uCom: Spatial Displays For Visual Awareness Of Remote Locations

Ana Luisa Santos
MIT Media Laboratory
20 Ames St.
Cambridge, MA 02139 USA
alsantos@media.mit.edu

V. Michael Bove, Jr.
MIT Media Laboratory
20 Ames St.
Cambridge, MA 02139 USA
vmb@media.mit.edu

CHI 2010, April 10–15, 2010, Atlanta, Georgia, USA
ACM 978-1-60558-930-5/10/04.

Abstract
uCom enables remote users to be visually aware of each other using "spatial displays"—live views of a remote space assembled according to an estimate of the remote space's layout. Remote video views from multiple viewpoints are shown individually or in a 3D collage representation that is faithful to the scene geometry. A multi-display setup integrates always-on visual connections of a remote site into the local space. This work applies an innovative spatial context to visual awareness between remote locations.

Keywords
Multi-display, 3D, collage, image-based, awareness.

ACM Classification Keywords
H.5.1 Information Interfaces and Presentation: Multimedia Information Systems, H.5.2 User Interfaces.

General Terms
Design, human factors.

Introduction
The deployment of communication technologies in the last decades has enabled direct interaction among individuals present in different locations. Yet we believe current available technologies could be further

explored to create a sense of proximity between distant coworkers, friends and loved ones. We particularly perceive opportunities to enhance real-time visual awareness among distant individuals using cameras and displays ubiquitous in the places in which we live and work.

We introduce the idea of "spatial displays", i.e., displaying multiple images of a location according to their spatial relationship. We use the spatial or geometric correspondence between remote views as a tool to achieve awareness across connected spaces. The application scenario is a collaborative activity engaging remote users who can benefit from seamlessly visualizing the remote room from multiple viewpoints. For instance, we can support two remote teams of coworkers on discussing ideas with focus on visual elements or objects. Even though our long-term goal is to facilitate remote collaboration, at this current stage of research we focus solely on the visualization aspect. By portraying the static geometric arrangement of a remote room, we aim at enhancing users' perception of the dynamics of the remote space, i.e., activities taking place there.

This paper begins by describing the system concept and related work. Then, we present our prototype's system architecture, followed by user interaction feedback. Finally, we explore future directions for our work.

Concept

uCom stands for "ubiquitous Cameras, observant monitors." It enables remote users to be visually aware of each other using "spatial displays," live views of a remote space assembled according to an estimate of the remote space's layout. The implementation portrays a collection of remote live video views in a 3D spatial context that provides a sense of the scene's geometry, and enables integrating different views of a remote space into the local physical environment.

We have three main design principles. First, the ideal system setup is composed of two physically separate architectural spaces, each one with multiple cameras and displays that mutually transmit live video. We provide mutual awareness to create a sense of connection between spaces, while minimizing privacy concerns. Second, we focus on visual awareness, not communication. We convey always-on visual awareness of a remote location, also integrating remote views into the physical environment. Third, the system setup is flexible in that it can easily create multiple geometric correspondences between two remote spaces, and also opportunistically use available equipment. We support most off-the-shelf video cameras or webcams, and most displays, whether monitors, TVs or portable devices. Equipment placement is flexible. Displays are freely placed to meet users' changing needs, while cameras are placed with few restrictions described in the Prototype section. Our focus on a flexible arrangement anticipates the growing use of low-cost or portable cameras and displays.

Related Work

Elements of our system draw inspiration from previous projects in areas such as "media spaces," awareness, videoconferencing, telepresence, Computer Supported Cooperative Work, groupware, and multi-display applications. Below we mention related work on four areas and describe how our system draws on them.

Spatial coherence between remote rooms
Even though several videoconference systems share our goal to convey a spatial connection between remote locations, they often focus on simulating stimuli common to face-to-face meetings, or on portraying a remote room as an extension of the local room. Common stimuli are: spatial sound, life-sized images of remote users [1], and eye contact [2]. A remote room is portrayed as an extension of the local room when both rooms have identical layout, furniture and lighting, or by simulating a shared meeting table, as per Cisco Telepresence™ or HP Halo™. Our system, in contrast, can be installed in a regular room, minimally interfering with its layout. Yet we acknowledge the discontinuity of our proposed multi-display interaction as we use variable display sizes, specifications and placement.

Image-based reconstructions of real-world scenes
Real-world scene reconstructions use images acquired from either viewpoints with a common center of projection or distinct viewpoints. We are interested in image-based scene reconstructions that use the latter, rather than synthesizing a photo-realistic view of the world. For that, we use the scene modeling method from Photo Tourism [3, 4] and draw inspiration from Microsoft Photosynth™ scene visualization. We extend their original focus on virtual tourism using Internet photos from touristic places, to our application on visual awareness across connected spaces.

Multi-display Applications
We draw inspiration from multi-display applications, such as interactive rooms or smart meeting rooms, which usually have different kinds of displays or projected areas, varying in size and position; and multiple user interfaces [5]. "The office of the future" [6] was envisioned to project onto any local surfaces live views of a remote office space. Yet it required a complex setup to dynamically estimate reflectance of all surfaces. Our system is different in that it can easily utilize almost any cameras or screen-enabled devices at flexible positions, and also scale in richness as further equipment becomes available.

Awareness Applications
Fostering awareness between remote workspaces has been extensively explored by "Media spaces" [7, 8, 9], audio- and video-based remote collaboration systems. They often focus on (1) fostering serendipitous communication or (2) supporting two or more remote users who are intensely collaborating on a common task. Media spaces usually create communication modes appropriate to specific situations [10]. uCom, in contrast, focuses on visually portraying the remote space itself. Yet it can be applied to any remote collaboration activity that can benefit from constantly visualizing the remote room from multiple viewpoints.

Prototype

The system architecture in each space comprises multiple cameras, displays and computers to mutually transmit live video views between two remote spaces.

Image Acquisition and scene computation
The system configuration stage computes the scene model using Bundler[1], the scene recovery algorithm from Photo Tourism [3, 4]. We acquire still images of the scene that overlap by at least 50%. Cameras must remain in the same position to later transmit live video. This unordered set of images is input into Bundler,

[1] http://phototour.cs.washington.edu/bundler/

which outputs a sparse 3D model of the scene geometry composed of: (1) feature points, both 3D position in the scene and 2D position in the respective images, and (2) relative position and orientation of cameras. Bundler computes correspondences between images by detecting and matching SIFT feature points, and the scene model by simultaneously estimating camera parameters and 3D location of feature points.

Image display and User interface

Local displays portray live video views of the remote space with two possible display formats: a 3D collage of multiple views, and individual views.

The 3D visualization shows live video views of a remote space in spatial context resembling the 3D collage from Microsoft Photosynth™. Users can navigate the 3D scene using the keyboard to zoom, pan and switch between cameras' points of view. Transitions between views use transparency gradients and follow paths faithful to the spatial relationship between images. We create this visualization using Bundler's outputs to estimate the relative positions of all images, assuming that most surfaces of the scene are walls. First, we compute a 3D plane by projecting the 3D feature points visible from the respective image onto the scene's ground plane. We fit the projected points into a line and raise a plane containing both this line and the scene's up vector [4]. Second, we compute the image position by projecting the position of the camera's image sensor through its center of projection, assuming the image sensor is parallel to the plane.

The visualization format for individual views integrates live video views into the local physical space, creating always-on visual connections with a remote site. Users assign a single remote live view to each local display using a remote control's arrow buttons. Up/down arrows scroll the unordered list of displays, and left/right arrows scroll the available remote views, while a visual beacon appears on the selected display. Displays can be freely placed throughout the space, subject to equipment type and room layout.

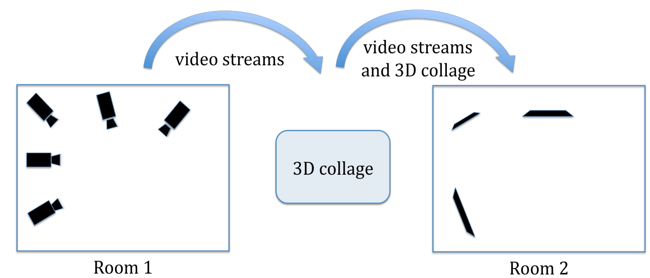

figure 1: System architecture in one direction.

Figure 1 shows the system setup in only one direction, i.e., cameras in one room transmit video to displays in another room. It illustrates the role of the 3D collage as a tool to aid users' understanding of the geometric arrangement of a remote room, also supporting users in deciding upon the displays' assignments and placements. Additionally, both visual representations can be present in the connected spaces.

We deploy our system's custom software using C and C++ to integrate Bundler with libraries such as: libVLC[2] to stream video; libSDL[3] to control the video buffer and user interface; OpenGL to render videos on a 3D GUI; Apple Quartz Composer ™ to assign video views to

[2] http://wiki.videolan.org/Developers Corner

[3] http://www.libsdl.org/

displays using an Apple remote control ™. Our software runs on the Mac OS 10.5 operating system.

Individual views shown in the 3D collage

(a)

(b)

figure 2: Experiment's 3D collage seen from two viewpoints.

User feedback

Our experiment assesses initial user feedback on two research questions: (1) if our system enables the understanding of a remote location's layout, and (2) how users react to incorporating remote views into their everyday physical environment. For the purpose of this initial experiment, the system was setup at a conference room with three movable displays and two stationary displays, seen in Figure 3. One subject at a time is located at the conference room to participate in the experiment. A total of eight adults not familiar with the layout of the spaces were recruited as subjects. They were presented to four still images of a common area of the same building in both individual and 3D collage display formats. The research method comprised observing a subject doing tasks such as: (1) attempting to describe the spatial layout of the remote space using both the 3D collage and a tiled view, and (2) arranging the individual remote views on the local displays. Upon each task completion, the investigator asked questions to assess the user's reaction.

figure 3: Experiment's multi-display setup.

The study started by showing the user the 3D collage from Figure 2, the least standard of the visualization formats. We then presented the tiled arrangement of the respective individual images. The user was asked to comment on how, or if, each format enabled the understanding of the remote location's layout. Users' feedback suggested a preference for the tiled arrangement (6 out of 8 users), as visual distortions on the 3D collage hindered identifying scene features to mentally match images. Yet most users (7 out of 8) understood the remote space's layout using the 3D model, especially due to its 3D navigation features.

A second part of the study comprised observing user interactions with the displays. We asked a user to assign individual remote views to the available displays, repositioning displays if desired. Although we did not suggest a motivation for this task, most users (6 out of 8) assigned remote views to displays according to the geometry of the portrayed scene. The remaining users either placed wide angle views at a farther distance or displayed only the views perceived to be interesting.

Conclusion and Future Work

Our main contributions derive from applying an innovative spatial context to visual awareness between remote locations. They include applying a 3D collage visualization to remote awareness, using available cameras and displays, and allowing users to easily incorporate remote views into the local environment. Additionally, uCom's potential is reflected in users' positive reactions to the remote space's representation using the 3D collage, and to the ease with which they could understand remote views in the local space.

Future work will evaluate other use case scenarios, such as: (1) user agreement while assigning and arranging live video views in a shared space, (2) its use in home or public environments rather than work settings, and (3) its use in the context of remote collaboration on specific tasks. Additionally, future system updates shall minimize 3D collage distortions.

Acknowledgments

We thank the support of Jim Barabas, Yannick Assogba, Quinn Smithwick and the MIT Media Lab sponsors.

References

[1] Gibbs, S.J., Arapis, C. and Breitenender, C. Teleport - towards immersive copresence. *Multimedia Systems*, 7 (1999), 214–221.

[2] Nguyen, D. and Canny, J. Multiview: Spatially faithful group video conferencing. *In Proc. CHI 2005*, ACM Press (2005), 799-808.

[3] Snavely, N., Seitz, S.M. and Szeliski, R. Modeling the world from internet photo collections. *Int. J. Comput. Vision 80*, 2 (Nov. 2008), 189-210.

[4] Snavely, N. Scene Reconstruction and Visualization from Internet Photo Collections. Doctoral thesis, University of Washington, Dec. 2008.

[5] Johanson, B., Fox, A. and Winograd, T. The interactive workspaces project: Experiences with ubiquitous computing rooms. *IEEE Pervasive Computing 1*, 2 (Apr.-June 2002), 67-74.

[6] Raskar, R., Welch, G., Cutts, M., Lake, A., Stesin, L. and Fuchs, H.. The office of the future: a unified approach to image-based modeling and spatially immersive displays. In SIGGRAPH, 179–188, 1998.

[7] Stults, R. Media space. *Xerox PARC Technical Report*, 1986.

[8] Dourish, P. and Bly, S. Portholes: Supporting awareness in a distributed workgroup. *In Proc. CHI 1992*, ACM Press (1992), 541–547.

[9] Fish, R.S., Kraut, R.E., Root, R.W. and Rice, R.E. Evaluating video as a technology for informal communication. *In Proc. CHI 1992*, ACM Press (1992), 37–48.

[10] Gaver, W., Moran, T., MacLean, A., Lovstrand, L., Dourish, P., Carter, K., and Buxton, W. Realizing a video environment: Europarc's rave system. *In Proc. CHI 1992*, ACM Press (1992), 27-35.