:

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 set-ups where the use 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 users 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 1-on-1 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 Connect-Board 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 1-on-1 meeting, we can deliver both genuine eye contact as well as correct gaze awareness.

2. PREVIOUS WORK

[2] recently used a light field display in combination with a real-time 3D capture system to deliver one-to-many eye contact. The light field display however 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 3D polygon avatars that are not photorealistic. [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 3D graphics to modify images to improve

978-1-4244-7492-9/10/$26.00 ©2010 IEEE

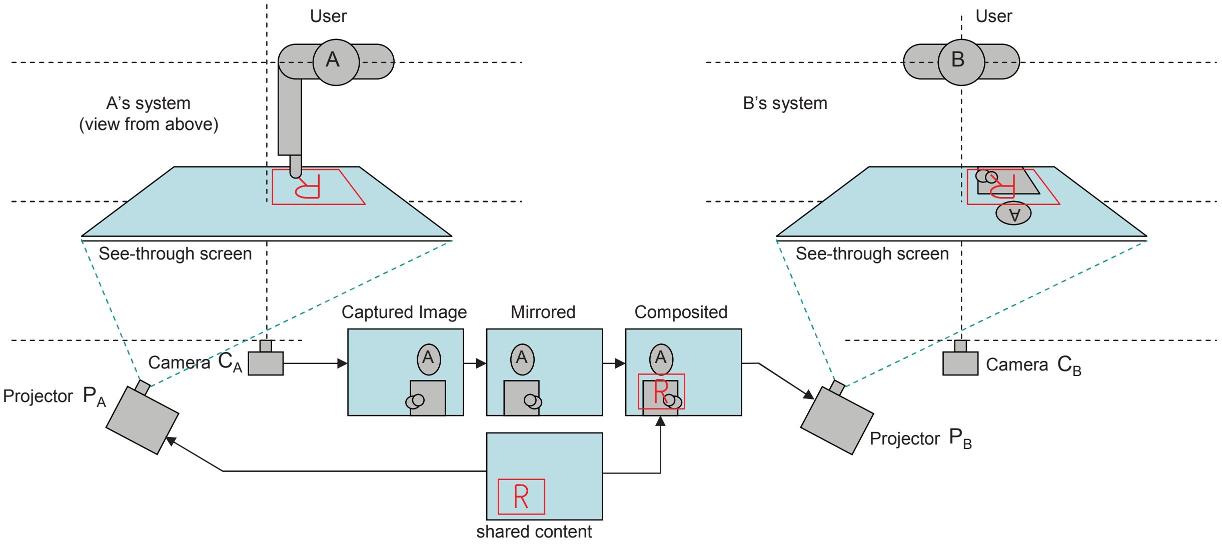

Fig. 1. Part of a 'see-through screen' system. User A creates some content R on the screen. Camera C_A captures A's image through the screen (but not the projected content). This image is mirrored left to right and combined with the shared content for display for user B (right). Not shown is the reverse path where B's image is captured by camera C_B, mirrored and displayed with the shared content on A's screen.

facial expressions and gaze awareness. [6] 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.

2.1. 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 say 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 mirror 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, 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, 14] is that currently available diffusers may not switch fast enough, 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 pre-specified 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 *backscat-*

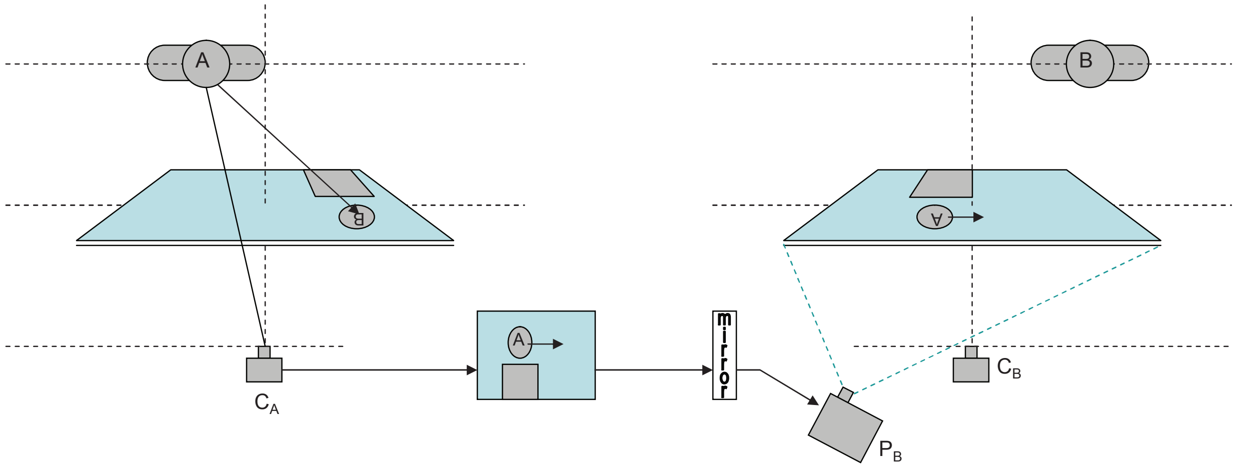

Fig. 2. The gaze problem. (left) User A is looking directly at user B's image. However, from camera C_A's viewpoint, A appears to be looking to their left. When this image is relayed to user B (right), B does not get the sense that A is looking directly at B. The reverse situation is similar, so neither party feels that eye contact is being established.

tered portion of the displayed content gets superimposed on the image of the user captured by the camera.

The HoloPort [14] and ConnectBoard [1] systems use similar transparent screens, together with various techniques to separate the light from the projected images from that captured through the projection screen by the camera. Holoport uses temporal multiplexing, alternating projected images with image capture. ConnectBoard uses *wavelength multiplexing*, in which different, non-overlapping sets of wavelengths are used for projection and capture. 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 3D movies [16].

These 'see-through screen' arrangements allow for capturing an image of the remote user from a viewpoint that corresponds to that of the local user, as if they were co-located 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. 1. 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. 2. In this paper we present a method for better communicating eye-contact in a see-through screen collaboration system, irrespective of user position.

3. OUR SOLUTION

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 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. 3. 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 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.

3.1. 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. As it turns out, 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. 4.

Recall that the magnitude of the shift is the same in both

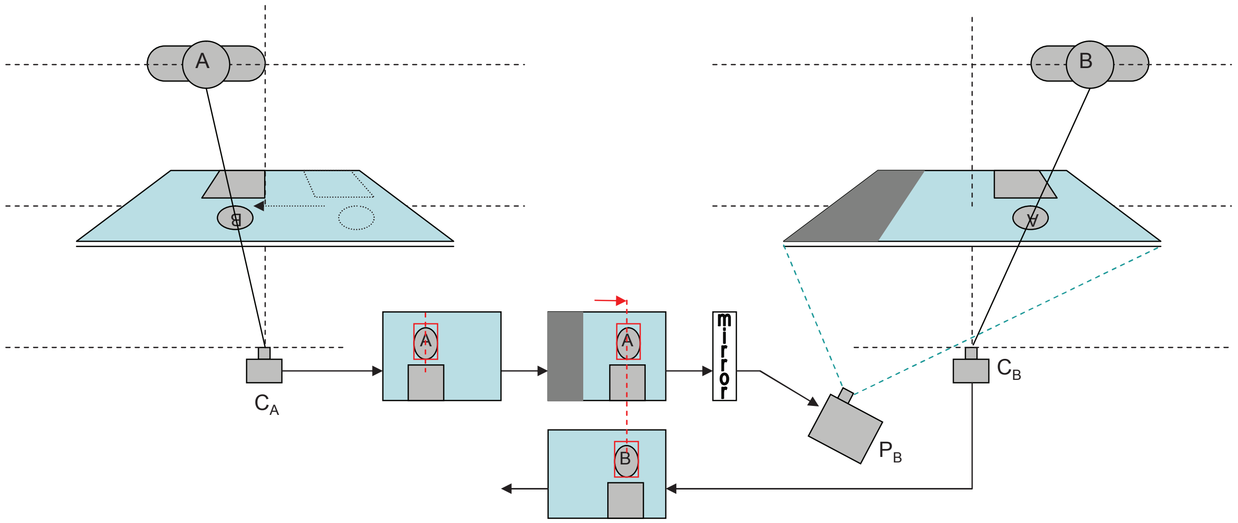

Fig. 3. Fixing the gaze problem. (left) The image of user B is shifted so that it is displayed on a line between user A's head and the camera C_A. In this way C_A captures an image of A looking directly into the camera. The mechanism for achieving this is shown for the display of A's image to B: face 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.

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. Shifting the screen image introduces a blank area as shown in Figs. 3 and 4. This actually has two useful properties. Firstly 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). Secondly 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.

3.2. 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. 5 is an illustration of our data flow processing pipeline.

Our prototype runs in real time and we were able to test the eye contact enhancement algorithms, as shown in Fig. 6. 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. 7.

4. DISCUSSION AND FUTURE WORK

We have presented a novel solution for delivering genuine eye contact and accurate gaze awareness in small 1-to-1 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.

For future work, a more detailed study of gaze perception on a see-through display set up 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' [22], [23], [17], and understand the limits of the 'Mona Lisa effect' [24],[25] to induce the perception that one is making eye contact. We are also looking at 'pseudo-3D' effects [26] that may enhance the user experience.

5. REFERENCES

[1] Kar-Han Tan, Ian Robinson, Ramin Samadani, Bowon Lee, Dan Gelb, Alex Vorbau, Bruce Culbertson, and John Apostolopoulos, "Connectboard: A remote collaboration system that supports gaze-aware interaction and sharing," in *Proceedings IEEE Workshop on Multimedia Signal Processing (***MMSP***)*, 2009.

[2] Andrew Jones, Magnus Lang, Graham Fyffe, Xueming Yu, Jay Busch, Ian McDowall, Mark Bolas, and Paul Debevec, "Achieving eye contact in

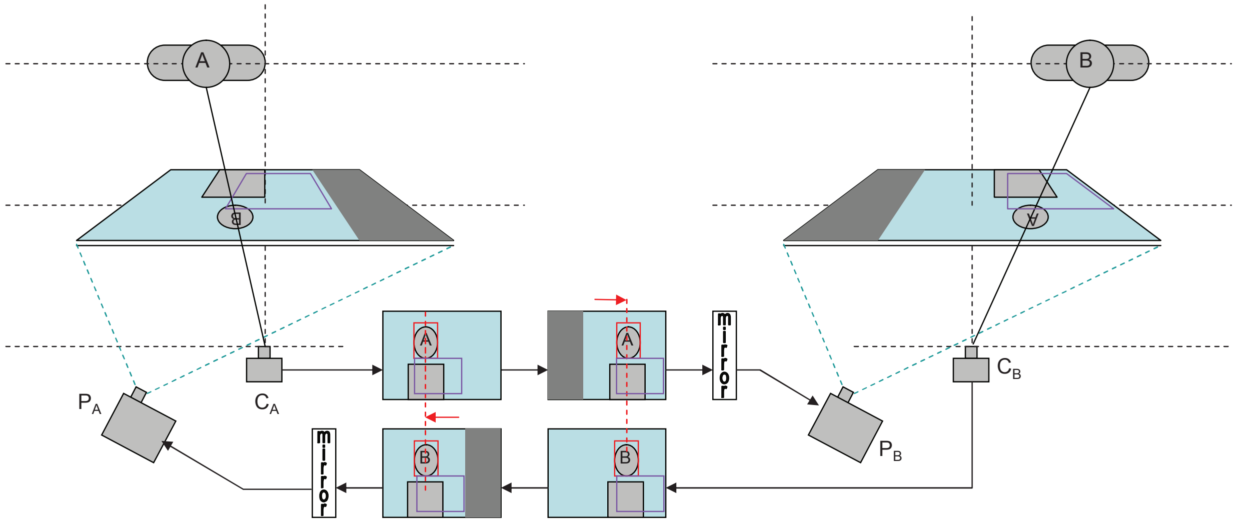

Fig. 4. Shifting shared content. Here the location of the shared content (purple rectangle) is shown in both video paths (note that the camera does not "see" the projected image, the content is digitally composited as shown in Fig. 1). Content is shifted by the same amount as the user image, thus preserving 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.

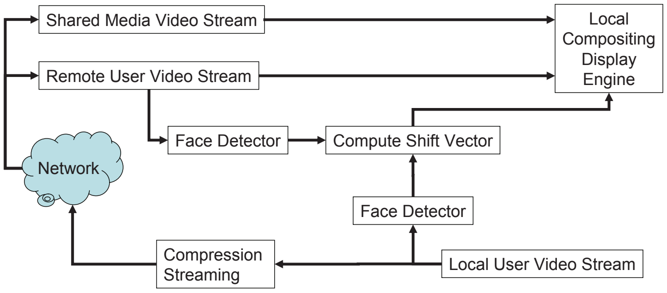

Fig. 5. Real-time processing pipeline.

a one-to-many 3d video teleconferencing system," *ACM Trans. Graph.*, vol. 28, no. 3, pp. 1–8, 2009, Proceedings of ACM **SIGGRAPH** 2009.

[3] David Roberts, Robin Wolff, John Rae, Anthony Steed, Rob Aspin, Moira McIntyre, Adriana Pena, Oyewole Oyekoya, and Will Steptoe, "Communicating eye-gaze across a distance: Comparing an eye-gaze enabled immersive collaborative virtual environment, aligned video conferencing, and being together," in *Proceedings IEEE Virtual Reality Conference*, 2009.

[4] H. Harlyn Baker, Nina Bhatti, Donald Tanguay, Irwin Sobel, Dan Gelb, Michael E. Goss, Bruce W. Culbertson, and Thomas Malzbender, "Understanding performance in coliseum, an immersive videoconferencing system," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 1, no. 2, pp. 190–210, 2005.

[5] Jim Gemmell, Kentaro Toyama, C. Lawrence Zitnick, Thomas Kang, and Steven Seitz, "Gaze awareness for video-conferencing: A software approach," *IEEE MultiMedia*, vol. 7, pp. 26–35, 2000.

[6] Milton Chen, "Leveraging the asymmetric sensitivity of eye contact for videoconference," in *Proceedings of the ACM SIGCHI conference on Human factors in computing systems (***CHI***)*, 2002, pp. 49–56.

[7] Jess Oppenheimer, "Prompting apparatus," US Patent 2883902, 1959, Filed Oct 1954.

[8] Stephen R. Acker and Steven R. Levitt, "Designing videoconference facilities for improved eye contact," *Journal of Broadcasting and Electronic Media*, vol. 31, no. 2, pp. 181–191, 1987.

[9] Bill Buxton and Tom Moran, "Europarc's integrated interactive intermedia facility (iiif): early experiences," in *Proceedings of the IFIP WG 8.4 confernece on Multi-user interfaces and applications*, Amsterdam, The Netherlands, The Netherlands, 1990, pp. 11–34, Elsevier North-Holland, Inc.

[10] Hiroshi Ishii and Minoru Kobayashi, "Clearboard: a seamless medium for shared drawing and conversation with eye contact," in *Proceedings of the ACM SIGCHI conference on Human factors in computing systems (***CHI***)*, 1992, pp. 525–532.

[11] Shinichi Shiwa and Morito Ishibashi, "A large-screen visual telecommunication device enabling eye contact," *SID Digest*, vol. 22, pp. 327–328, 1991.

[12] Markus Gross, Stephan Würmlin, Martin Naef, Edouard Lamboray, Christian Spagno, Andreas Kunz, Esther Koller-Meier, Tomas Svoboda, Luc Van Gool, Silke Lang, Kai Strehlke, Andrew Vande Moere, and Oliver Staadt, "blue-c: a spatially immersive display and 3d video portal for telepresence," in *Proceedings ACM SIGGRAPH*, 2003, pp. 819–827.

[13] Shahram Izadi, Steve Hodges, Stuart Taylor, Dan Rosenfeld, Nicolas Villar, Alex Butler, and Jonathan Westhues, "Going beyond the display: a surface technology with an electronically switchable diffuser," in *Proceedings 21st annual ACM symposium on User interface software and technology (***UIST***)*, 2008, pp. 269–278.

[14] Martin Kuechler and Andreas Kunz, "Holoport - a device for simultaneous video and data conferencing featuring gaze awareness," in *Proceedings of the IEEE conference on Virtual Reality*. 2006, pp. 81–88, IEEE Computer Society.

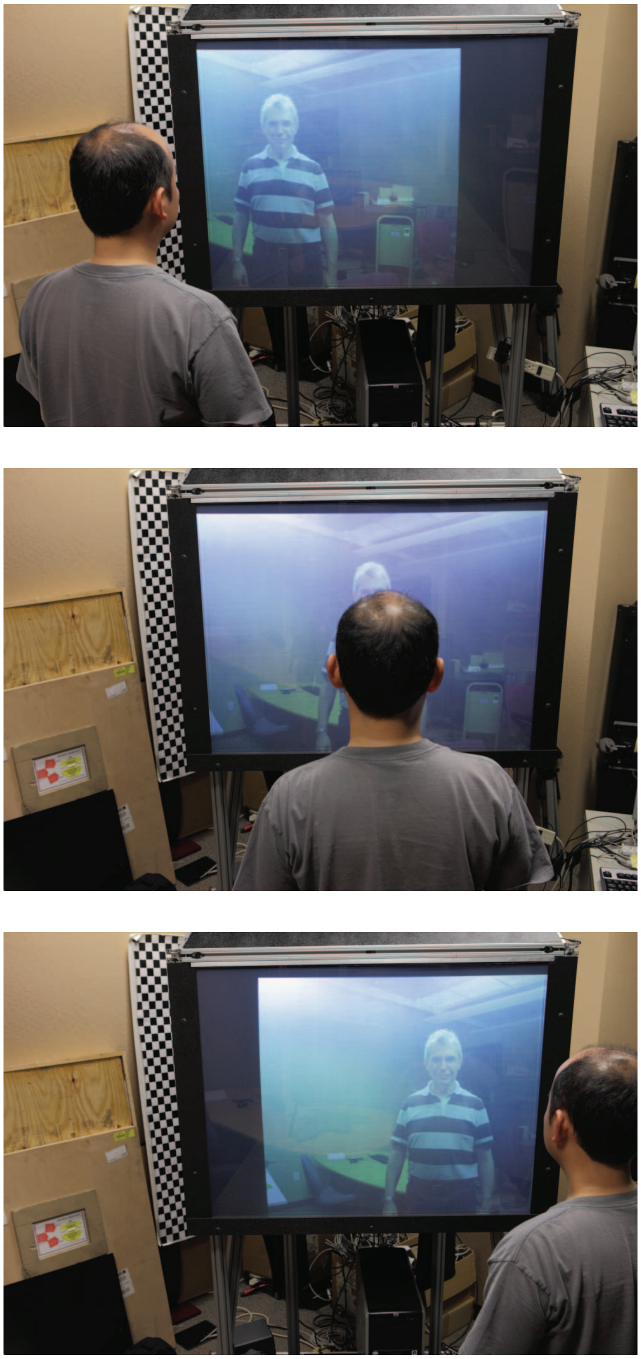

Fig. 6. Enabling eye contact. As the local user moves from the left side of the display to the right, the image of the remote user is automatically shifted so that the camera, local user's face, and remote user's face is always collinear. This way when the local user looks at the face of the remote user, 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.

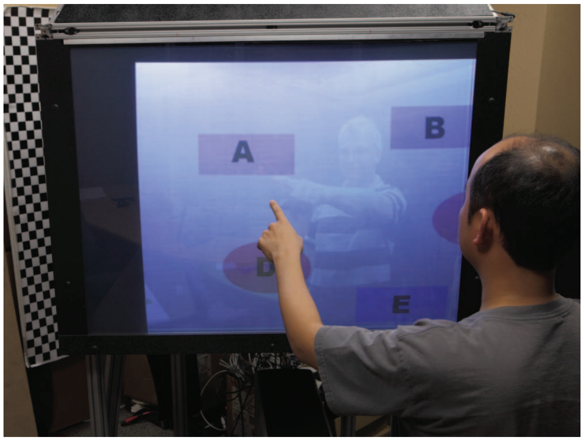

Fig. 7. Ensuring gaze awareness. As the remote user's image is shifted, the shared media (shown here digital composited) layer is also shifted correspondingly to ensure a correct rendering of the remote user's gaze direction with respect to shared content. At the same time, the users can make eye contact if they look at each other in the eye.

[15] Andrew Wilson, "Touchlight: An imaging touch screen and display for gesture-based interaction," in *Proceedings International Conference on Multimodal Interfaces (***ICMI***)*, 2004.

[16] H. Jorke and M. Fritz, "A new stereoscopic visualization tool by wavelength multiplex imaging," in *Proceedings Electronic Displays*, Sep 2003.

[17] Jan J. Koenderink, Andrea J. van Doorn, Astrid M. L. Kappers, and James T. Todd, "Pointing out of the picture," *Perception*, vol. 33, pp. 513–530, 2004.

[18] William Hyde Wollaston, "On the apparent direction of eyes in a portrait," *Philosophical Transactions of the Royal Society of London*, vol. 114, pp. 247–256, May 1824.

[19] Paul Viola and Michael Jones, "Robust real-time object detection," *Compaq CRL Technical Report*, , no. CRL-2001-01, February 2001, http://www.hpl.hp.com/techreports/Compaq-DEC/CRL-2001-1.html.

[20] Darryl Greig, "Video object detection speedup using staggered sampling," in *Proceedings IEEE Workshop on Applications of Computer Vision (***WACV***)*, 2009.

[21] Donald Tanguay, Dan Gelb, and H. Harlyn Baker, "Nizza: A framework for developing real-time streaming multimedia applications," HP Labs Tech Report HPL-2004-132 http://library.hp.com/techpubs/2004/HPL-2004-132.html, 2004.

[22] Jan J. Koenderink, Andrea J. van Doorn, Astrid M. L. Kappers, and James T. Todd, "Ambiguity and the 'mental eye' in pictorial relief," *Perception*, vol. 30, pp. 431–448, 2001.

[23] James T. Todd, Augustine H. J. Oomes, Jan J. Koenderink, and Astrid M. L. Kappers, "On the affine structure of perceptual space," *Psychological Science*, vol. 12, no. 3, pp. 191–196, May 2001.

[24] Dejan Todorović, "Geometrical basis of perception of gaze direction," *Vision Research*, vol. 46, pp. 3549–3562, 2006.

[25] S.M. Anstis, J.W. Mayhew, and Tania Morley, "The perception of where a face or television 'portrait' is looking," *American Journal of Psychology*, vol. 82, no. 4, pp. 474–489, 1969.

[26] Cha Zhang, Zhaozheng Yin, and Dinei Florencio, "Improving depth perception with motion parallax and its application in teleconferencing," in *Proceedings IEEE Workshop on Multimedia Signal Processing (***MMSP***)*, 2009.