OneSpace: Shared Depth-Corrected Video Interaction

David Ledo, Bon Adriel Aseniero, Sebastian Boring, Anthony Tang
University of Calgary
2500 University Dr NW, Calgary, AB
davidledo89@gmail.com, bon_adriel@hotmail.com, {sebastian.boring | tonyt}@ucalgary.ca

ABSTRACT
Video conferencing commonly employs a video portal metaphor to connect individuals from remote spaces. In this work, we explore an alternate metaphor, a *shared depth-mirror*, where video images of two spaces are merged into a single shared, depth-corrected video. Just as seeing one's mirror image causes reflective interaction, the shared video space changes the nature of interaction in the video space. We realize this metaphor in OneSpace, where the space respects virtual spatial relationships between people and objects, and in so doing, encourages cross-site, full-body interactions. We report preliminary observations of OneSpace in use, describing the role of depth in our participants' interactions. Based on these observations, we argue that the depth mirror offers new opportunities for shared video interaction.

Keywords
Video communication, media spaces.

INTRODUCTION
Enabling synchronous interaction between people separated by physical distance has long been a principal concern for CSCW research. The core vision underlying considerable work in this space is to support interaction with remote people as if they were co-present. Attempts to realize this vision have relied primarily on video transmission. Current approaches have largely been shaped by the demands of particular usage contexts. Due to the expense of technology, for example, initial approaches were focused on the communication demands of business. To support "face-to-face" conversation and meetings, the most common approach has been to provide an audio-video window into a remote space (i.e. video conferencing). Many early media spaces, born out of an interest in collaborative group work, also focused on providing a shared workspace. These fused workspaces (frequently through video cameras pointed at that workspace) provided people with a shared space to conduct joint activity (e.g. drawing or design tasks) while being able to gesture fluidly (e.g. [4,10,11]).

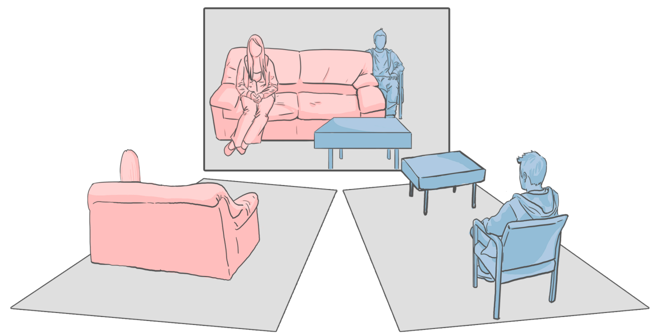

Figure 1. OneSpace integrates two remote spaces (bottom right and left) into a single space (top) by presenting a combined depth mirror of both spaces.

Our interest here is in revisiting the question: how can we design technology that allows people to interact as if they were co-present—allowing them to share a space together, and to feel connected through their shared interactions in that space? As illustrated in Figure 1, the metaphor we explore in this work is a *shared depth-mirror*: people see themselves and interact with others in a shared video scene that looks like a mirror, with objects and people from remote sites are overlaid with correct depth cues. We were inspired by the video-based interfaces introduced by Krueger [6], and more recently popularized by video game systems, where people interact with a mirrored video image of themselves. This approach creates a virtual "stage" for interaction, and as we will see, fundamentally changes how people interact with one another.

OneSpace's approach to integrating remote spaces draws attention away from the problem of supporting conversation or other specific tasks, and instead brings focus to the kinds of interactions we can enable through video spaces. Our preliminary observations show that the depth-corrected feed allows for a broad range of rich, playful interactions that go beyond a traditional chroma-keyed implementation [8]. Specifically, the depth cues provide people with a shared, negotiated *stage* for their shared interactions, where the negotiation occurs merely through one's closeness with the video scene (just like in a mirror—only one person can be "in front" at once).

This paper outlines the design of OneSpace, and the technical approach that we used to realize our design goals. We present the range of interactive possibilities supported by OneSpace, and describe preliminary experiences and reactions to the system. The main contribution of this work is that it rethinks how we can connect remote spaces to support interaction, and concretely illustrates how a depth-corrected mirror metaphor does so.

RELATED WORK

Researchers have long used video as a means to allow people to interact with one another as if they were in a collocated space. Yet, researchers' orientation to this vision, and even how to employ video has varied. We briefly outline this work, discussing how video has been used as a portal between spaces, and as a means to bring interaction into a shared space.

Conversation through a portal. The video conferencing model employs a camera paired with a video of the remote feed, is the prevailing approach in most work. Here, video acts as a portal or tunnel that connects remote spaces, primarily for conversation. Beyond video conferencing, we have seen this applied to a wide range of contexts, such as design work [10], informal awareness [2], and play [9].

Shared workspaces for tasks. Rather than focusing specifically on conversation, video has also been used to fuse two separate workspaces into a single shared workspace for task work. These generally project a video feed from the remote workspace onto the local space (e.g. [4,5,10,11]). The result is a single "workspace" that allows people to interact through shared artifacts (or drawings). Importantly, the use of video allows people to interact with one another through gesture, rather than solely through conversation. MirrorFugue [15] explores this interaction within a musical context, where the focus is on the placement and movement of fingers over a shared/mirrored piano keyboard.

Full-body communicative interaction. Krueger's original Videoplace work realized a vision to connect remote spaces through full-body silhouettes that were simultaneously projected onto a large wall-sized display [6]. These silhouettes supported full-body communicative gestures, and interaction through computational elements. This virtual third space was realized through video in HyperMirror [8]. Here, video capture of remote spaces are fused through chroma-keying effects, with the resulting fused image (akin to a mirror) projected onto a large display. This mirror metaphor allows for self-reflection, and accordingly, a more relaxed conversational environment. Hill et al. [3] also explored this metaphor, but used virtual embodiments rather than video.

Our work builds on the ideas introduced in HyperMirror [8]. As in HyperMirror, rather than focusing on particular tasks, or conversation as in prior work, our emphasis was on integrating remote spaces to support people's interactions: to allow them to interact as if they were co-present. We revisit the mirror metaphor, and go beyond HyperMirror by adding depth information to the video feed. As we will see, this substantially changes the space of possible interactions.

ONESPACE

OneSpace integrates remote spaces through a shared depth-mirror metaphor. From this vision, we outline several goals that guided our design process:

Engagement through visualization of one's own video embodiment. Seeing one's embodiment in video (as in a mirror) encourages self-reflection, and, as observed by others, we reorganize and change our behaviours when we are made aware of them.

Provide opportunities for interpersonal interaction rather than focusing on specific tasks. Rather than emphasizing particular tasks, or discussion, we wanted to design the system so people would feel as if they were interacting in the same space, be it a real or virtual one.

Support full-body interaction. We wanted to go beyond the desktop video conferencing model (where the focus is face-to-face conversation) and instead afford people the ability to make use of the physical space around them to better support full-body communicative interaction.

Respect spatial relationships. A crucial aspect of OneSpace's design was to respect the location, distance, and orientation between people and objects in the shared space. In addition to supporting the left-right relationships as illustrated in HyperMirror [8], we also sought to support front-back relationships. As we will see, these relationships become a key factor in physical play activities, or showing objects to others.

OneSpace integrates multiple physically separated spaces through the shared depth-mirror metaphor. It allows for any number of real and virtual spaces (we have tested it with four environments) while respecting the spatial relationships of people and objects in the space: things and people who are closer to the mirror appear in front of those who are further away. People are able to interact through the manipulation of physical objects in the space, and through body movement and motion in the space.

IMPLEMENTATION

We implemented OneSpace as a distributed application using a client/server architecture. As illustrated in Figure 2, thin clients send RGB and depth data collected from connected Microsoft Kinects. The server then processes and merges this data before sending it back to the clients to be displayed. In our current setup, we use large, whiteboard-sized displays to show the resulting video. OneSpace is implemented in C# WPF with the Microsoft Kinect SDK.

OneSpace's server integrates the color video frames it receives from clients (Figure 2). On a per-pixel basis, it uses the depth information to extract the front-most color pixels to create a new video frame which is then sent back to all the clients for display. This process provides people with a mirrored image of themselves, and preserves the spatial relationships of every person and object in each space, allowing for occlusions and overlaps to occur in the final video frame. We apply standard image processing techniques to smooth the depth information, to help the resulting image appear smoother and more seamless.

Krueger's VIDEOPLACE provided a number of video effects on people's video embodiment [6] that allowed people to engage in expressive, video-based "embodied" interaction. Inspired by the opportunities for interpersonal interaction enabled by these video filters, we also designed a number of effects for OneSpace, as illustrated in Figure 3:

- *Background Effects*. OneSpace can use three different kinds of scenes as the "background" for the interaction: (a) it can use the scene from one of the sites; (b) it can use a static background from a still image, or (c) it can employ a pre-recorded 3D scene (with both colour and depth information). This allows people to virtually share a background—real or virtual.
- *Shadows and traces*. As with Krueger's original implementation, we can also draw foreground objects as silhouettes, allowing people to interact as shadow puppets rather than as video embodiments. We can also apply a trace effect, where ghostly trails of people's motions are overlaid atop one another. These effects encourage unique forms of interaction where people's bodies can be merged into one.
- *Recorded scenes*. As in Looking Glass [1], we can also record a 3D video of a scene, and then embed live users into the playback of that scene. This allows people to relive previous scenes and interactions with one another—potentially those that involve others altogether.

PRELIMINARY OBSERVATIONS OF USE

We made OneSpace available to several members of our institution to understand the kinds of interactions OneSpace afforded. For these tests, we connected two remote spaces through a Gigabit Ethernet connection. Each site had its own large display (whiteboard-sized) and Kinect camera, and the two spaces were connected through a separate audio link. Typically, these tests involved groups of four people—two people per site. Beyond describing the basic technical features of the system, we did not guide their interactions, and simply allowed them to use the system.

Virtual physical and visual play. While we expected that people would still use the system for conversation, we were surprised to see very little conversation at all (although there was a lot of laughter). Instead, interaction focused on the shared scene being displayed on-screen, with participants focused on how their video embodiment (i.e. their reflection) interacted with/shared the scene with video embodiments of people from the remote site on the shared "stage." Where speech did occur, it was to coordinate or guide these interactions.

These scenes were striking, as we saw our participants engage spatially with one another in ways that they would not if they were actually physically co-present. That is, they allowed their visual embodiments to interact and virtually "touch" one another in ways that would be unusual or uncomfortable in real life. For instance, a common interaction (perhaps a statement about our society) was to enact mock fist-fights with participants from the remote site. These fist fights made use of the depth-cues—for example, a punch might begin from "behind" a user, and follow through into the foreground. Here, the target would feign being hit in that direction. Perhaps as a response to these fist-fights, our participants also hugged one another, as the system would create the visual effect of these interactions in the mirror without actual physical contact. Notably, none of these participants had gotten into fist-fights or hugged one another in real life before.

Staging visual interaction. Participants also carefully staged the visual interaction with one another. In many of the fist-fights, people who were "not involved", would move out of the scene. In other cases, we observed several participants playing "headless horseman" with one another. Here, two people would stand "atop" one another in the scene, with one person leaning his head back, while the other would lean his head forward. The resulting scene would produce a humourous combination "person" with the body of one person, and the head of another. Here, the depth cues allow for interactions that would not be otherwise possible with a chroma-key solution.

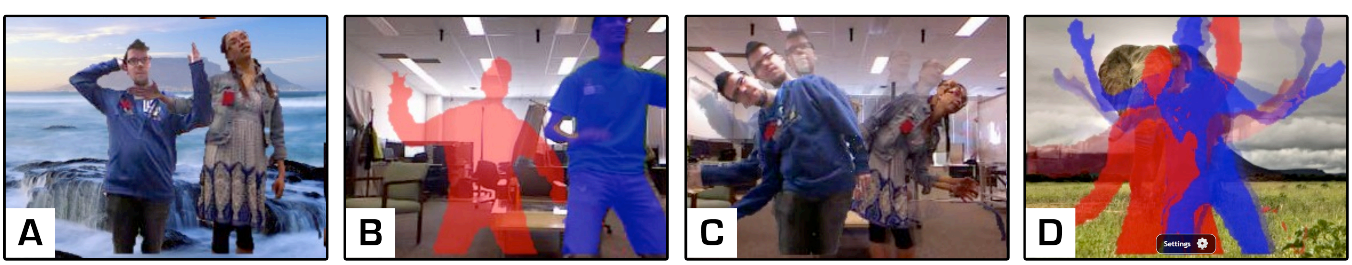

Figure 3. In OneSpace, one can see different visual effects: (a) shows a static background; (b) shows silhouettes; (c) shows traces, and (d) shows a combination of several effects (here: silhouettes, traces, and a static background).

We see here then that people are negotiating the use of the "stage" in two ways: in the first, people who are not involved move out of the way, while in the second, correcting the shared scene for depth allows people to alternate who takes "the stage." This stage is a flexibly negotiated space, since it merely means moving closer to the camera. Yet, it is not binary, as it would be in a chroma-keyed approach: as we saw in the "headless horseman" example, this stage is a blended area, where people can choose what "part" of their body is in front. The feedback provided by seeing one's own embodiment enables this active negotiation.

Engagement and enjoyment. Participants clearly enjoyed using our system. Much as in Social Comics [7], participants took pleasure in making one another laugh through the shared visual scene, and to create scenes that would be absurd, unusual or even impossible to enact in real life. The size of our display and capture area allowed for full-body interaction, and the shared depth-mirror metaphor allowed our participants to exploit spatial relationships. We saw them engaging in play, and immersing themselves in the activities that they created. For these reasons, we believe our system to be particularly useful for play environments and also useful to bring people together to have fun.

CONCLUSIONS AND FUTURE WORK

Standard video conferencing will likely remain the dominant form of interaction across remote spaces. From our explorations, however, we have seen that OneSpace's shared depth mirror metaphor blends spaces in a way that is fundamentally different from the video portal approach (e.g. [4,10,11]). In particular, the "stage" of interaction is not only shared, but because it is based only on depth cues, it becomes a space negotiated by one's proximity to the camera. As a consequence, people interact through the system in a qualitatively different manner from prior systems (e.g. [3,8])—rather than being controlled by an unseen force (as would be with chroma-key approaches), people control these features, and use it in their interactions with one another. We see these types of playful interactions as an opportunity for further exploration.

For instance, Yarosh et al. [13] indicate that to support play between children, blending the representations of remote children with the play space—something that OneSpace does—would be a useful avenue to explore. OneSpace affords differently sized toys as well: whereas VideoPlaydate employed smaller toys (e.g. toy cars), and focused on the "play space", we would expect that OneSpace could be used with larger, physical toys (e.g. baseball bats). To support this kind of exploration, we expect to make a number of technical improvements, for instance, increasing the frame rate of the system, or to provide shared virtual toys that children could play with in the mirror. Further, this kind of video space can provide a means to support physiotherapy, where the depth cues can aid teaching certain movements and poses.

In this paper, we introduced OneSpace, a system that performs depth-corrected integration of multiple spaces. The system supports a number of variations on the visual output, including static and 3D scenes, as well as silhouette and trace effects. Our preliminary observations of the system in use suggest that people understand and appropriate the depth-mirror metaphor for physical and visual play. We have seen that this metaphor encourages forms of shared interactions that go beyond current efforts in video conferencing, and presents a unique set of opportunities for shared video interaction across remote spaces.

REFERENCES

1. Adriel, B., and Sharlin, E. (2011). The looking glass: visually projecting yourself to the past. In *Proc. ICEC 2011*, 282-287.
2. Dourish, P. and Bly, S. (1992). Portholes: supporting awareness in a distributed work group. In *Proc. CHI 1992*, 541-547.
3. Hill, A., Bonner, M. N., and MacIntyre, B. (2011). ClearSpace: mixed reality virtual teamrooms. In *Proc. HCI International 2011*, 333-342.
4. Ishii, H., and Kobayashi, M. (1992). ClearBoard: a seamless medium for shared drawing and conversation with eye contact. In *Proc. CHI 1992*, 525-532.
5. Junuzovic, S., Inkpen, K., Blank, T., and Goopta, A. (2012). Illumishare: sharing any surface. In *Proc. CHI 2012*, 1919-1928.
6. Krueger, M. W. (1991) *Artificial Reality II*. Reading, MA: Addison-Wesley.
7. Lapides, P., Sharlin, E., and Sousa, M. C. (2011). Social comics: a casual authoring game. In *Proc. BCS HCI 2011*, 259-268.
8. Morikawa, O. and Maesako, T. (1998). HyperMirror: toward pleasant-to-use video mediated communication system. In *Proc. CSCW 1998*, 149-158.
9. Mueller, F., Gibbs, M. R., and Vetere, F. (2009). Design influence on social play in distributed exertion games. In *Proc. CHI 2009*, 1539-1548.
10. Tang, J.C. and Minneman, S. (1991). VideoWhiteboard: video shadows to support remote collaboration. In *Proc. CHI 1991*, 315-322.
11. Tang, J.C. and Minneman, S.L. (1991). Videodraw: a video interface for collaborative drawing. *ACM Trans. Inf. Syst. 9*(2), 170–184.
12. Xiao, X. and Ishii, H. (2011). MirrorFugue: communicating hand gesture in remote piano collaboration. In *Proc. TEI 2011*, 13–20.
13. Yarosh, S., Inkpen, K.M., and Brush, A.J. (2010). Video playdate: toward free play across distance. In *Proc. CHI 2010*, 251–1260.