FACE2FACE - A SYSTEM FOR MULTI-TOUCH COLLABORATION WITH TELEPRESENCE

Jörg Edelmann, Peter Gerjets

Knowledge Media Research Center
Schleichstrae 6, 72076 Tübingen

Philipp Mock, Andreas Schilling,
Wolfgang Strasser

WSI/GRIS
University of Tübingen
Sand 14, 72076 Tübingen

ABSTRACT

One major benefit of multi-touch interaction is the possibility for passive observers to easily follow interactions of active users. This paper presents a novel system which allows remote users to perform collaborative multi-touch interaction in a remote face-to-face situation with shared virtual material seamlessly integrated in a videoconference application. Each user is captured by a camera located behind the screen and touch interactions are thus made visible to the remote collaborator. In order to enhance the immersion of the setup, the system provides realistic stereoscopic 3D video capture and presentation. We highlight the concept and construction of the system and show sample applications which allow for collaborative interaction with digital 2D and 3D media in an intuitive way.

Index Terms— multi-touch, telepresence, stereoscopic 3D, natural user interface, CSCW

1. INTRODUCTION

Multi-touch technology has become a common human-computer interface available for handheld devices like smartphones or tablet computers, but also for larger devices like tabletop installations (e.g., Microsoft Surface) to provide an intuitive and direct access to digital information. One benefit of multi-touch capable displays with larger screen estate is the possibility to allow multiple users to interact with screen content simultaneously. It is therefore a favored interface for co-located computer supported collaborative work [1, 2]. Since the input is performed directly on the screen, other users can easily understand and follow the interaction performed by the active user.

In this paper, we present *Face2Face*, a system for collaborative multi-touch interaction with shared digital material between remote users in a virtual face-to-face situation. In order to enhance the illusion of co-presence, the video acquisition and presentation support stereoscopic 3D. We describe a feasible hard- and software solution based on transparent projection screens and optical touch sensing that allows for gaze awareness and recognition of touch interaction by remote users. Sample applications for collaborative interaction with 2D and 3D content are outlined.

2. RELATED WORK

An early successful example for collaboration with digital material integrated in a face-to-face conversation is ”Clearboard” by Ishii et al. [3]. This system allows remote participants to create digital drawings in a collaborative manner. Since the camera captures the user from above with a half-mirror installation, real eye-contact with the remote collaborator and gaze awareness is provided. However, artifacts occur when captured objects ,e.g., the user’s hands, are visible both directly in the camera image and in the mirrored image. The authors could show that a seamless integration of shared digital material in a face-to-face video conference application can reduce the amount of eye and head movements in comparison to desktop or whiteboard collaboration. Tang et al. present a system for remote collaboration with horizontal tabletop displays [4]. Interactions upon the tabletop are captured with an additional camera to provide the other team members a shadow-like representation of remote users’ arms on the display. Additional displays are located around the tabletop where a video stream of each participant is shown in order to create the illusion of the other users being located around the tabletop system. Wilson presented with ”Touchlight” [5] an interactive optical touch screen based on a transparent holographic projection surface. To identify touch interactions upon the screen, an infrared stereo camera setup captures the screen area through the transparent display. In this system the rectified stereo image pair is multiplied and bright regions are identified as interactions upon the screen. Jones et al. introduced a one-to-many 3D-videoconferencing system providing eye contact [6]. Here, the user’s face is scanned in 3D with a structured light approach and is presented to the remote audience on an autostereoscopic horizontal-parallax 3D display. However, collaborative interaction with additional virtual material is not in the scope of this system.

978-1-4673-0898-4/12/$31.00 ©2012 IEEE

3. SYSTEM DESCRIPTION

Face2Face is based on the same collaboration metaphor as ”Clearboard”, but extends the interaction space by providing a multi-touch interface and enhances the illusion of co-presence with stereoscopic acquisition and presentation of the remote user. The hardware setup (Fig. 1) comprises a transparent projection with optical multi-touch acquisition similar to ”Touchlight”, but uses monocular touch sensing. In the following, the system components are presented in detail.

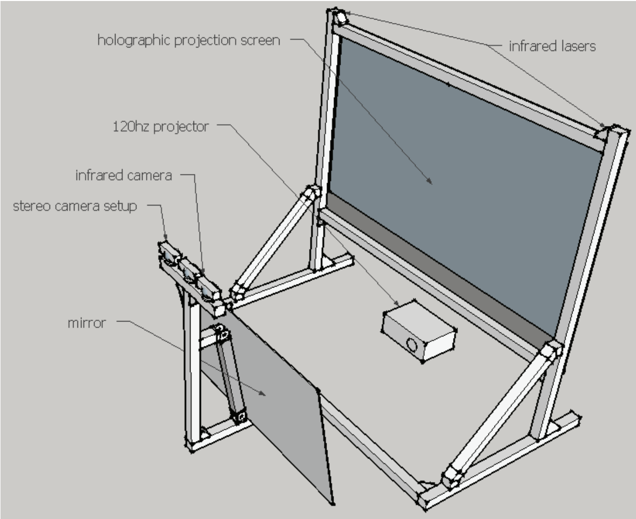

Fig. 1. Face2Face system components of a single client installation

3.1. Multi-touch Sensing

In order to detect the location of the users' fingers upon the screen, we use an optical *Laser-Light-Plane* (LLP) setup: infrared lasers are placed at each corner of the screen. The laser beams are split into thin planes of light by special lenses mounted to each laser. All of the four lasers are aligned in order to merge to one thin plane that runs parallel close to the screen surface. An infrared camera mounted behind the transparent screen senses objects that intersect the laser plane which appear as bright regions in the camera image. The image acquisition of the infrared camera is not affected by the projection that is thrown onto the transparent screen, since it operates in a different light spectrum. A range of image processing filters similar to [7] are applied to the camera image in order to identify connected components, which are tracked over time. These are used as input for multi-touch applications. Compared to other optical methods for touch sensing [8], LLP does not require a diffuse screen material and is therefore favored for *Face2Face*.

3.2. Projection and Video Capturing

The system is designed to capture the user's fingers from behind the screen. To allow for gaze awareness and eye contact, the cameras capturing the user and the display have to be co-axial. Accordingly, the cameras are placed behind the screen. In order to capture through the display, a transparent rear-projected screen has to be used. Unfortunately, in this arrangement, the projection can interfere with the camera images since some of the light is reflected from the projection screen. For the best possible image quality of the user videos, the transparency of the used screen material is crucial. Accordingly, different transparent screens were tested. In the end, a screen with holographic optical elements was used, because of it's high gain-factor and good image quality. In addition, the projected image is hardly visible from the cameras point of view under daylight conditions. Furthermore, different illumination levels and polarizing filters were tested. Opposing polarizing filters for cameras and projector can be used to render the light of the projector invisible to the camera. However, this comes at the cost of losing about half of the projected light due to polarization. With the incorporated holographic screen, appropriate lighting conditions for camera capturing turned out to be the most important factor for satisfying results. The cameras are calibrated to the display and the image is warped accordingly to ensure that remote user interactions match the screen content. The captured video is digitally mirrored before being displayed to the other user. This way, on-screen content can be displayed true-sided on both clients with touch interactions matching the corresponding user video stream at the same time.

3.3. Stereoscopic 3D

In order to enhance the illusion of having an conversational partner actually present behind the transparent screen, the users are captured with a stereoscopic camera setup. The stereoscopic image is created by two toed-in cameras that both focus the center point of the screen. The amount of perceived depth correlates to the camera placement relative to the screen. For a given distance (determined by the fixed camera lens), the interocular distance between the two cameras is the decisive unknown. A formula for an adequate camera distance has been derived from [9]:

$$A = \frac{S\, d_F\, F'}{F' - Z'} \quad \text{with} \quad S = \frac{W'}{W}$$

Here, S is the scaling factor between virtual and actual screen size (W' and W), F' is the distance to the farthest object behind the screen, Z' is the distance between camera and virtual display and d_F is the maximum disparity (depending on F', Z' and the camera distance A). Since the image regions that correspond to the actual display are displayed in full screen, S is equal to 1. The basic geometry using this simplification is shown in Fig. 2. In the presented setup, this results in an optimal distance of about 7,5cm for correct depth perception for the expected viewer position (centered and located about 80cm distance from the screen). It shall be noted, however, that this might lead to excessive disparity for more distant objects.

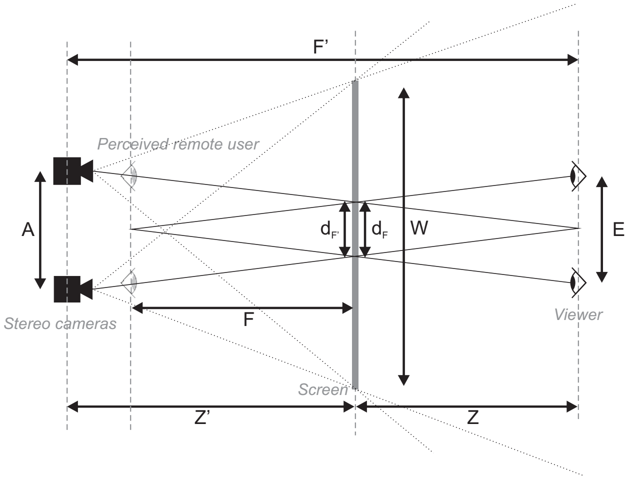

Fig. 2. Stereo geometry: F and F' are the distances from stereo camera to screen and to the user, d_F and $d_{F'}$ are the disparities on the display and the virtual display. The interocular distance and stereo base are denoted as E and A. Z and Z' are the distances from screen to user and camera. W is the screen width.

3.4. Synchronization

The client applications have to be synchronized in order to maintain the system's inherent metaphor. For the presented system, this is achieved by synchronizing the touch input for both clients: Touch events are transmitted from each client installation via TUIO protocol [10] to a multiplexing server. This server maps the IDs of currently active fingers to discriminative ranges and distributes the altered events to all connected clients. Therefore, all clients receive equal input and touch sources are distinguishable. This is necessary to avoid fingers from separate sources to form joint interactions. In our case, locks are created on active objects and successive touches from other sources are ignored. This method of synchronization does not account for packet loss which can lead to unsynchronized client applications and more sophisticated methods have to be implemented for real-world usage. However, for the prototype installation where the clients are connected via ethernet, packet loss is negligible.

4. APPLICATIONS

To demonstrate the capabilities of *Face2Face*, two sample applications were implemented to illustrate the interaction and collaboration behavior with *Face2Face* and to highlight some of the system's main features.

4.1. 3D Model Viewing

The illusion of a transformable virtual object between two remote users is exemplified by a 3D model viewing application. The virtual workspace shared by the two remote users consists of a textured model, which is rendered upon the camera image of the respective user (figure 3). The shared 3D object can be scaled, moved and rotated by one- or two-finger multitouch interactions. Since the 3D model should appear as in between the participants, it is presented from opposite sides to the users.

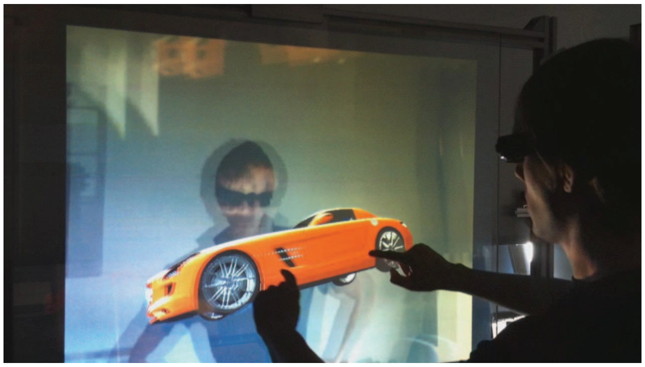

Fig. 3. The 3D model viewing application with stereoscopic 3D enabled.

4.2. Collaborative Media Application

To visualize the interaction with 2D media items, an application for collaborative interaction with image and video objects has been created (see Fig. 4). The shared workspace consists of a collection of two-dimensional media objects, which can be composed of images, videos or textual information. Both users can move, rotate and resize the media items with multitouch interaction. As one user touches an object it gets locked and can't be manipulated by the other user in order to avoid inferences. To make pointing gestures to larger objects, which would occlude the remote user, more comprehensible, the images become sightly transparent as soon as they get touched. It is possible to finger-draw onto the screen to create notes or group the data. Text appears true sided on both clients because of mirrored video streams.

5. CONCLUSION AND FUTURE WORK

We have introduced a system for remote multi-touch interaction with shared digital material, integrated in a natural face-to-face conversation situation. On top of a feasible hardware solution based on transparent holographic projection screens

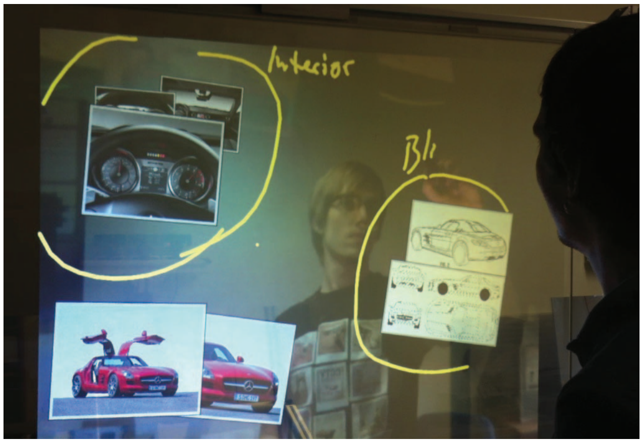

Fig. 4. The collaborative media application in practice. On screen text appears true-sided for both participants.

combined with optical multi-touch sensing, we presented two types of collaborative applications which allow for remote interaction with digital 2D and 3D content. Though formal user studies have not been carried out so far, informal user testings showed that the illusion of having digital material in between the users, which is accessible from both sides, works very well. In case of shared 3D material, all users assumed that the opponent user perceives the respective other side of the model. We see this as an indicator that the intended illusion of the overall system is well established and intuitively understood. The value of stereoscopic 3D integration is twofold: one the one hand it seems to enhance the immersion of the system, but on the other hand it degrades gaze awareness since the users has to wear shutter glasses. Another problem with 3D material occurs with touch interaction when stereoscopic projection is enabled: since the screen is transparent it is difficult to estimate the location of the display for touch interaction.

For future work we therefore plan to integrate an additional in-air gesture interface. First tests with a Microsoft Kinect depth sensor have shown that this could be a feasible approach since this device can be operated through the holographic projection screen. This would also allow for a correct compositing, accounting for occlusion of the user video with the virtual scene. In order to gain insight on how this system can improve remote collaboration, formal users studies have to be carried out. Especially the interaction with 3D content and the value of stereoscopic presentation are topics we want to focus on.

6. REFERENCES

[1] Saleema Amershi and Meredith Ringel Morris, “Cosearch: a system for co-located collaborative web search,” in *Proceeding of the SIGCHI conference*, New York, NY, USA, 2008, pp. 1647–1656, ACM.

[2] Björn Hartmann, Meredith Ringel Morris, Hrvoje Benko, and Andrew D. Wilson, “Pictionaire: supporting collaborative design work by integrating physical and digital artifacts,” in *Proceedings of the CSCW*, New York, NY, USA, 2010, pp. 421–424, ACM.

[3] Hiroshi Ishii and Minoru Kobayashi, “Clearboard: a seamless medium for shared drawing and conversation with eye contact,” in *Proceeding of the SIGCHI conference*. 1992, pp. 525–532, ACM.

[4] Anthony Tang, Michel Pahud, Kori Inkpen, Hrvoje Benko, John C. Tang, and Bill Buxton, “Three’s company: understanding communication channels in three-way distributed collaboration,” in *Proceedings of CSCW*, New York, NY, USA, 2010, CSCW ’10, pp. 271–280, ACM.

[5] Andrew D. Wilson, “Touchlight: an imaging touch screen and display for gesture-based interaction,” in *Proceedings of Multimodal interfaces*, New York, NY, USA, 2004, ICMI ’04, pp. 69–76, ACM.

[6] Andrew Jones, Magnus Lang, Graham Fyffe, Xueming Yu, Jay Busch, Ian McDowall, Mark Bolas, and Paul Debevec, “Achieving eye contact in a one-to-many 3d video teleconferencing system,” in *ACM SIGGRAPH 2009 papers*, New York, NY, USA, 2009, pp. 64:1–64:8, ACM.

[7] Jörg Edelmann, Sven Fleck, and Andreas Schilling, “The dabr – a multitouch system for intuitive 3d scene navigation,” in *3DTV CON - The True Vision*, 2009.

[8] Johannes Schoening, Jonathan Hook, Nima Motamedi, Patrick Olivier, Florian Echtler, Peter Brandl, Laurence Muller, Florian Daiber, Otmar Hilliges, Markus Loechtefeld, Tim Roth, Dominik Schmidt, and Ulrich von Zadow, “Building interactive multi-touch surfaces,” *journal of graphics, gpu, and game tools*, vol. 14, no. 3, pp. 35–55, 2009.

[9] Graham Jones, Delman Lee, Nicolas Holliman, and David Ezra, “Controlling perceived depth in stereoscopic images,” in *STEREOSCOPIC DISPLAYS AND VIRTUAL REALITY SYSTEMS VIII*. 2001, pp. 200–1, SPIE.

[10] M. Kaltenbrunner, T. Bovermann, R. Bencina, and E. Costanza, “Tuio: A protocol for table-top tangible user interfaces,” in *Proc. of the The 6th International Workshop on Gesture in Human-Computer Interaction and Simulation*, 5/1/2006 2005.