Enhancing Expressiveness in Reference Space

Anthony Tang
University of British Columbia
tonyt@ece.ubc.ca

Aaron Genest
University of Saskatchewan
aaron.genest@cs.usask.ca

Garth Shoemaker
University of British Columbia
garths@cs.ubc.ca

Carl Gutwin
University of Saskatchewan
gutwin@cs.usask.ca

Sid Fels
University of British Columbia
ssfels@ece.ubc.ca

Kellogg Booth
University of British Columbia
ksbooth@cs.ubc.ca

Abstract
In the broad design space of telepresence systems, we are interested in contexts where users collaborate over a shared workspace. Our work involving connecting distributed touch surfaces (e.g. distributed tabletops) has shed light on the problem of supporting *reference space*—the ability of collaborators to point at, and refer to objects in the workspace. A promising approach to support this gestural communication has been to capture video of users' arms as they work over the surface, transmitting and overlaying that video at remote workstations. The problem with this approach is that the video image is a flat projection of users' 3D bodies, limiting the expressiveness of users' gestures, and occasionally providing false information. In our most recent work, we have begun exploring "non-photorealistic" visualizations of users' bodies to support expressiveness in reference space.

Introduction
We employ Buxton's three-space conceptual framework for telepresence to ground our research (Buxton, 2009). This framework specifies *person-space* (where verbal and facial cues are used for expression—typically realized as video and audio connection), *task-space* (where the work appears—typically realized as a shared workspace application), and *reference-space* (where remote parties use body language to refer to work). The focus of this position paper is on our investigations into supporting reference space—specifically, enabling *gestural communication* such as pointing, tracing or workspace gestures that are common to collaborative workspace activity. In designing and building collaborative workspace systems employing touch surfaces such as SMARTBoards or Microsoft Surfaces, we have adopted the design approach realized by Tang & Minneman (1990) to support reference space. This approach captures video of user's bodies as they interact atop the surface, transmitting the processed video to the remote workstation, where it is then overlaid atop the shared workspace.

Figure 1 illustrates such a collaborative system at work (Tang et al., 2010). Remote collaborators are represented with physical surrogates consisting of a screen, webcam, microphone and speaker (facilitating person-space), while the tabletops themselves are running a shared workspace (supporting task-space). Finally, an overhead camera (not visible) captures the arms of collaborators, and we overlay them into remote

Figure 1. The telepresence system presented by Tang et al. (2010) enables person-space (through the video surrogates), task-space (through the connected tabletop application), and reference-space (through the video-captured representation of B's arms).

workspaces to support reference-space. We have called this embodiment technique alternatively VideoArms (Tang et al., 2006) and Arm Shadows (Tang et al., 2010), but the underlying goal of supporting rich gestural communication remains the same.

These embodiments support both what we call "foreground" (intentional) as well as "background" (unintentional) gestural communication. Examples of foreground gestures include explicit pointing and deictic gestures (for example, when someone points at an object and says, "This one"), or pathing gestures (where a shape or outline is traced). By background gestures, we are referring to what has been called consequential communication, or the information that is "given off" by a user's interactions with the workspace. That is, when users are interacting with a workspace, we can observe and infer their actions because of how their bodies are moving in relation to objects in the workspace. Without such embodiments, distributed collaboration can appear stilted and awkward as the actions of remote collaborators become unpredictable. These rich embodiments support the workspace awareness in distributed systems that we take for granted in everyday collocated activity.

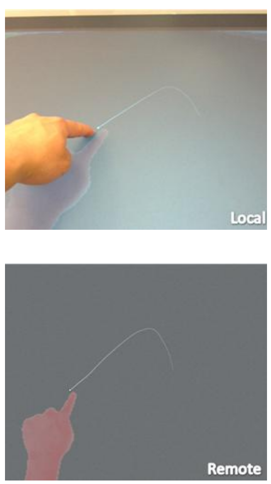

Figure 2. The video embodiment of the user is enriched by adding: colour, a contact trace, and an emphasized contact point.

Yet, the expressiveness of those bodily gestures is still limited by the particular capture technology that we are using: the camera provides only a planar, "photo-realistic" embodiment on the workspace, whereas we can already capture and potentially visualize a far richer range of information about collaborators' actions (for instance, where they are touching, or what they are trying to communicate). That is, we can accentuate characteristics about the user's activities that carry meaning, or de-emphasize characteristics of the embodiment when they carry little additional value. Our present research examines this "non-photorealistic rendering" space, exploring how we can provide additional expressiveness to embodiments in reference space. Figure 2 illustrates early iterations of this idea, where not only the video of the arm is transmitted: the arm is tinted to help aid with identity, contact traces are left behind, and the contacts (i.e. where the screen is being touched) are visualized as sizable dots.

In relation to the workshop to which we are presenting this work, we pose the organizers and attendees the simple question: given that we have "maxed out" our efforts in achieving "photo" and "spatial" realism in telepresence systems such as HP's Halo, can there be a role for "non-photorealistic" representations (NPR) of remote parties? As stated earlier, NPR representations can accentuate important and de-emphasize unimportant aspects of reference space. Figure 3 illustrates the utility of NPR in a medical context. Imagine using these images to learn about the different

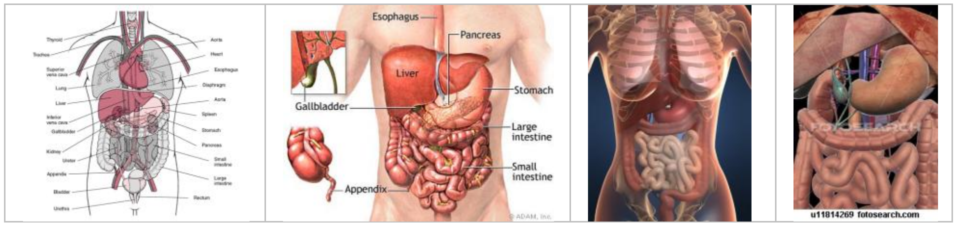

Figure 3. Each of these images illustrate the placement, size and shape of the human organs. Depending on the specific information needs of the user, the additional detail (e.g. 3d contour, realistic colour, cutaway skin, ribcage) is useful or distracting.

organs of the human body: we see that in many cases, the additional detail (3D contour, realistic colour) actually does not provide additional meaningful information, and in many cases, obscures the point altogether. Similarly, how can moving away from "realism" enhance telepresence systems?

Reference-Space Embodiments

Our use of video-based embodiment technique was based on a set of design requirements focused primarily on being able to capture and deliver awareness information about remote collaborators to enable smooth shared work. In contrast to the use of telepointers, for example, video-based embodiments (particularly on touch-sensitive surfaces such as electronic whiteboards or tabletops) allow for: (1) gestures and motion to be created in-context (with respect to the work), (2) explicit communicative gestures, and (3) consequential communication, where information about users' activities and intended actions are communicated (by virtue of being able to see what they are doing).

Figure 4. The video-based embodiment technique attempts to recreate a bird's eye-view of collocated collaborators.

Since the early work with analog video cameras and half-silvered mirrors (e.g. Tang & Minneman, 1990), many researchers have explored the approach using digital technologies. These researchers have generally indicated that the reference space embodiments were useful in aiding the collaborative activity (e.g. Wilson, 2005; Luff et al., 2006; Pauchet et al., 2007; Tuddenham & Robinson, 2009). In practice, however, there still exist subtle problems arising from the technology in use, and more fundamentally, the mental model inherent in this approach.

Philosophically, the approach attempts to recreate the visual scene of the tabletop *as if* remote parties were also co-present (e.g. Figure 4). This recreation has many subtle consequences: the arms are distracting since take up a considerable amount of space on the table; in a given moment of time, a user can only be represented by an image of his/her arms (as opposed to the trace-approach that we illustrate in Figure 2), which is problematic because of the rapidness of some gestures; the hand itself may obscure the view of the objects being manipulated. The capture technology (a camera) also presents some technical challenges: first, the flat, planar projection of the image can create the impression of interactions with the surface when in fact the user may simply be waving his/her hands over the table; the false-projection also does not take care of per-user parallax as in the physical world; the camera magnifies objects that are close to it (above the table), making what should be insignificant extremely large, and finally, the "view frustum" of the camera fails to account for significant volume of space above the tabletop.

Nevertheless, there is good reason to keep using the video-based embodiments technique to support reference space: namely, from the *production* side it is easy to use hands and arms to generate gestures (it is what we do in everyday life), and at least in the context of touch-sensitive surface computing, they are a rich source of information. A simple contrast is to consider the "gesture" system that is common in video games (such as World of Warcraft or Second Life): here, canned gestures can be "executed" in the system through a few keystrokes or a button press. This mediated form of gestures, while interesting, likely lacks the expressiveness of unmediated corporeal hand gestures.

Our hypothesis is that on the *consumption* side, the visual representations of these arms do not need to be photo-realistic, and we are in fact exploring and studying alternate visual representations that emphasize the important aspects of those gestures while de-emphasizing the distracting attributes. In effect, we are hoping to support a reference space that is just as expressive as or more expressive than the equivalent face-to-face scenario.

Current Work

We have embarked on two related threads of inquiry to support expressiveness in reference space. In the first project, we have conducted a study of collaborative behaviour around maps in order to identify the different types of hand gestures, and what aspects of these gestures need to be emphasized. Building from these ideas, we have begun sketching embodiment representations that fulfill the needs as outlined from the mapping study.

Study of Gestures: Goals and Practice

We have recently performed a series of qualitative examinations of natural gesture interactions in reference-space with co-located participants. The examinations were in two parts: The first was a series of artificial information-sharing tasks between two participants in a lab setting and the second was a set of observations performed during natural information elicitation and planning activities between professionals. In both cases, video of the interactions was coded for deixis type and each gesture was characterized in a variety of categories, from the practical, physical components of the gesture, to the intent of the participant in making the gesture.

Although we found a wide range of gestural behaviour, we identified two items of interest with respect to the development of NPR techniques. First, approximately 80% of all gestural interactions observed could be described as index or two-finger pointing at a single target. Of the remainder, the majority were either variations on pointing or were open-hand gestures, which were primarily used to indicating areas on the surface. This finding simplifies the problem of developing usable NPR techniques, since it suggests that effective NPR techniques need only model two, relatively easily identified deictic gestures to be effective for a large proportion of deixis under these conditions. Second, we identified several characteristics of gestures critical for nuanced natural communication but that are difficult to represent in NPRs. Examples include the height of the gesture from the surface; subtle movements such as wiggles, pressure changes, or tapping; occlusion adjustments; and references to targets out of the boundaries of the display, above the surface, or that have a height component (e.g. a bridge over a river).

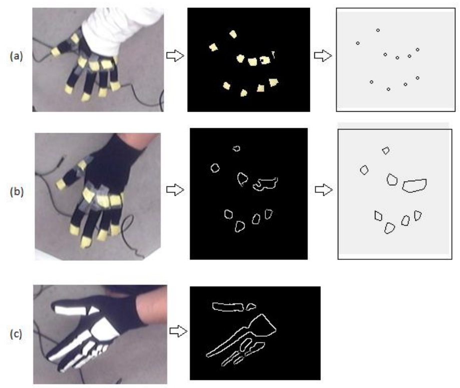

Figure 5. Some different representations we have been exploring.

Based on our observations, we have begun developing a taxonomy of deixis with respect to collaborative interactions. This taxonomy has emphasised the wide range of gestures, that, based on narrative context, physical location, target qualities, and personal tendencies, can express the same conversational goal – a clear challenge for the creation of any NPR of deixis.

Sketching Visual Embodiments
We have begun exploring the design space for visual embodiments. As illustrated in Figure 2, we have tried contact traces, contact points, alpha blending, and colour tinting. To take a further step into the NPR space, we have also tried a number of more radical representations, illustrated in Figure 5.

Figure 5a illustrates a "pointillist" representation we were considering. With motion, this representation is considerably more convincing, and relies on perceptual motion gestalt, where our minds can effectively reconstruct the "body of the hand" even when it is missing. The psychological concept that we rely on here is called "biological motion," and this effect has been demonstrated in the context of gross body motion (e.g. running), as well as interpretations of facial motion (Vatikiotis et al., 1998). In Figure 5b, we restore some of the contour of the markers, and this gives a better view when the hand is tilted to the side. Figure 5c gives a slightly different view of the same idea (albeit with a different type of marker). Some of the questions these sketches explore are question such as: What is the necessary/important part of the gesture that needs to be communicated? What is the role of the "body" of the hand? Is it sufficient to send points of data rather than an entire video? Many of these questions can be answered by inspection, and in large part, our thinking has been informed by study of collaborative mapping behaviour described above.

We see the sketches in Figure 5 to be a starting point—we have yet to consider deeply the rendering side (even though the images actually confound the issue of capture and rendering). For instance, there is the possibility of using temporal traces of the entire visualization, and placing emphasis on certain edges or events (e.g. Wigdor et al., 2009); further, we may explore contour renderings (like pencil sketches) to achieve notions of depth.

Workshop Goals

In participating in this workshop, it is our goal to essentially put forth this notion of moving *away* from "realism" in order to achieve our ultimate goal, which is to support work, presence, and collaboration. While the predominant theme in telepresence work (as exemplified by HP Halo) has been to build experiences where "*you feel like you're there*," our thinking is that we need to focus more on the goal (i.e. supporting work and communication), which means we need to explore the broader design space. In art and graphics research, researchers have found compelling uses for NPR—perhaps telepresence research can similarly benefit from this slight shift in focus.

References

[1] Buxton, W. Mediaspace – meaningspace – meetingspace. In Harrison, S. (Ed) *Media Space 20+ Years of Mediated Life*, Springer, 2009.

[2] Luff, P., Heath, C., Kuzuoka, H., Yamazaki, K., and Yamashita, J. Handling documents and discrimination objects in hybrid spaces. In *Proc CHI 2006*, (2006), 561-70.

[3] Pauchet, A., Coldefy, F., Lefebvre, L., Picard, S., Bouguet, A., Perron, L., Guerin, J., Corvaisier, D., and Collobert, M. Mutual awareness in collocated and distant collaborative tasks using shared interfaces. In *Proc INTERACT 2007*, 59-73.

[4] Tang, A., Neustaedter, C., and Greenberg, S. VideoArms: embodiments for mixed presence groupware. In *Proc British-HCI 2006*, Springer (2006), 85-102.

[5] Tang, A., Pahud, M., Inkpen, K., Benko, H., Tang, J. C., and Buxton, B. Three's company: understanding communication channels in three-way distributed collaboration. In *Proc CSCW 2010*, (2010).

[6] Tang, J. and Minneman, S. VideoDraw: a video interface for collaborative drawing. *ACM Transactions on Information Systems 9*, 2 (1991), 453-469.

[7] Tuddenham, P. and Robinson, P. Territorial coordination and workspace awareness in remote tabletop collaboration. In *Proc CHI 2009*, (2009), 2139-2148.

[8] Vatikiotis-Bateson, E., Kuratate, T., Munhall, K. G., & Yehia, H. C. The production and perception of a realistic talking face. In *Proc LP'98*, Item order in language and speech 2 (pp. 439-460).

[9] Wigdor, D., Williams, S., Cronin, M., Levy, R., White, K., Mazeev, M., and Benko, H. 2009. Ripples: utilizing per-contact visualizations to improve user interaction with touch displays. In *Proc UIST 2009,* 3-12.

[10] Wilson, A. PlayAnywhere: A Compact Tabletop Computer Vision System. In *Proc. UIST 2005,* (2005). 83-92.