3D-Board: A Whole-body Remote Collaborative Whiteboard

Jakob Zillner[1], Christoph Rhemann[2], Shahram Izadi[2], Michael Haller[1]

[1]Media Interaction Lab, University of Applied Sciences Upper Austria
[2]Microsoft Research

ABSTRACT
This paper presents *3D-Board*, a digital whiteboard capable of capturing life-sized virtual embodiments of geographically distributed users. When using large-scale screens for remote collaboration, awareness for the distributed users' gestures and actions is of particular importance. Our work adds to the literature on remote collaborative workspaces, it facilitates intuitive remote collaboration on large scale interactive whiteboards by preserving awareness of the full-body pose and gestures of the remote collaborator. By blending the front-facing 3D embodiment of a remote collaborator with the shared workspace, an illusion is created as if the observer was looking through the transparent whiteboard into the remote user's room. The system was tested and verified in a usability assessment, showing that *3D-Board* significantly improves the effectiveness of remote collaboration on a large interactive surface.

Author Keywords
Remote Collaboration, Shared workspace, Teleconferencing, Interactive Whiteboard, Reconstruction

Figure 1: The virtual 3D embodiment of a remote user superimposed on top of the digital content of the interactive whiteboard.

INTRODUCTION
Interactive whiteboards in combination with video conferencing systems are becoming evermore popular. In current solutions, the video image of a geographically distributed user and the shared workspace are separated resulting in a cumbersome collaboration. Systems such as [20, 12] demonstrate that superimposing the life-sized video of a remote user on top of the shared content can improve the cooperation. These related solutions mainly focus on 2D video images, showing a rear-view of the distributed participants. This paper presents *3D-Board*, a whiteboard featuring a novel visualization of the collaborator enhancing the experience of working remotely.

With *3D-Board*, we facilitate social presence in a shared working environment by blending a front-facing 3D embodiment of the remote user with the digital content of the interactive whiteboard (cf. Figure 1). Our solution creates the impression as if the remote collaborator would stand behind the shared, interactive whiteboard. Summarizing, the main contributions of this paper are:

- Enabling remote face-to-face collaboration on a large, interactive surface without the need for a holographic screen or half-silvered mirror, with a simple and new front-facing visualization technique.
- A study that demonstrates the benefits of our visualization technique that supports full body 3D virtual embodiment of the remote user. Our study shows that the 3D front facing approach improves the user awareness and provides additional features that makes remote collaboration tasks more efficient.
- Robustly segmenting the remote user from the background scene to avoid occlusion when superimposing the transparent, virtual embodiment on top of any digital content.
- And finally, our hardware setup is a self-contained unit, since all cameras can be easily mounted on the edges of a whiteboard and do not have to be behind or in front of the screen.

The design and results of this study are discussed after setting the context in the related work section and presenting our implementation. Finally, the overall findings are discussed alongside future work.

RELATED WORK
Facilitating awareness for the remote user is one of the most important challenges in computer supported cooperative working environments. Distributed tabletop and whiteboard applications often provide user embodiments to improve the awareness among collaborators [2]. Therefore, deictic gestures and gestures in general are highly essential for a distributed conversation [9, 14, 8].

Although the video image is serving as the connection between the distributed users [3], the video has to fuse with the actual shared content into a single, common workspace [21, 18, 25, 16, 26]. ClearBoard [15] is one of the most inspiring setups that combines face-to-face conversation and shared drawing activities into one screen. Based on the ideas of ClearBoard, Gumienny et al. developed Tele-Board [12]. Tele-Board features video- and data-conferencing by overlaying the transparent content of an interactive whiteboard over the full-screen rear or side video view of the remote collaborators. Similarly, CollaBoard [20] separates the user from the background to superimpose the full-sized back-view of the upper body on top of the digital data. While this approach provides accurate deictic gestures, natural interaction is limited since the face-to-face communication is restricted.

Our work is also inspired by Onespace [21], which utilizes a depth-sensing camera to segment users from the background and to get their position in 3D space. Therefore, a 2D representation can be placed in a virtual environment and the distributed partners can collaboratively interact with their digital surroundings. The front-view makes gaze-awareness and gestures possible, but the depth-sensing cameras force users to keep a certain distance to the screen. In addition, a mirrored representation of the co-located person needs to be displayed in order to raise the awareness of one's position in 3D space.

Maimone et al. [22] use multiple depth-sensing cameras to capture a scene in 3D. The reconstructed scene can then be viewed by utilizing a head-tracking system and an auto stereoscopic display. Similarly, Beck et al. present an immersive group-to-group telepresence setup by using multiple Kinect sensors [1]. TeleHuman [19] is a videoconferencing system that uses a cylindrical 3D display portal to feature a $360°$ view of a geographically distributed user. By supporting stereoscopic 3D and motion parallax, the system is capable of conveying accurate hand pointing gestures and body poses. Edelmann et al. focus on the interaction with digital content [7]. The system features a stereo camera setup behind a semi-transparent holographic multi-touch screen to capture the remote scene. This allows for a natural face-to-face collaboration on the interactive surface. However, the need for shutter glasses and the opaque holographic screen disrupt the telepresence experience. In addition, the remote user is not separated from the background, thus the video occludes much of the workspace.

The visualization and presentation of the virtual embodiment has a vast influence on the workspace awareness. Different possibilities have mostly been evaluated in the context of tabletops [6, 10, 27, 23]. Doucette et al. [6], for instance, used different levels of abstraction for arm embodiments while interacting with the tabletop. The remote collaborator therefore sees the local user's arms as either just a thin line, a colorized arm, or as a real picture. In the context of a tabletop setup, the results clearly show that almost all digital arm embodiments, regardless of the level of abstraction, do not significantly raise the awareness for the remote user. However, when dealing with interactive whiteboards, the presence of a virtual embodiment is highly relevant since the full body of the collaborator is visible. Gaze awareness [28], facial expressions, and hand pointing [17] play an important role while interacting with another person on such devices. By utilizing a large see-through display and face detection, Tan et al. [24] developed a system supporting eye contact and gaze by offsetting the video stream based on the viewer's position.

In contrast to the related work, our system is encumbrance free, featuring a 3D view of the remote collaborator without the need for shutter glasses. Utilizing the full size of the screen, the virtual embodiment can be placed over any digital content. Moreover, *3D-Board* features a natural face-to-face visualization of the remote user, without having to place a camera behind an opaque screen.

3D-BOARD

Creating the impression of looking through an interactive whiteboard into the remote collaborator's room requires capturing every pose and gesture of a person directly in front of the whiteboard. The recorded scene is then reconstructed and rendered in 3D at the local user's screen. In the following sections, we highlight the key system components and discuss in detail the implementation of our approach.

Prototype Setup and Hardware

Currently, our setup consists of two interactive whiteboards: one for capturing the remote user in front of the screen and the other for observing the virtual embodiment. Both boards have a size of $1.6m \times 1.2m$ with a resolution of 1280×800 pixels and were operated by a Vivitek D795WT short throw projectors, with input through Anoto digital pens (ADP 601) [13]. To avoid networking issues, both whiteboards are driven by the same PC with an Intel Core i7-3770 CPU, 8GB of RAM and a Nvidia Quadro K2000, 2GB GDDR5.

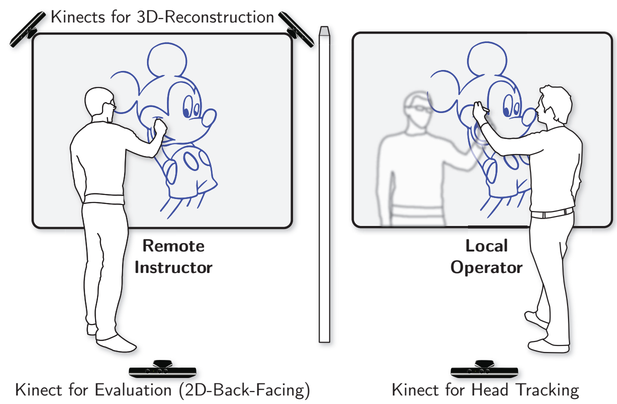


Figure 2: The prototype of *3D-Board*. The two upper Kinects capture the interaction of the *instructor* with the whiteboard (left). The *operator* (right) can observe the *instructor's* virtual embodiment in 3D by utilizing a third Kinect for head tracked motion parallax. A fourth Kinect, placed behind the *instructor*, was solely used for another technique (*2D-Back-Facing*) during the evaluation and is not part of *3D-Board*.

In total, four Microsoft Kinect (v1) cameras are used (cf. Figure 2). Two are attached to the top left and top right corner of the remote user's whiteboard. The data acquired from these depth sensors are used to reconstruct a 3D virtual embodiment of the geographically distributed person. Rotated downwards at a $45°$ angle, the two Kinects capture the user's actions from directly on the whiteboard's surface up to $1.2m$ away from the screen. Both sensors are operating in near-mode in order to cover the whole width and height of the board with only a minimal loss of about $10cm$ in the upper corner regions of the whiteboard. At the other (local) workspace, another Kinect sensor is capturing the full size of the second whiteboard from the back of the room. This camera tracks the head of the local observer to support motion parallax by changing the viewpoint of the rendering accordingly. The fourth Kinect is not part of the *3D-Board* setup. It was solely used to capture a 2D view of the back of the remote user for the second technique, discussed later in the evaluation section of this paper. With all

Kinects operated at a color and depth resolution of 640×480 pixels, the setup is running with approximately 30 Hz.

Camera Calibration

In order to merge the output of the two Kinect sensors attached to the upper corners of the remote user's whiteboard, the cameras need to be calibrated. The calibration is done using OpenCV's[1] implementation of the method of Zhang [29]. At first, multiple images of a checkerboard calibration target are taken from the RGB-camera in order to compute the camera intrinsics (focal length, principal point, distortion coefficients). Then the relative translation and rotation of the two cameras towards each other are determined using stereo camera calibration (by imaging the same calibration target simultaneously from both viewpoints). By computing the intrinsic and extrinsic camera parameters, the point clouds from the Kinect sensors can be transformed into a common coordinate space and rendered as a merged output.

Novel View Synthesis

Given the color and depth streams from multiple Kinect cameras located around the whiteboard, our goal is to render the user that is interacting with the board from a viewpoint of a virtual camera located behind the screen. This requires fusing the live color and depth data from all Kinect sensors in real time and rendering of the merged foreground data from a novel viewpoint. For accessing the Kinect's data, the official driver and SDK are used. The output is rendered in *DirectX* and displayed in a *Windows Presentation Foundation* (WPF) window.

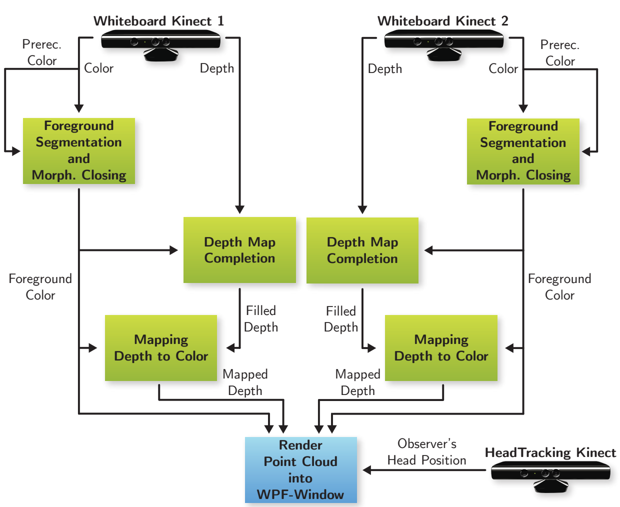


Figure 3: The rendering pipeline of *3D-Board*.

The data processing and rendering pipeline (cf. Figure 3) of *3D-Board* comprises of three subsequent steps:

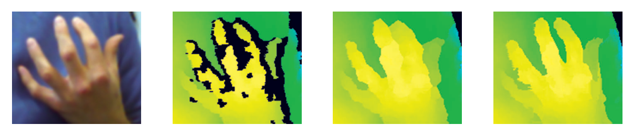
1. *Segmentation*: In order to separate the user from the background, a foreground mask is computed from the viewpoint of each Kinect camera.

2. *Depth Map Completion*: Since the depth output is often incomplete, we estimate the geometry (depth) for all foreground masks to fill small holes.

3. *Depth Map Fusion & Rendering*: The foreground geometry from the two sensors is fused into a single, complete 3D point cloud that can be rendered from a virtual viewpoint by utilizing the previously discussed camera parameters.

Segmentation Given the color image C from the Kinect sensor, we aim to separate the dynamic foreground regions from the static background via frame differencing. To this end, we pre-record an image C_{clean} that shows the scene without foreground objects. Every pixel i in the live color image C that differs more than a threshold θ from C_{clean} in color space is added to the set of foreground pixels Ω_F. We fill small holes in the foreground mask Ω_F by applying a morphological closing operation.

Depth Map Completion Next, we infer the geometry (i.e. depth) for each pixel inside the foreground mask Ω_F. Each Kinect sensor provides a depth map D that can be rendered from the viewpoint of the color camera of the Kinect. However, depth maps are usually incomplete due to errors in triangulation as well as due to interference of multiple sensors. Since the overlap between the two Kinect cameras in the *3D-Board* setup is very small, interference can be handled by our geodesic depth map filling algorithm and does not require an additional hardware solution [5].



(a) (b) (c) (d)

Figure 4: Geodesic depth map filling: (a) RGB input image (b) Depth map from Kinect (c) Euclidean NN filling (d) Geodesic NN filling

We denote the set of pixels with valid depth as Ω_D and the set of foreground pixels with valid depth as $\Omega_{FD} = \Omega_F \cap \Omega_D$. For pixel i inside the foreground region without a valid depth, we find the geodesic nearest neighbor NN_{geo} in Ω_{FD}. We then copy the depth from the nearest neighbor to pixel i: $D(i) = D(NN_{geo}(i))$. The geodesic nearest neighbor NN_{geo} in set Ω_{FD} is defined as:

$$NN_{geo}(i) = \arg \min_{j \in \Omega_{FD}} d(j, i), \tag{1}$$

where

$$d(j, i) = \min_{P_{j,i}} \sum_{k \in P_{j,i}} W(k). \tag{2}$$

$P_{j,i}$ is a path connecting pixels i and j and the weight $W(k)$ is defined by the image gradient at pixel k. Intuitively, the nearest geodesic neighbor is found along a path along which the color in the RGB image varies only slightly. This is in contrast to the euclidean nearest neighbor that is found along a straight line. Since the geodesic nearest neighbor takes the color image data into account, a better completion of the depth map can be achieved. An example of depth map filling using geodesic distances is shown in Figure 4 which depicts a hand in front of a person's chest (cf. Figure 4 (a)). The corresponding depth map from the Kinect sensor (cf. Figure 4 (b)) shows missing depth values between the fingers (shown in black in Figure 4 (b)). Filling the missing values with the nearest neighbor in

[1] `http://opencv.org/`

euclidean space (cf. Figure 4 (c)) leads to wrong depth values between the fingers. Filling with the closest pixel in geodesic space yields superior results (cf. Figure 4 (d)). We use the approximate algorithm of [4] to compute the geodesic nearest neighbor efficiently.

Depth Map Fusion & Rendering The registration of the filled depth image to the color image is done by the Kinect SDK. The resulting mapped depth texture and the foreground color texture can then be visualized as a colored point cloud in the shader pipeline.

The point cloud is constructed by projecting each pixel of the filled and mapped depth texture into 3D space by multiplying the depth pixels with the inverse of the camera intrinsics matrix. The resulting 3D point is expanded into a fixed sized quad in the geometry shader. The color of the quad is sampled from the foreground color texture. To keep the RGB values consistent across all cameras we set the exposure time and the color temperature for each Kinect sensor to the same fixed value. In addition, the size of each billboard is enlarged with increasing depth to avoid gaps in the visualization.

This efficient point-based rendering technique allows us to visualize multiple point clouds on top of each other without having to merge their surfaces. The point clouds of both Kinects are fused by projecting them into the same coordinate space. Therefore each point is multiplied with the inverse of the extrinsic camera matrix.

The merged point cloud is then rendered from a virtual camera located at the observer's point of view. The viewpoint of the observer is computed for each frame by tracking the user's head with the third Kinect sensor that is mounted at the back side of the local user's room. The final output is rendered into a transparent (75%) WPF window that does not capture any input events from the mouse, the keyboard or the digital Anoto pens. Thus, the virtual 3D embodiment of the remote user can be superimposed on top of any content and is independent of the underlying application.

EVALUATION

Two empirical studies were conducted to explore the benefits and limitations of our proposed *3D-Board* technique. Both experiments were carried out with the same techniques, participants, and apparatus.

Apparatus

The studies were conducted in a quiet room with two interactive whiteboards, as illustrated in the implementation section (cf. Figure 2). Each of the whiteboards was dedicated for one of the two participants. The workplaces were separated from each other by a metal frame with blinds for visual cover. Thus, participants were unable to see the other whiteboard, but could communicate via voice commands and the technique in use.

Participants

In total, 12 paid volunteers (6 female, 6 male) from the local university participated. Their age ranged from 20 to 29 years ($M = 24.08$, $SD = 2.53$). Participants performed both studies in pairs of two. One was acting as an *instructor*, giving commands to the *operator*. The *operator* on the other hand, had to interpret the given instructions and acted accordingly. An important aspect of both experiments was to investigate the impact of a certain technique on the users' perception of their personal space. Thus, to increase the effects on the physical and social distance between users, we made sure that the paired up participants had never met before.

While 75% of the participants reported not being familiar with interactive whiteboards, 6 out of 12 had more experience with pen-based interfaces. Finally, 67% of the participants were using voice-over-IP application on a daily/weekly basis.

Techniques

Three different techniques were tested to compare the performance as well as the accuracy of our proposed technique:

- *3D-Board*: When using the *3D-Board* technique a front-facing 3D embodiment of the remote user was displayed, as described in the implementation section.
- *2D-Back-Facing*: This technique was chosen as a comparison with the recent related work of *CollaBoard* [20]. We maintain the same physical setup, with the camera placed behind the *instructor* (cf. Figure 2) to highlight the limitations of a 2D back-facing approach. The 2D image of the back of the user is captured and projected onto the remote whiteboard. Therefore, the body of the user can occlude the arms, forcing the users to find a good posture with their arms visible to the camera.
- *Co-located*: In the baseline condition, both participants were collaboratively working in front of the same whiteboard, thus simulating a co-located, face-to-face situation. This reflected best the everyday usage of a situated interactive display as used in both experiments.

EXPERIMENT 1: ABSTRACTED ENVIRONMENT

The first experiment investigated the performance of the techniques in an abstracted pointing task.

Task

Our task environment, based on the experiment of Grossman et al. [11], consisted of a grid of targets with the highlighted target objects being candidates for selection (cf. Figure 5).

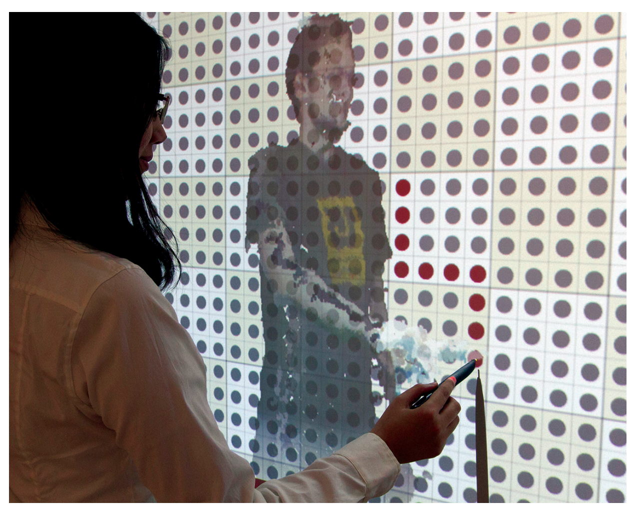


Figure 5: In the first experiment, the *instructor* pointed at target structures of different complexity levels. The *operator* had to mark the target points accordingly.

At the beginning of each trial, all targets were inactive (black). Next, a set of targets turned red, only visible for the *instructor*. Now the *instructor* had to point to all highlighted (red) targets as quickly as possible using finger/hand gestures. Following these instructions, the *operator* on the other side had to find out which of the targets to select. During the first experiment, the participants were not allowed to talk or correct each other. Targets were marked by tapping the corresponding objects once with the pen. The set of targets varied in terms of complexity. Two different levels of complexity (*Simple* and *Complex*) for the target shapes were tested. Simple targets were single line-chains, while complex objects consisted of three chains connected over two corners. The diameter of the target at the interactive whiteboard under the simple condition was $3.5cm$ and $2cm$ under the complex condition. Both participants were told to complete each trial as fast as possible. Trial times were measured between pressing the start button (on the *instructor* side) and the end button (on the *operator* side).

Design

A repeated measures within-subject design was used. Technique *3D-Board*, *2D-Back-Facing*, *Co-Located* and Complexity (simple, complex) were used as independent variables. The presentation order for the techniques was counterbalanced. For each turn, a total of three trials had to be completed. Summarizing, each group completed a total of 36 trials (3 techniques $\times$ 2 complexities $\times$ 3 trials $\times$ 2 users). In addition, qualitative feedback was collected after each technique using user experience questionnaires. Participants had to rate each technique on a 5-point Likert scale using the criteria: easy-to-use, accuracy and pointing awareness. Task load ratings were collected with NASA TLX questionnaires. After completing one full cycle of trials, participants were also asked to switch from the *instructor* to the *operator* role and vice versa. The whole test, including training sessions and questionnaires, lasted for approximately 40 minutes for each group.

Hypotheses

The experiment was designed and conducted with the following hypotheses in mind:

- *Hypothesis 1*: *3D-Board* is at least equally fast as the Co-Located technique.
- *Hypothesis 2*: *2D-Back-Facing* is the slowest technique.
- *Hypothesis 3*: *3D-Board* is more accurate than *2D-Back-Facing*.

Quantitative Results

Error rates and trial completion times were analyzed using a repeated measures ANOVA ($\alpha = 0.05$) separately for each level of complexity. The Greenhouse-Geisser correction was used if the assumption of sphericity was violated. Post-hoc analyses on the main effects were conducted in order to confirm/reject the formulated hypotheses. These consisted of paired-samples t-tests with family wise error rate controlled across the test using Holms sequential Bonferroni approach. For all bar charts, the error bars indicate the range of two standard errors of the mean (above and below the mean).

Trial Completion Time & Hit rate A repeated measures analysis of variance showed main effects for the *Technique* ($F_{2,10} = 20.057, p < 0.0001$) as well as for the *Complexity* ($F_{1,11} = 82.227, p < 0.0001$). Figure 6 depicts the overall mean time for each technique. Post-hoc analyses showed that the *3D-Board* technique was significantly faster ($M = 13.5s, SD = 3.3s$) than *2D-Back-Facing* ($M = 20.98s, SD = 6.97s$) with $p < 0.001$ for the complex task. One reason was that *operators* could instantly recognize the pointing gesture since *instructors* were not occluding the projected target with their bodies. On the other side, participants performed slightly faster in the *Co-Located* scenario than in the *3D-Board* setup. A pairwise comparison, however, showed no statistical significance ($p = .596$). This confirms *Hypothesis 1* since *3D-Board* can be considered as fast as the *Co-Located* technique. With *2D-Back-Facing* being the slowest technique, *Hypothesis 2* is confirmed as well. The same results were found for the simple task (cf. Figure 6).

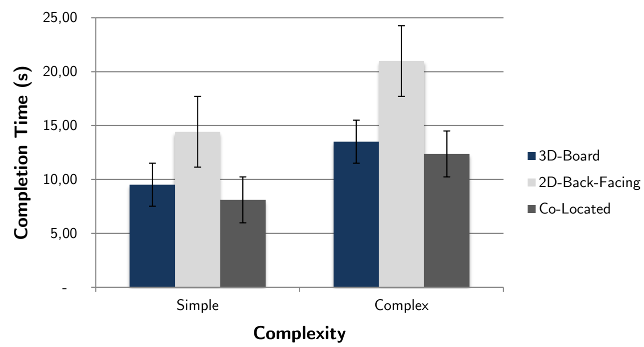


Figure 6: Overall completion time by Techniques and Complexity (Simple vs. Complex).

In addition to the trial times the number of errors were counted. On average 7.44 points were candidates for selection per trial for the complex target structures. 7.22 ($SD = 1.08$) targets have been selected correctly using the *3D-Board* technique compared to 7.25 ($SD = 1.01$) targets using the *Co-Located* technique, and 7.06 ($SD = 1.47$) using the *2D-Back-Facing* technique respectively. Given such small error rates, no significant difference could be found for selecting the correct targets. Thus, *Hypothesis 3* needs to be rejected since all three techniques appear to be very accurate.

Qualitative Results

After each technique block, participants were asked to rate the different techniques. The rating was based on a 5-point Likert scale (1 = strongly agree; 5 = strongly disagree).

Overall, 91.67% of the participants found the *3D-Front-Facing* technique either very easy or easy to use. In contrast, 75.0% of the participants rated the *2D-Back-Facing* to be difficult or very difficult to be used. Most of the participants found *3D-Board* to feel as natural as the *Co-Located* setup. The *instructor* could directly point onto the whiteboard and the *operator* would immediately understand. When using the *2D-Back-Facing* technique, we noticed that the *instructors* often tended to move sidewise with their body to avoid occluding their arms with the torso. 8 out of 12 participants also had the impression to know exactly where their partner was pointing under the *3D-Board* condition. This was in contrast to the *2D-Back-Facing* technique, where 11 out of 12 disagreed to know where their partner was pointing to.

Figure 7 depicts the task load ratings (1 = very low; 5 = very high) for all techniques, focusing on Physical Demand, Performance, Effort, and Frustration, where significant differences have been found. No significant main effects could be found for both Mental and Temporal Demand.

Interviews and Observations

According to the participants, the main advantages of *3D-Board* was the direct connection between the digital content and the superimposed 3D embodiment. This created the im-

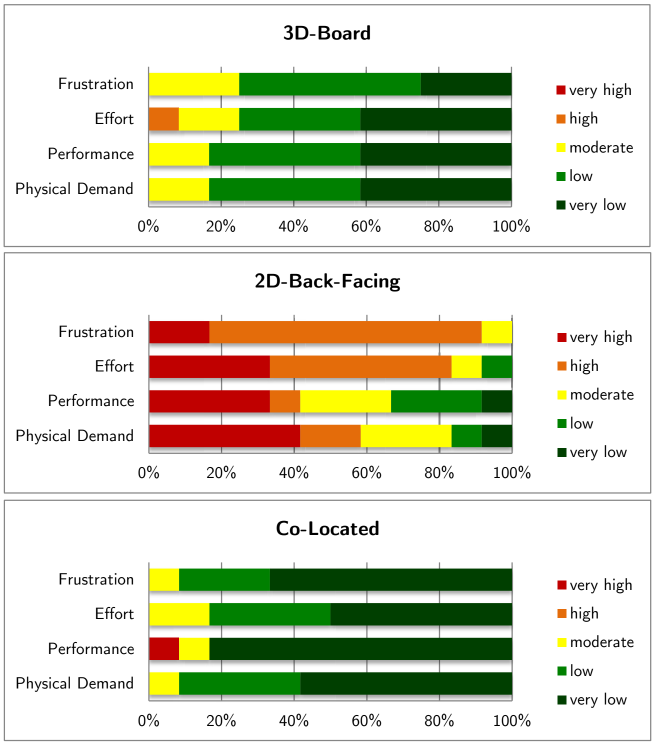


Figure 7: Task load ratings of the first experiment with significant differences for *3D-Board*, *2D-Back-Facing*, and *Co-Located*.

pression of working face-to-face with the remote user only separated by a glass window.

> *While using the 3D-Board technique, it was easy to see where the* instructor *was pointing to, because it gave me the impression that my partner was behind the wall. The interaction with someone who is behind the wall is exciting. –Participant 2A*

A lot of participants stated the *2D-Back-Facing* technique *needs a lot of practice*. One participant stated that acting as an *instructor* was a lot more mentally demanding and they had to hold the fingers in an ideal position to point to the target correctly.

The most convincing argument for the *3D-Board* was to work face-to-face with the other user.

> *I liked the feeling of communication in the 3D-Board condition, because the other person was looking at me, which made me feel more in contact with him. –Participant 5B*

Finally, they also appreciated it to see the other person in full size.

> *It was strange and at the same time felt good to see the other partner in full size. –Participant 6A*

We noticed increased interaction between users under the *3D-Board* condition. In addition, users tended to use more gestures and acted as if the remote participant was physically present.

EXPERIMENT 2: INTERIOR DESIGN

While in the first experiment we were mainly focusing on pointing gestures, in the second experiment, we investigated the performance of the *3D-Board* technique in a realistic scenario to test the natural interaction between users.

Task

In the second task, the *instructor* was presented a floor plan and adequate furniture, which had to be selected, scaled, rotated, and translated accordingly by the *operator* based on the instructions provided. In addition to rearranging the furniture, the *operator* had to draw three inner walls by sketching them on the surface (cf. Figure 8). Thus, the *instructor* had to use a big variety of gestures in order to give the *operator* precise guidelines. The *instructor* was also allowed to use voice commands, creating a realistic scenario of collaboration. The task was finished once the *instructor* was satisfied with the floor-plan.

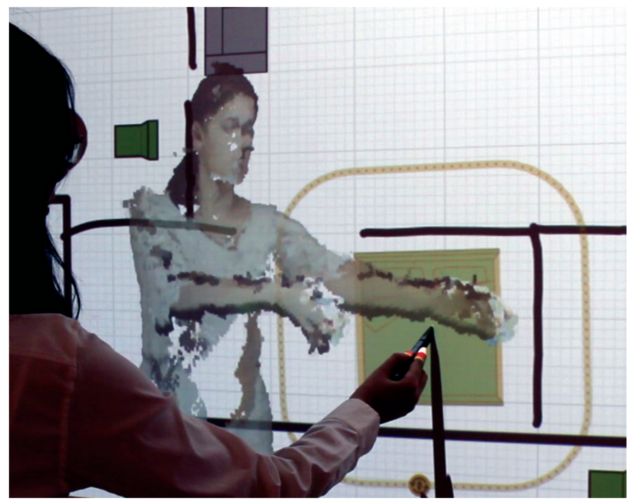


Figure 8: In the second experiment both participants had to work simultaneously while drawing and rearranging furniture on a floor plan.

Design

A within-subject design was used. The presentation order for the three techniques was counterbalanced. Per technique one unique floor plan had to be completed. The participants could familiarize themselves with the application before the trials began. After completing the trials for all techniques the *instructor* and *operator* switched roles. It took one group about 20 minutes to complete all trials. User satisfaction, ease-of-use, and engagement were gathered through interviews and questionnaires. Task load ratings were collected using the NASA TLX questionnaires.

Hypotheses

With the user experience in mind these additional hypotheses were formulated for the second experiment:

- *Hypothesis 4*: *Co-Located* is the preferred technique.
- *Hypothesis 5*: *3D-Board* raises workspace awareness.
- *Hypothesis 6*: The *3D-Board* embodiment does not violate the collaborator's personal space.

Qualitative Results

Qualitative user feedback showed that participants were highly positive about the concept of *3D-Board*. Half of all participants preferred *3D-Board* over all other techniques, and no one of them mentioned *3D-Board* as their least favorite. Thus, *Hypothesis 4* has to be rejected since *3D-Board* and *Co-Located* were equally popular.

On a five-point Likert scale (1 = very easy/strongly agree; 5 = very difficult/strongly disagree), participants found *3D-Board* easy to use ($ME = 1.5$) and easy to learn ($ME = 1.0$). In contrast, the *2D-Back-Facing* technique was rated as being moderate to use ($ME = 3.0$) and moderate to learn ($ME = 2.5$). In addition to that, 66.67% of the participants also felt a high or very high personal engagement when using *3D-Board*. The *Co-Located* scenario achieved better results with 83.33% while only 58.33% felt high or very high engagement when using *2D-Back-Facing*. A Wilcoxon signed-rank test showed significant differences for both usability ($z = 2.375, p = 0.018$) and learnability ($z = 2.326, p = 0.020$), for the comparison of *3D-Board* and *2D-Back-Facing*.

Figure 9: Task load ratings of the second experiment with significant differences for *3D-Board*, *2D-Back-Facing*, and *Co-Located*.

Figure 9 depicts the task load ratings for all techniques, focusing on Physical Demand, Performance, Effort, and Frustration, where significant differences have been found. As in the first experiment, no significant main effects could be found for both Mental and Temporal Demand. The post-condition questionnaire data (1 = very low; 5 = very high) were analyzed using two related samples Wilcoxon signed-rank tests. Overall, participants found *3D-Board* ($ME = 2.0$) less physically challenging than the 2D technique ($ME = 3.5$), $z = -2.414, p = 0.016$. In addition, most of the participants also felt the performance was better when using *3D-Board* ($ME = 4.5$) than with *2D-Back-Facing* ($ME = 4.0$), $z = -2.449, p = 0.014$. Also the effort was rated lower for *3D-Board* ($ME = 2.0$) than for *2D-Back-Facing* ($ME = 3.0$), with $z = -2.972, p = 0.003$. Finally, participants felt less frustration using the 3D technique ($ME = 2.0$) than using the 2D technique ($ME = 3.0$), $z = -2.739, p = 0.006$.

Awareness Once asked about how easy it was to make oneself well understood as an *instructor*, the majority of the participants voted in favor of the front-facing *3D-Board* technique ($ME = 1.0$) compared to the back-facing technique ($ME = 2.0$), $z = 2.714, p = 0.007$. When acting as the *operator*, the majority of the participants knew better what the *instructor* was doing when using *3D-Board* ($ME = 1.5$) compared to *2D-Back-Facing* ($ME = 2.0$), $z = 2.810, p = 0.005$. However, the *Co-Located* technique outperformed *3D-Board* ($ME = 1.0$) in terms of awareness with $z = -2.236, p = 0.025$. Thus, although *3D-Board* can raise the workspace awareness better than *2D-Back-Facing*, *Hypothesis 5* needs to be rejected, since *Co-Located* is still the preferred technique in this case.

Territoriality Similar to [6], we asked the participants to rate their agreement with the statements, "I often had the feeling that my personal space was invaded by my partner", and "I often had the feeling of invading my partner's personal space". While approximately 75% of the participants disagreed or strongly disagreed with the first statement of their personal space being invaded by the partner under the *3D-Board* condition, only 58.33% of all participants disagreed under the *Co-Located* condition. Similar results were observed with the second question. Again, 75% of the participants disagreed in the *3D-Board* setup, while 66.67% of all participants disagreed in the *Co-Located* setup. A Wilcoxon signed-rank test showed no main effects. In our experiment, participants had complete freedom to choose where to stay and how to interact with their remote partner. We noticed that the *operator* and *instructor* often stood face-to-face and made use of working simultaneously on the same spot of the whiteboard when using *3D-Board*.

Interviews and Observations
At the end of the study, participants were asked about the advantages *3D-Board* has compared to the *Co-Located* condition and what they liked best about it.

Two participants found that the technique provides more space, which would also *allow more people to work on the whiteboard simultaneously*. In addition to that, a user will automatically have more private space. It was interesting to see that even though the remote collaborator was directly in front of a person, the *instructor* was not noticed as a disturbance as it would be the case in a *Co-Located* setup, where people would step on each other's toes.

> *The collaboration was easier once you see the remote person in front of you - it gives you the feeling to work really face-to-face. And this, without treading on someone's foot –Participant 5A*

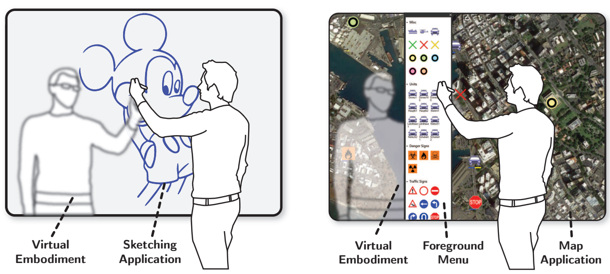
Participants appreciated the front-facing approach a lot, which provided a better gestural interaction, while working on the same content.

> *The facial expression was highly relevant for the interaction and although the remote participant was superimposed, I could clearly see the* instructor's *gestures. –Participant 4B*

On the other side, three participants found no advantages using the *2D-Back-Facing* technique. One participant found that the *2D-Back-Facing* technique was *exhausting* and lacked facial expression. Similarly, another participant also missed the gestural interaction. Overall, most of the participants found it more difficult to interact without seeing the remote worker face-to-face. Summarizing, these observations confirm *Hypothesis 6* since users showed a tendency of acting more confident when using *3D-Board*.

APPLICATION SCENARIOS

We have tested *3D-Board* in two scenarios (cf. Figure 10). In the first scenario, we superimpose the entire video image on top of a full-screen application. In the second scenario, important content, such as on-screen menus, are rendered as the topmost element to avoid occlusion and keep them visible at all time.



(a) (b)

Figure 10: While in the first demo application the entire data screen is occluded by the video image (a), specific content (e.g. menus) is not superimposed by the video image in the second demo application (b).

Remote sketching application

For the first demonstrator, we implemented a simple sketching tool. As depicted in Figure 11, the superimposed remote collaborator is occluding parts of the underlying application. However, because of the size of the surface, the transparency of the virtual embodiment and the nature of the content the users never felt that the virtual embodiments was distracting from the actual content.

Figure 11: In the remote sketching application, the entire working space is occluded by the transparent (75%) video image.

Remote Map Surveillance

The main idea for 3D Board came from a close collaboration with the local police. The police needed an intuitive and easy-to-use remote Common Operational Picture (COP) based on an interactive whiteboards. Larger operations often require additional support from other groups situated in geographically distributed control rooms. To provide an effective operational picture, it is important that all data is merged in a single view as quickly as possible without overwhelming the officers.

Figure 12: In the Common Operational Picture application, police vehicles had to be placed accordingly. Important windows (e.g. dialogs) are always on top and not occluded by the video image.

In our scenario, police officers utilize an interactive wall to monitor a large city map. The map is augmented with GPS-tracked police vehicles that had to be coordinated by multiple co-located and remotely located police officers.

While the main COP environment is again superimposed by the remote collaborator (similar to the sketching application), it was important avoid occlusion for important items. Thus, menus for icon selection are rendered as the topmost elements to avoid occlusion.

CONCLUSION & FUTURE WORK

In this paper we have presented *3D-Board*, a digital whiteboard featuring life-sized 3D virtual embodiments of geographically distributed users. By blending the front-facing 3D embodiment of a remote collaborator with the shared workspace, an illusion is created as if the observer is looking through the transparent whiteboard into the remote user's room. Our approach is characterized by a simple setup and front-facing rendering technique, providing an easy-to-understand and easy-to-learn face-to-face experience for remote collaboration.

The system was tested and verified in a usability assessment, showing that *3D-Board* significantly improves the effectiveness of remote collaboration on a large interactive surface (especially compared with a *2D-Back-Facing* setup). We also noticed that participants explicitly appreciated the face-to-face communication. It was interesting to see participants working in the same area of the shared workspace simultaneously, since they experienced the remote collaborator as standing behind the screen. During the study we could also observe that the participants were more involved and a lot more activity happened. In contrast, participants would keep a certain distance to each other in the *Co-Located* setup.

There are several areas that we would like to improve. The current image quality is inferior and should be further enhanced. Although the study results were highly promising and participants did not mention the rendering to be have negative effects for our approach, the noisy look has to be improved with more advanced depth data processing techniques. Moreover, we would like to improve the frame rate and implement more rendering stages on the GPU. In addition, we would like to test the setup using newer depth sensors and try alternative setups with the overall goal to embed the sensors in the frame of the interactive whiteboard.

Acknowledgments
The research leading to these results has received funding from the Austrian Research Promotion Agency (FFG) within the research program KIRAS under grant agreement #29335581 (ERiC).

REFERENCES

1. Beck, S., Kunert, A., Kulik, A., and Froehlich, B. Immersive group-to-group telepresence. *IEEE Trans. Vis. Comput. Graph. 19*, 4 (2013), 616–625.
2. Benford, S., Bowers, J., Fahlén, L. E., Greenhalgh, C., and Snowdon, D. User embodiment in collaborative virtual environments. In *CHI '95*, ACM Press/Addison-Wesley Publishing Co., 1995, 242–249.
3. Bly, S. A., Harrison, S. R., and Irwin, S. Media spaces: Bringing people together in a video, audio and computing environment. *Commun. ACM 36*, 1 (Jan. 1993), 28–46.
4. Borgefors, G. Distance transformations in digital images. *Comput. Vision Graph. Image Process. 34*, 3 (June 1986), 344–371.
5. Butler, D. A. et al. Shake'n'sense: Reducing interference for overlapping structured light depth cameras. In *CHI'12*, ACM, 2012, 1933–1936.
6. Doucette, A., Gutwin, C., Mandryk, R. L., Nacenta, M., and Sharma, S. Sometimes when we touch: How arm embodiments change reaching and collaboration on digital tables. In *CSCW '13*, ACM, 2013, 193–202.
7. Edelmann, J., Peter, G., Mock, P., Schilling, A., and Strasser, W. Face2Face - a system for multi-touch collaboration with telepresence. In *ESPA'12*, IEEE, 2012, 159–162.
8. Fussell, S. R. et al. Gestures over video streams to support remote collaboration on physical tasks. *Hum.-Comput. Interact. 19*, 3 (Sept. 2004), 273–309.
9. Genest, A., and Gutwin, C. Characterizing deixis over surfaces to improve remote embodiments. In *ECSCW'11*, Springer, 2011, 253–272.
10. Genest, A. M., Gutwin, C., Tang, A., Kalyn, M., and Ivkovic, Z. Kinectarms: A toolkit for capturing and displaying arm embodiments in distributed tabletop groupware. In *CSCW '13*, ACM, 2013, 157–166.
11. Grossman, T., Baudisch, P., and Hinckley, K. Handle flags: Efficient and flexible selections for inking applications. In *GI'09*, 2009, 167–174.
12. Gumienny, R., Gericke, L., Quasthoff, M., Willems, C., and Meinel, C. Tele-Board: Enabling efficient collaboration in digital design spaces. In *CSCWD'11*, IEEE, 2011, 47–54.
13. Haller, M. et al. The nice discussion room: Integrating paper and digital media to support co-located group meetings. In *CHI '10*, ACM, 2010, 609–618.
14. Hauber, J., Regenbrecht, H., Billinghurst, M., and Cockburn, A. Spatiality in videoconferencing: Trade-offs between efficiency and social presence. In *CSCW'06*, ACM, 2006, 413–422.
15. Ishii, H., and Kobayashi, M. Clearboard: A seamless medium for shared drawing and conversation with eye contact. In *CHI'92*, ACM, 1992, 525–532.
16. Izadi, S. et al. C-slate: A multi-touch and object recognition system for remote collaboration using horizontal surfaces. In *Tabletop*, IEEE, 2007, 3–10.
17. Jota, R., Nacenta, M. A., Jorge, J. A., Carpendale, S., and Greenberg, S. A comparison of ray pointing techniques for very large displays. In *GI'10*, Canadian Information Processing Society, 2010, 269–276.
18. Junuzovic, S., Inkpen, K., Blank, T., and Gupta, A. Illumishare: Sharing any surface. In *CHI '12*, ACM, 2012, 1919–1928.
19. Kim, K., Bolton, J., Girouard, A., Cooperstock, J., and Vertegaal, R. Telehuman: Effects of 3d perspective on gaze and pose estimation with a life-size cylindrical telepresence pod. In *CHI'12*, ACM, 2012, 2531–2540.
20. Kunz, A., Nescher, T., and Küchler, M. Collaboard: A novel interactive electronic whiteboard for remote collaboration with people on content. In *CW'10*, IEEE, 2010, 430–437.
21. Ledo, D., Aseniero, B. A., Greenberg, S., Boring, S., and Tang, A. Onespace: Shared depth-corrected video interaction. In *CHI '13 Extended Abstracts*, ACM, 2013.
22. Maimone, A., and Fuchs, H. Encumbrance-free telepresence system with real-time 3d capture and display using commodity depth cameras. In *ISMAR*, IEEE, 2011, 137–146.
23. Pinelle, D., Nacenta, M., Gutwin, C., and Stach, T. The effects of co-present embodiments on awareness and collaboration in tabletop groupware. In *GI'08*, Canadian Information Processing Society, 2008, 1–8.
24. Tan, K.-H., Robinson, I. N., Culbertson, W. B., and Apostolopoulos, J. G. Connectboard: Enabling genuine eye contact and accurate gaze in remote collaboration. *IEEE Trans. on Multimedia 13*, 3 (2011), 466–473.
25. Tang, J. C., and Minneman, S. Videowhiteboard: Video shadows to support remote collaboration. In *CHI '91*, ACM, 1991, 315–322.
26. Tuddenham, P., and Robinson, P. Remote review meetings on a tabletop interface. *CSCW'06* (2006).
27. Tuddenham, P., and Robinson, P. Territorial coordination and workspace awareness in remote tabletop collaboration. In *CHI '09*, ACM, 2009, 2139–2148.
28. van Eijk, R., Kuijsters, A., Dijkstra, K., and IJsselsteijn, W. Human sensitivity to eye contact in 2D and 3D videoconferencing. In *QoMEX'10*, IEEE, 2010 2010, 76 – 81.
29. Zhang, Z. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell. 22*, 11 (2000), 1330–1334.