Preserving Non-verbal Features of Face-to-Face Communication for Remote Collaboration

Jörg Edelmann[1], Philipp Mock[2], Andreas Schilling[2], and Peter Gerjets[1]

[1] Knowledge Media Research Center,
Schleichstr. 6, 72076 Tübingen, Germany
[2] University of Tübingen,
Sand 13, 72076 Tübingen, Germany

Abstract. Distributed working groups rely on collaboration systems that promote working on a project cooperatively over a distance. However, conventional systems for remote cooperative work do not transport important non-verbal cues of face-to-face communication like eye-contact or gaze awareness that would be available in co-located collaboration. Additionally, reference material and annotation tools should be readily accessible for all users. The screen layout should moreover create awareness for the transmitted video of remote participants and reference material alike and allow users to easily follow both at the same time. This paper describes how the presented system Face^{2}Face meets these requirements and thereby supports the collaborative design process. Furthermore, the performance of the system is evaluated in order to validate its practical applicability.

Keywords: remote collaboration, telepresence, face-to-face, multi-touch, awareness.

1 Introduction

Face-to-face conversations contain a large amount of non-verbal communication like eye contact, gestures and facial expressions which are substantial for fluid and natural discussions. However, project team members today often work distributed at different locations and depend on computer-mediated communication. Therefore, a collaboration tool for remote and synchronous interactions should try to preserve these non-verbal aspects, as they are an integral component of successful communication and facilitate collaboration at the same time. In the context of computer supported cooperative work (CSCW), interactive displays have shown to incorporate smoothly in co-located collaboration with small groups (e.g., [1][2]). This is because the technology can easily be integrated into the communication flow, as everyday gestures can be used to include digital material into the discussion. Accordingly, interactive displays naturally enable many aspects of workspace awareness [3].

However, in remote scenarios many of the above mentioned communication aspects get lost due to technological or conceptual shortcomings. In this paper,

Y. Luo (Ed.): CDVE 2013, LNCS 8091, pp. 27–34, 2013.

we present our ongoing work with Face^{2}Face, a system for video based remote collaboration with multi-touch interaction. The system aims at making the benefits that interactive displays provide for co-located collaboration applicable for remote scenarios. In previous work, we focused on the system setup [4] and depicted improvements in image quality and application scenarios [5]. Here, we highlight the practical implementation and how the system supports the collaborative design process for distributed groups.

2 Related Work

Understanding activities of other group members is an hugely important factor of collaboration and is commonly referred to as "awareness" in the CSCW context [6]. Gutwin and Greenberg [3] point out the importance of workspace awareness for distributed collaboration and provide a detailed analysis of different communication channels. For co-located scenarios, Hornecker *et al.* showed that the positive indicators of awareness can be increased with interactive displays in comparison to traditional mouse input, as the users can follow other users physical actions on the display more effortless [7]. Different hardware setups were proposed to transfer the benefits of interactive displays for co-located collaboration to a remote scenario. Tang *et al.* proposed a system based on multiple tabletop devices which not only synchronizes the application state among the participating sites but also provides an embodiment of other users' actions on the screen [8]. This is realized by capturing each tabletop device with an additional camera from above. The hand contours are then extracted and visualized for the other users as colored shadows. Additional displays located on chairs around each tabletop show a video stream of the respective other users. By separating the collaborative workspace from the video embodiment, the users have to switch attention explicitly between interaction and conversation. Face^{2}Face in contrast integrates the shared application seamlessly into the video image of the remote participant and thereby follows the concept of "Clearboard" [9]. This system allows two remote participants to create digital drawings in a face-to-face situation. Here, the user is captured from above with a half-mirror installation, thereby gaze directions and eye contact is supported. However, artifacts occur for objects above the display, e.g., the users' hands, as both the objects themselves and their reflection appear in the camera image. Therefore, we use holographic projection screens to capture the user in front of the display through the screen. A similar screen camera setup was used with "TouchLight" presented by Wilson [10]. Here, the transparent projection screen was used to capture user interactions on and before the screen with an infrared stereo camera setup from behind the display. "ConnectBoard" [11] and "HoloPort" [12] use this configuration of camera and see-through display to create systems which provide the remote collaboration features of "ClearBoard". Face^{2}Face enhances these concepts by integrating collaborative multi-touch which naturally extends the interaction capabilities. In order to enhance immersion and to intensify the impression of co-presence, we additionally integrated 3D stereo capturing and display.

3 System Description

Each client installation of Face^{2}Face is basically a vertically mounted touchscreen. The display is transparent so that the user and all of his or her interactions, even on the display surface, can be captured through the screen (Fig. 1). For remote collaboration, two clients from different locations are interlinked: The video streams of each site are transmitted and rectified to the screen geometry of the remote installation. Shared digital material on the screen is synchronized and can be interacted with virtually from both sides using multi-touch gestures. By superimposing the shared workspace onto the transmitted video stream, the illusion of having one transparent interactive workspace in between the two sites is created. The camera is placed behind the screen on eye-level. That way, display and camera are on one axis, when a user stands directly in front of the screen. As a result, gaze directions are preserved and can be displayed correctly on the other client. Moreover, as the whole screen area is captured, on-screen interactions are also captured as a whole.

Touches are recognized using an optical approach called "laser light plane" (LLP): for a LLP setup, several infrared (IR) lasers equipped with a line generator lens span a thin plane of IR light directly in front of the screen surface. An additional camera with an IR bandpass filter is mounted behind the display. Fingers touching the screen are then illuminated with IR light and thus become visible in the camera image. The user video and visual markers are captured with regular consumer cameras (for further details and extension to stereo 3D, see [4]).

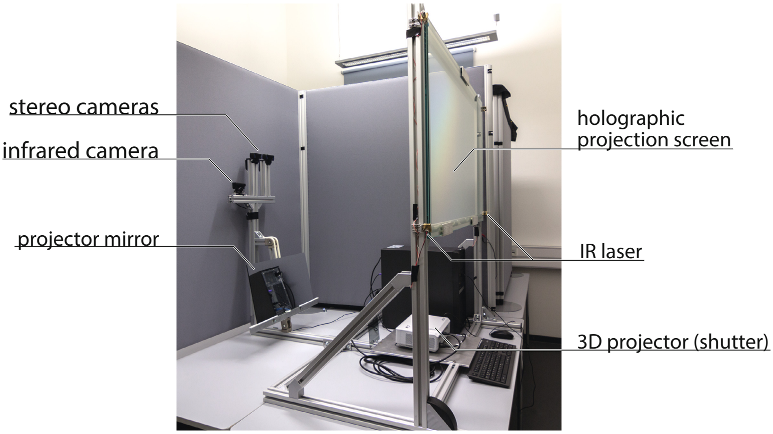


Fig. 1. Hardware components of Face^{2}Face

The application content is synchronized using Adobe® Media Server (AMS) (Fig 2). For each item in the shared workspace, a remote shared object exists on the server and is synchronized with each client. Each of those remote shared objects encapsulates all information that is necessary to describe and share an item: transformation matrix and an annotation layer which is stored as a bitmap. A manipulation on one of the clients triggers the synchronization for the active item which locks the element for the respective other user. Thereby conflicts and accidental interferences are avoided. Different property changes of the remote shared objects are updated separately. To keep the required bandwidth for transmitting annotations low, the data is updated only in adjustable intervals of several milliseconds while one user is drawing. The resulting lag is slightly visible, yet it is the best compromise between quality and induced network load.

Video streams are also transmitted via the AMS. The rectification and image enhancement is performed by the receiving client according to the initially exchanged calibration information of each installation [5]. In order to reduce computational time, image processing of the video stream is performed utilizing the graphics processing unit. A discussion of video latencies with varying compression quality settings and video resolutions can be found in the results section of this paper.

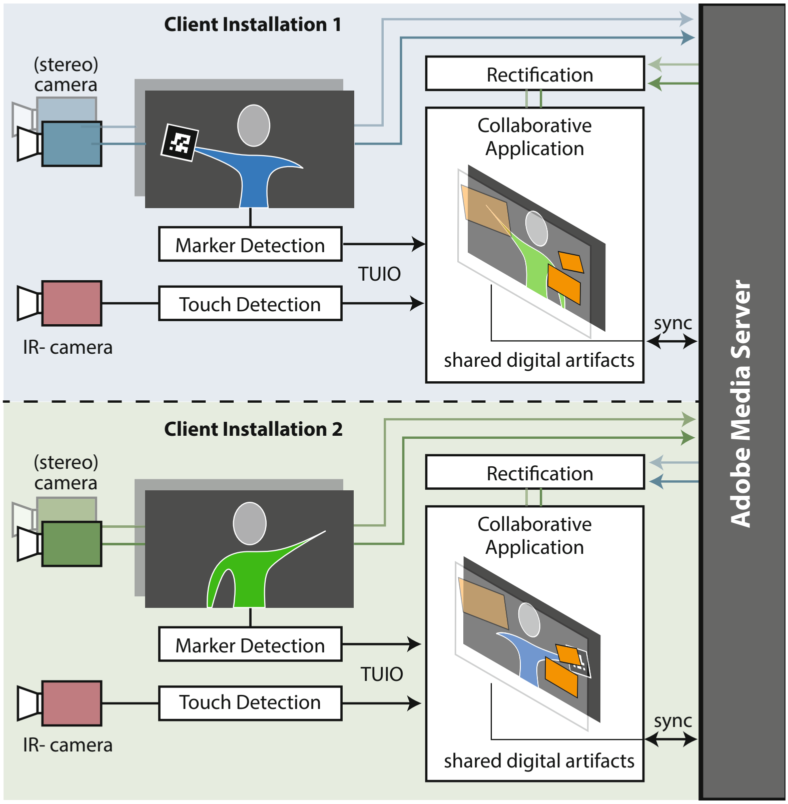


Fig. 2. Schematic system overview and synchronization workflow

4 Supporting the Collaborative Design Process

With Face^{2}Face two remote sites can be connected. To depict the benefits of system for cooperative design, we exemplify the workflow in the context of car engineering: the constructor responsible for the car body (in this example "user blue", see Fig. 2) wants to discuss latest design changes with two co-workers responsible for engine installation space ("users green"). As the team members work at different locations they set up a meeting using Face^{2}Face. As the system requires a fairly complex installation and is quite space consuming, we foresee the system to augment meeting rooms rather than regular workplaces. Due to large screen size and multi-touch interaction, multiple users per site can collaborate simultaneously. In order to integrate Face^{2}Face into the workflow of a distributed group, we propose a simple way to bring data to a digital meeting. User blue can collect digital material like renderings, technical documentations or 3D models in shared cloud storage at his or her workplace. In order to *transport* the content from the workplaye to the Face^{2}Face system, a visual marker that encodes the URL of the shared material location can be genrated on a portable device system that encodes the url of the shared material and that can be generated on a portable device. As the users establish connection, user blue shows the marker to the integrated camera and the shared material is loaded and presented on both installations. Now the users can discuss intuitively the design of the car body. Additional annotations can be applied to the visual material which appear true-sided for all participants.

In order to store the annotations or share the results with other team members via email additional user interface elements are required. In order to avoid consuming further screen space and occluding the video stream of the other user, tangible objects with unique visual markers can be used to select additional program functions. E.g., a tangible object referring to the email function can be used to send annotations to members of the work group that did not attend the meeting. In contrast to a common video conference application, Face^{2}Face extends remote collaboration with the following awareness features:

Presence: For group meetings, it is fundamental to understand who is present and who performs which action. As Face^{2}Face utilizes live video transmission, the presence and identity of participating group members is clear and transparent. In contrast to regular video conferencing, users actions in the shared workspace are also directly visible and thereby authorship of interactions can be determined intuitively.

Eye Contact: With the co-axial arrangement of camera and screen, Face^{2}Face supports eye contact. Thereby, conversation flow is improved as the next speaker can be negotiated by making eye contact. This also works for multiple users per site: Users green can differentiate which person user blue is currently looking at.

Gaze Awareness: The system correctly reproduces gaze directions also for the overlaid digital material. Thereby, the users are implicitly aware of which elements are in the others' focus of attention. This can serve as a visual evidence that the participants are referring to the same material (Fig 3). In order to avoid false sided digital material, the video stream of the respective other installation is vertically mirrored. Thereby, the digital material can be presented true sided on both clients and gaze awareness is still provided (Fig 3)

Gesture / Interaction Awareness: An important aspect of Face^{2}Face is the visibility of other users interactions with the digital content. Also, pointing gestures can be naturally used and serve as deictic reference and thereby reduce the need for verbal coordination.

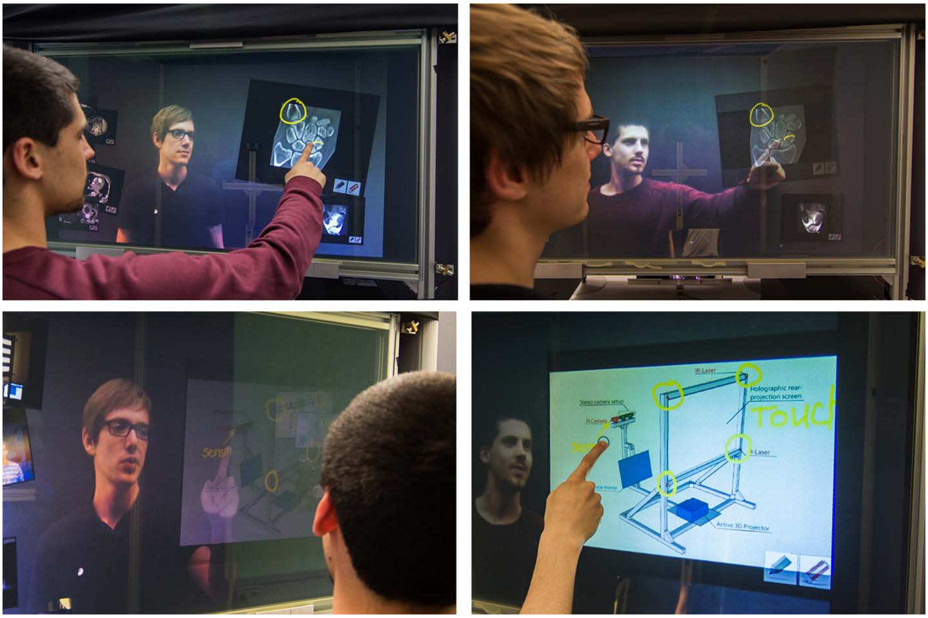
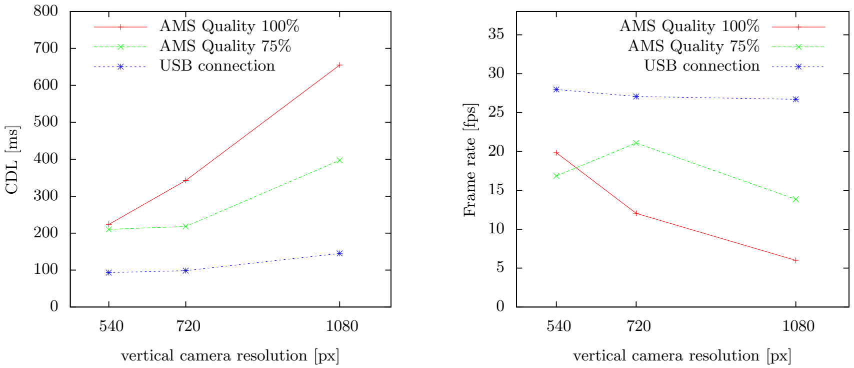


Fig. 3. Gaze direction and pointing gestures are preserved with Face^{2}Face. Annotations appear true sided on both client installations.

5 Results

For a persuasive remote collaboration, Face^{2}Face requires real-time high-definition video transmission and fluid workspace synchronization. Therefore, latency due to network transmission has to be reduced. In order to determine the general practicability of the concept and the system prototype, we measured the video performance transmission for different typical camera resolutions and compression quality settings made available by the AMS. We used the Tool “vDelay” [13] which allows measuring the capture-to-display latency (CDL) and the frame rate with a software based approach by capturing and detection of barcodes which encode the current system time. The measurements were performed using a switched gigabit Ethernet network. One of the two client installations additionally run the AMS. For comparison, the cameras were additionally crossplugged, thereby providing a direct USB transmission to the respective other installation.



(a) Capture-to-display latency (b) Video frame rate

Fig. 4. Performance of video transmission for different resolutions and quality settings

The results as depicted in Fig. 4 show the relationship between video resolution and CDL and video rate respectively. As the video stream has to be rectified a higher camera resolution leads to superior visual quality of the other user. With a direct USB connection the video latency of up to $145ms$ is not observable. In addition, with direct camera connection it can be shown, that the Face^{2}Face applications can potentially process and display also high resolution video streams at high frame rates. Of course, in a real scenario with distributed collaborators, the video stream has to be transmitted using a wide area network. With the AMS, a good compromise between display quality and latency is obtained with a camera resolution of $1280 * 720px$ with a compression quality setting of 75%. Here, the video latency of $217ms$ and the video rate of $21fps$ provide a satisfying user experience.

The bandwidth required for workspace synchronization is much smaller in comparison and is thus negligible. However, in case of higher video latency, the workspace synchronization should be delayed accordingly so that interactions of the respective other users are in sync with the video output. For the prototype implementation we used one of the clients to run the AMS in parallel to the Face^{2}Face application. Therefore, we expect performance improvements by using a dedicated server.

6 Conclusion and Outlook

In this paper we depicted how Face^{2}Face supports the collaborative design process by transporting important aspects of non-verbal communication and thereby increasing workspace awareness. Due to the hardware setup, there is no need for artificial embodiments. The setup creates a natural reproduction of a real face-to-face conversation and incorporates digital material seamlessly. Our performance evaluation moreover showed that the system is suitable for real-world application. A limitation is that Face^{2}Face is constraint to connecting two sites at a time. However, by using larger screen sizes it is possible to have more than one person interact simultaneously at a single client and thus increase the number of participants. For future work we plan to perform user experiments to gather empirical data on how the characteristics of this system affect workspace awareness. Aside from that we plan to implement and evaluate the proposed interaction techniques using the marker system and incorporate a variety of additional collaboration features in order to broaden the system's functional range.

References

1. Morris, M.R., Fisher, D., Wigdor, D.: Search on surfaces: Exploring the potential of interactive tabletops for collaborative search tasks. Information Processing & Management 46(6), 703–717 (2010); Collaborative Information Seeking
2. Isenberg, P., Fisher, D., Paul, S., Morris, M., Inkpen, K., Czerwinski, M.: Co-located collaborative visual analytics around a tabletop display. IEEE Transactions on Visualization and Computer Graphics 18(5), 689–702 (2012)
3. Gutwin, C., Greenberg, S.: The importance of awareness for team cognition in distributed collaboration. Technical Report 2001-696-19, University of Calgary, Alberta, Canada (2001)
4. Edelmann, J., Mock, P., Schilling, A., Gerjets, P., Straßer, W.: Face2face - a system for multi-touch collaboration with telepresence. In: Proc. of ESPA, pp. 159–162. IEEE (2012)
5. Mock, P., Edelmann, J., Schilling, A., Straßer, W.: Direct 3D-Collaboration with Face2Face – Implementation Details and Application Concepts. In: Proc. 3DTV – The True Vision (3DTV CON). IEEE, New York (2012)
6. Dourish, P., Bellotti, V.: Awareness and coordination in shared workspaces. In: Proceedings of the CSCW 1992, pp. 107–114. ACM, New York (1992)
7. Hornecker, E., Marshall, P., Dalton, N.S., Rogers, Y.: Collaboration and interference: awareness with mice or touch input. In: Proc. of CSCW, ACM, pp. 167–176. ACM (2008)
8. Tang, A., Pahud, M., Inkpen, K., Benko, H., Tang, J.C., Buxton, B.: Three's company: understanding communication channels in three-way distributed collaboration. In: Proc. of CSCW, pp. 271–280. ACM (2010)
9. Ishii, H., Kobayashi, M.: Clearboard: A seamless medium for shared drawing and conversation with eye contact. In: Proc. of the SIGCHI, pp. 525–532. ACM (1992)
10. Wilson, A.D.: Touchlight: an imaging touch screen and display for gesture-based interaction. In: Proceedings of Multimodal Interfaces, ICMI 2004, pp. 69–76. ACM, New York (2004)
11. Tan, K.H., Robinson, I.N., Culbertson, W.B., Apostolopoulos, J.G.: Connectboard: Enabling genuine eye contact and accurate gaze in remote collaboration. IEEE Transactions on Multimedia, 466–473 (2011)
12. Kuechler, M., Kunz, A.: Holoport - a device for simultaneous video and data conferencing featuring gaze awareness. In: Proc. of the IEEE Conference on Virtual Reality, VR 2006, pp. 81–88. IEEE (2006)
13. Boyaci, O., Forte, A., Baset, S., Schulzrinne, H.: vdelay: A tool to measure capture-to-display latency and frame rate. In: 11th IEEE International Symposium on Multimedia, ISM 2009, pp. 194–200 (2009)