TWO APPROACHES TO CASUAL INTERACTION OVER COMPUTER AND VIDEO NETWORKS

Alan Borning and Michael Travers

Rank Xerox EuroPARC
61 Regent Street
Cambridge CB2 1AB
United Kingdom

ABSTRACT
We describe two systems that use interactive computer-controlled video for shared awareness and casual communication. *Polyscope* lets users monitor a large number of video sources simultaneously. Observers are provided with a window containing a collection of frame-grabbed bitmap images or animations. These images can be used to access additional video services, such as videophone. *Vrooms* is a follow-on system, which employs a strong spatial metaphor. Users can enter and leave virtual rooms. Once in a virtual room, users can see and be seen by all the other occupants, and have easy access to other video, audio, and text-based communication tools.

KEYWORDS: group work, collaboration, casual interaction, video, virtual spaces

INTRODUCTION

Despite the ongoing telecommunications revolution, physical proximity is still extremely important to co-operative work [4]. The everyday life of people in organizations depends upon the chance encounters and easy access that result from people sharing physical space. The ultimate aim of our project is to create the same atmosphere of casual awareness and informal interaction between people at sites that might be physically separate.

We are interested in supporting shared awareness: distribution of general information about the environment, both physical and social. Such information includes who's here, what they are doing (if they want this to be known), whether they are available for interactions, and what's happening in the common areas. We also want to support informal interactions: the sorts of conversations that occur around coffee pots, mailboxes, in the hallways, and the like. In such interactions, the purpose, duration, and degree of involvement of the interaction is not planned in advance (in contrast to e.g. a scheduled meeting), but is negotiated in a subtle and dynamic fashion during the course of the interaction. We believe that support for these two activities should be intertwined, since awareness is a prerequisite for informal interaction.

We have constructed two systems to explore these issues: *Polyscope* and *Vrooms*. Both make use of the extensive collection of audio-video gear at EuroPARC, which includes video cameras and monitors, microphones and speakers, a framegrabber, and workstations with bitmap displays. The video and audio equipment is all interconnectable via a computer-controlled switching network [2], and the workstations are on an Ethernet computer network. Every researcher's office at EuroPARC is equipped with video and audio input and output devices, as are many of the common areas, so we have a unique opportunity in the Laboratory to experiment with ubiquitous audio-video.

PRIVACY CONCERNS

Ubiquitous video raises numerous concerns about privacy. How do people feel about having cameras in their

---

Authors' current addresses:
Alan Borning
Dept. of Computer Science and Engineering, FR-35
University of Washington
Seattle, Washington 98195 USA
borning@cs.washington.edu

Michael Travers
MIT Media Laboratory E15-494
20 Ames St.
Cambridge, Massachusetts 02139 USA
mt@media-lab.media.mit.edu

offices and in public areas? What sorts of controls are felt to be necessary? Will a significant number of people simply not want cameras in their offices at all? To a considerable extent we and other researchers at EuroPARC have been attempting to answer these questions empirically, by putting systems in place and gathering peoples' reactions. However, early in the course of this project we proposed two abstract design principles for video systems: *control* and *symmetry*.

I should be able to *control* whether or not others can see me, and under what conditions. Thus, I should be able to choose whether, when, and with what sort of notification I allow a frame-grabbed image to be sampled, or a full video connection to be made. Alternatively, I should be able to in effect post an electronic notice that says "do not enter," and be assured that it will be respected.

Symmetry means that I can see you if you can see me. This principle doesn't mean that I *must* see you if you can see me, only that I have that option if I should choose to exercise it. Further, the kinds of information should be comparable: if you can see full live video of me, I should be able to see full live video of you; if you can see frame-grabbed images of me, I should be able to see frame-grabbed images of you; and so forth.

These principles were arrived at through a combination of *a priori* concerns (a search for simple rules that would satisfy a desire for video privacy), and discussion with potential users. In the discussions, control was universally regarded as important; symmetry was controversial. We decided to provide both as options, then see how well they fared in the light of implementation and actual use. (See the "Experience" section later in the paper.)

POLYSCOPE

Polyscope allows an observer to monitor a large number of video sources conveniently and simultaneously, by providing a two-dimensional array of frame-grabbed video images in a workstation window. The observer can select which of the available sources to monitor; these images are then updated periodically at an interval determined by the observer. On the source side, potential sources of video (e.g. researchers with video cameras in their offices) can decide whether or not to make video available.

There are two types of Polyscope windows: Observers and Sources. An Observer window (Figure 1) displays some number of screen images in a rectangular grid.[1] Each image is is labeled with its name and last update time. The display is tailorable in various ways. The user can select a subset of the available sources, can control whether or not they are animated and how often they are updated, can rearrange the display, and can decide whether or not new sources are added automatically as they become available.

An interesting feature of the system is that we can provide a simple form of animation. Rather than a single video image, we can grab several images in rapid succession, and then display them repeatedly. This sort of animation has proven effective in showing at a glance whether there is activity in an area being observed, and is fun to watch.

A Polyscope observer window is also an interface to other services provided by the audio-visual system. Clicking on an image will pop up a menu of options, including *glance* (a brief one-way connection with feedback—the analog of poking one's head in someone's office) and *videophone* (a longer-term, two-way, negotiated connection).

A Source window (Figure 2) controls a particular video source, typically a camera on a desktop in a researcher's office. Source windows implement the control and symmetry principles described in the "Privacy" section in the following manner.

The Mode line in a source window controls what information is being made available: no information, a short text message, or manual or automatic video. In manual video mode, an image is grabbed only at the request of the owner of the source window. In automatic video mode, images are automatically grabbed under control of the observers (typically once a minute). Grab Now, Save, and Restore are used in conjunction with manual video mode. Grab Now records images when pressed. Save and Restore write and read images to and from files. For example, if a user has a useful image or animation—say of being on the phone, or of typing furiously—he or she can save it and use it again later.

The Feedback line specifies what sort of feedback is desired: Off, Names Only (a list of names of all observers watching this source), or Video (video images of all observers watching this source). Finally, the Video Symmetry Required line specifies whether or not symmetry is required. If no, then any observer can view this source, irrespective of whether or not that observer in turn makes video information available. If yes, then to see video from this source, an observer must in turn make video available.

EXPERIENCE WITH POLYSCOPE

Polyscope was in experimental use at EuroPARC for two months by a dozen researchers and administrators.

[1] Due to space limitations in this paper Figure 1 shows only 4 images, but typically 15–20 sources are available and can be displayed simultaneously. The images shown are 200 by 150 bits, with no gray scale.

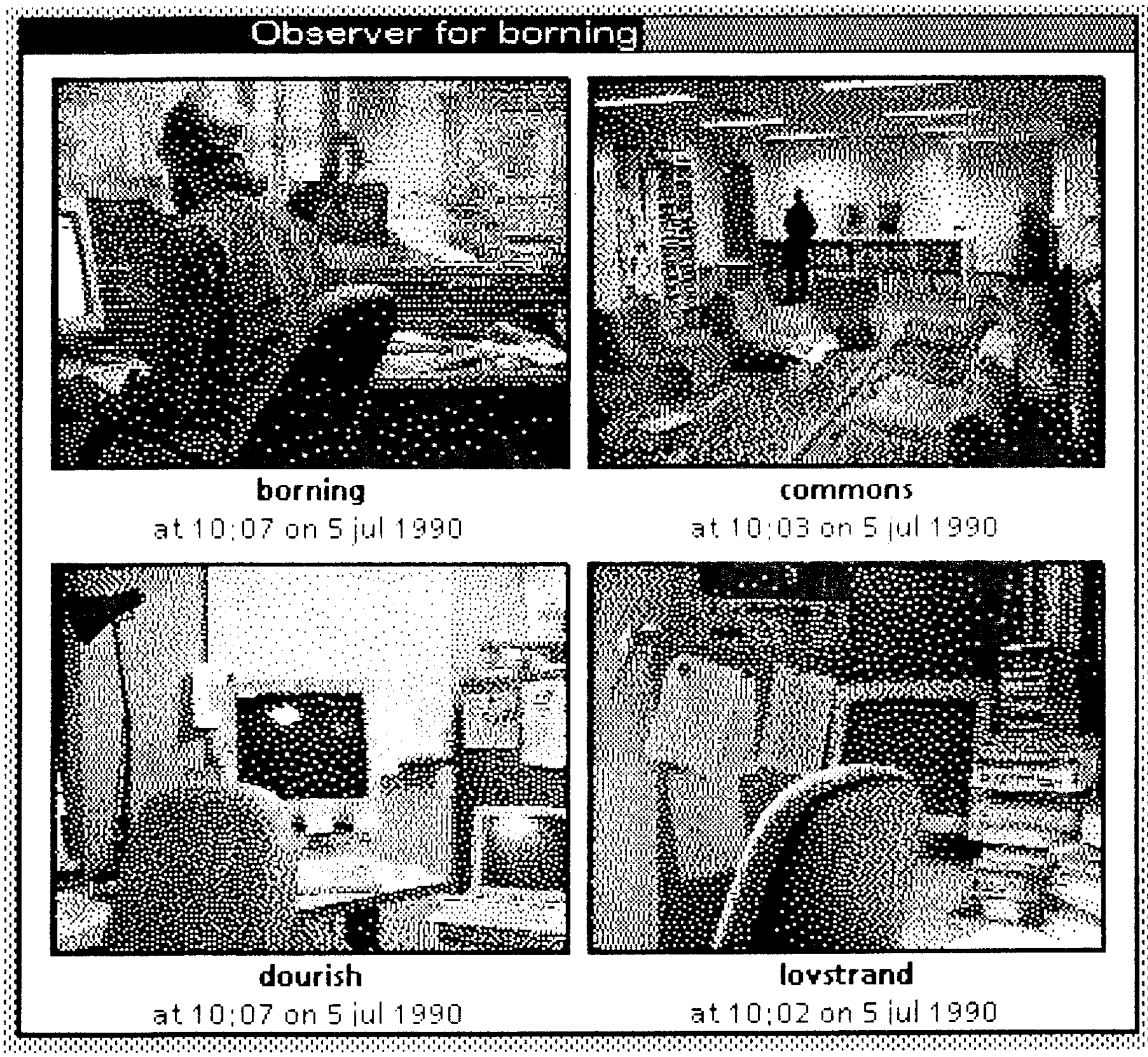

Figure 1: A Polyscope observer window

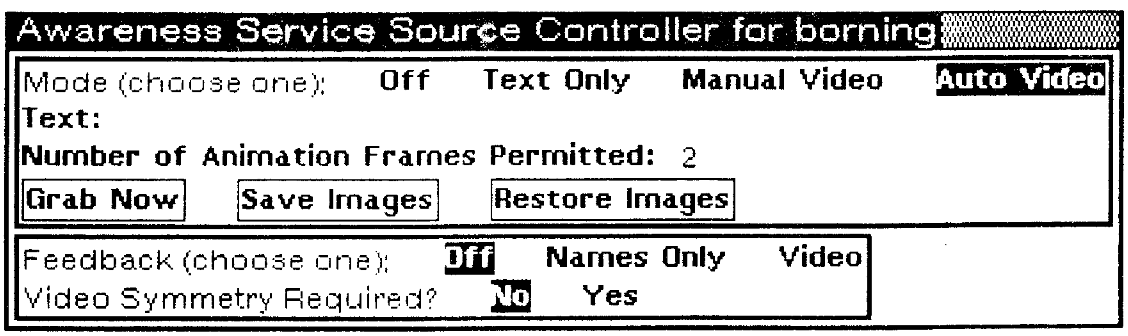

Figure 2: A Polyscope source window

Users commented favorably on the facility it provides to view many video sources simultaneously—a facility not otherwise available with current EuroPARC systems—and on its interface to other AV systems. They commented unfavorably about the reliability of the initial release (subsequently much better), and about its speed (still not great).

The manual video facility proved unexpectedly popular. When the system was first released, there was a rash of "home movies" on display, for example, a researcher mutating into Elvis Presley. In response to user requests, in a second release of the system we added Save and Restore buttons to make it easier to provide these custom images. Once the novelty factor had worn off, the number of home movies diminished, but not to zero. However, these home movies are moving away from the original Polyscope goal of awareness of what is going on *now*. In future systems, then, it would be worthwhile to allow home movies and up-to-date frame-grabbed images to co-exist. (See the "Future Work" section below.)

We instrumented the system to gather statistics about its use; we were particularly curious about how the abstract control and symmetry principles fared in a real implementation. The results for symmetry were surprising. In the abstract, people often commented that feedback and symmetry were important. In practice, Names Only feedback was used approximately 22% of the time, and Video feedback less than 1%; feedback was off the other 77%. (These statistics are for a month and a half of sporadic use by 13 users.) People did indicate that they liked having the capability of requesting the feedback. Video symmetry was almost never requested, and despite its abstract appeal, after actually using the system few people thought it useful.

A fundamental problem with Polyscope concerns the way it satisfies the symmetry principle, and to a lesser extent, the control principle. The intent was to enforce a basic property of face-to-face interaction, namely, the mutual accessibility and visibility of participants. This goal was satisfied, but at the cost of an interface with various options, which forces users to make explicit decisions about their accessibility and visibility.

Another weakness of the symmetry principle is that one user having Auto Video on doesn't necessarily mean he or she is providing qualitatively similar information to another user with Auto Video on. For example, the camera positioning might be different, or the lighting might provide a less clear picture. (For a time one researcher had his camera focussed on his keyboard, providing an image of typing hands or nothing. This gave information, but not the same sort as a wider view into an office.) Finally, the actual experience with the Video Symmetry switch implies either that this isn't a useful way of satisfying the symmetry principle, or that this principle is unimportant in this application. (The small bitmap images provided in Polyscope are much less intrusive than full live video; people seem to feel more strongly about symmetry in this latter case.)

VROOMS

Our second system, Vrooms, deals with some of the social and interface issues raised by Polyscope, in particular the way Polyscope satisfies the symmetry and control principles by using control panels with explicit options.

Our metaphor here is one of virtual rooms (vrooms). A virtual room is a place for social interaction and conversation. Each vroom is a workstation window containing a Polyscope-like collection of bitmap images. This presentation metaphor allows us to dispense with some of the complexity of Polyscope's interface. Instead of many options, we make use of people's tacit knowledge of what it means to be in a room. Being in a vroom with others means that you can see them and they can see you, and that both have the option of initiating a conversation, and that this conversation will usually be more casual than one whose initiation involves a purposeful trip by one person to the other person's location.

Virtual rooms differ from physical rooms in that one can be "in" more than one vroom at a time. This implies that being in a virtual room involves less commitment than being in a physical room. Further, the slow refresh rate and low resolution of our frame-grabbed images mean that less information is available about the real activity of the person behind the virtual presence—not altogether a negative property, since it is less intrusive.

The current implementation uses a client-server architecture. A server program maintains a database of all existing vrooms, their occupants, and other details. A client program runs on the machine of each vrooms user, and exchanges information with the server using remote procedure calls.

A user sees a vroom as a window on his or her workstation. A vroom has small animated, frame-grabbed images of each of its occupants, positioned arbitrarily within the vroom. The vroom can also hold various sorts of shared objects, such as communication devices. In the current implementation the only such sort of shared object is a text box, in which users can type statements. Discounting communication delays, all users with a window open on the vroom see the same contents. If any user makes a change—for example moving his or her image or typing in a statement in a text box—the change is sent to the server, which up-

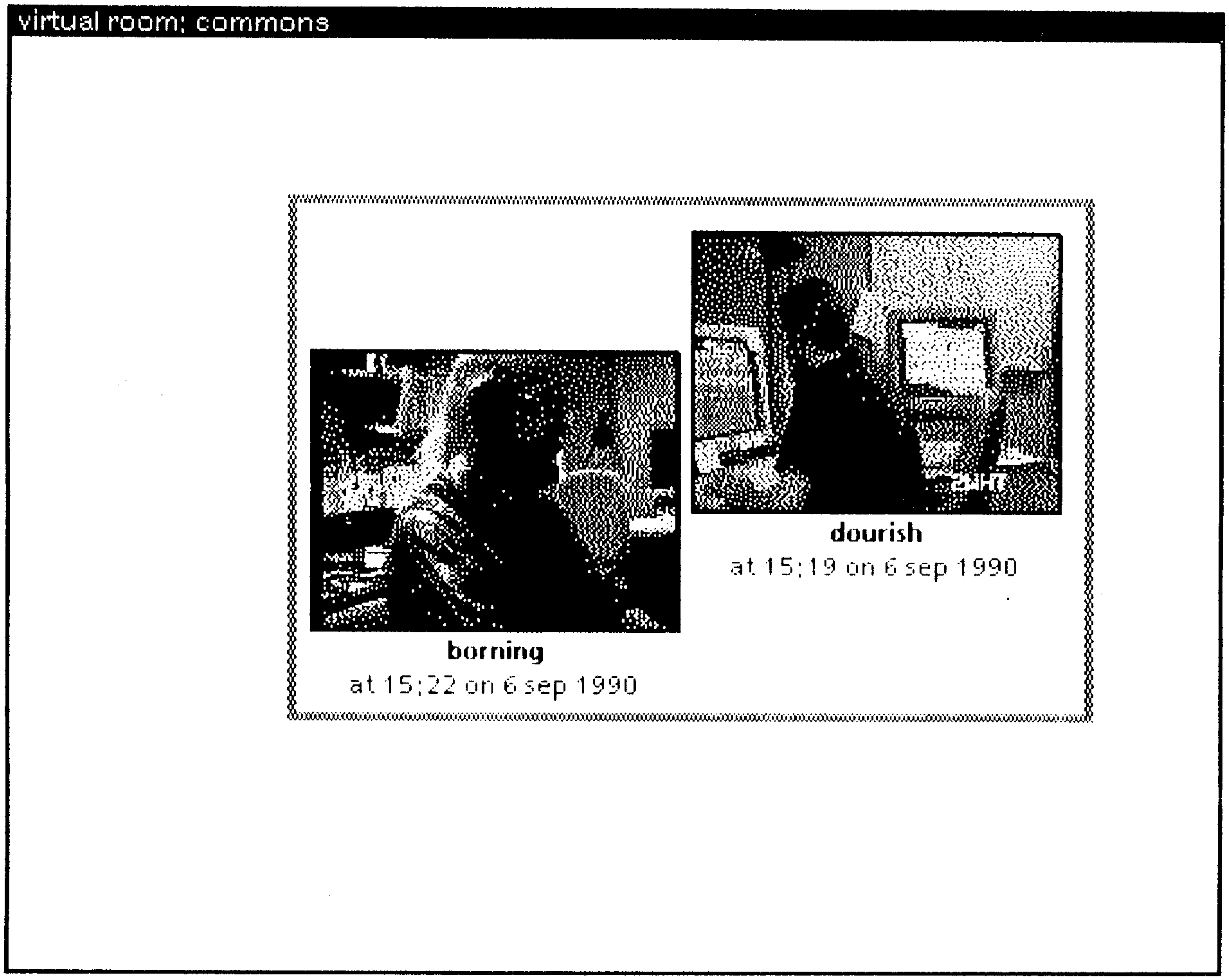

Figure 3: Setting up a full video and audio connection in a vroom

dates its own database and also broadcasts the change to all other users with an open window on that vroom.

Top-level access to vrooms is via several buttons [6] on the user's display: *enter existing vroom*, *create new vroom*, and *delete vroom*. Pressing *enter existing vroom* creates a new window on a vroom; *create new vroom* creates a new vroom as well as a window on it. A user's image remains in the vroom as long as the corresponding window is open on his or her screen; when the window is closed, the user automatically leaves the vroom. Closing the window doesn't delete the vroom from the server's database, however; others may still be in it. Such a deletion can be performed using the *delete vroom* button.

Thus, in Vrooms the symmetry principle is automatically and unobtrusively satisfied. The actions of opening a window on a vroom and entering it are bound together by the user interface, as are those of closing the window and leaving a vroom. Once I'm in a vroom, I can always see all the other occupants, and they can see me.

Once in a vroom, an occupant can move his or her image around in the room, and can create, delete, or move text boxes. More importantly, an occupant has several ways of communicating with other occupants. If an occupant moves close to the image of another occupant, a full two-way video and audio connection is immediately set up between the two on a separate monitor (Figure 3). When either person moves away, the connection is broken. The vrooms system provides feedback that a connection will be set up—if an image is moved close enough to another, a thick grey box will be drawn around the two images of the people in conversation. This box appears in the windows of all occupants of the vroom, to let others know that a conversation is taking place. In addition, any user can type text into a shared text box.

The Vrooms interface for setting up a full video/audio connection is simpler than the videophone protocol used by Polyscope and other systems at EuroPARC. This standard protocol requires explicit acts by the caller to to select the person being called, by the called person to accept the call, and by either or both to terminate the call. In Vrooms, the act of entering a vroom sets up an implicit agreement that one is willing to be contacted by other people in that vroom, thus obviating the need for these more explicit acts.

RELATED WORK

The Cruiser [7] project at Bell Communications Research shares many of the goals of Polyscope and especially Vrooms. It uses multi-media, including video and audio, and a virtual space metaphor to support casual communication. Cruiser provides virtual hallways which one can cruise, and in the process encounter other people or peek into offices. It implements the symmetry and control principles. Perhaps the most significant difference between Cruiser and Vrooms is that, for privacy reasons, Cruiser explicitly does not support continuous video or activity monitoring. In contrast, in Vrooms users are encouraged to hang out in one or more vrooms (represented there by a frame-grabbed image). Based in part on the experience with Polyscope, we believe that these small images represent a small enough intrusion and commitment that it would be acceptable to remain for long periods in a vroom.

The use of virtual rooms as places to communicate was inspired by the multi-user adventure game TinyMUD [1], as well as earlier programs such as the Plato system's Talkomatic. The metaphor of rooms has been used in other interfaces as well, e.g. Rooms [3]. Shared manipulation spaces are provided by such systems as Colab [8], VideoDraw [9], Shared ARK, and Shrdit.

The text box mechanism in Vrooms is related to a window-based talk facility used at Project Athena at MIT. Originally designed to broadcast systems messages about printer availability and the like, the facility is now used to provide informal consulting, to arrange late-night Chinese restaurant trips, and many other sorts of informal interactions [5]. We find this usage pattern very encouraging for Vrooms. Also related are the Unix "talk" program, and the Talk facility in Xerox Interlisp.

FUTURE WORK

Both authors were visiting researchers at Rank Xerox EuroPARC while this work as being done, and have now returned to their respective home institutions. However, we hope that others will continue work in this area. One important activity would be involving users in the design and evaluation of future systems of this sort. As described above, we did have some discussions with potential users before constructing Polyscope, and gathered some preliminary statistics and user reactions to it. However, more extensive user studies would be appropriate. Vrooms became operational just before the end of the visit, and it would be important to gather user experience with it, and then involve these users in a re-design if in fact it seemed to fulfill a need in the workplace.

An interesting question is whether people's intuitions about social space (i.e., how close to approach someone that you know slightly in order to suggest that you want to converse) will work in a non-embedded view such as Vrooms provides. A Vrooms window provides a "god's-eye view" of the social world, in which you appear in the same format as everyone else, in contrast to ordinary life, in which you view other people from a radically different perspective than that from which you view yourself. This issue may be an important one in the design of computer systems that support social interaction.

Our full design for Vrooms was more elaborate than the system that we had time to implement. One important additional feature in this full design is a mechanism to allow transitions from two-way to multi-way conversations using split-screen images and audio mixers. We would like to provide additional sorts of shared communication objects, such as shared text editors, shared drawing programs, and posted announcements or pictures. (Our implementation uses object-oriented techniques, so as long as these other objects obey the correct protocol they could be accommodated.) Given the popularity of Polyscope's manual video and animation facilities, we'd also like to have "art vrooms" specifically for users to display interesting custom animations that they have produced.

Finally, a "door" mechanism would allow moving among vrooms. A door to another vroom could be placed in a vroom; when the user clicked on it, he or she would leave the current vroom (thus deleting its window from the screen), and enter the new one. One could optionally remain in the existing vroom as well as entering the new one. (A feature of being in virtual rather than physical space is that one can be in several places at once.) Doors could also be used to control access to some vrooms, for example by checking visitors against a list of allowed occupants, by asking permission to enter of those already in the vroom, by locking a vroom from inside, and so forth. Each person's office would have a corresponding vroom as well. Borrowing an idea from Cruiser [7], doors to these offices could also be placed in a corridor vroom. Individuals would of course control access to their own offices: we envision this being done by altering the door icon to be closed, open a crack, half-open, fully open, and so forth. (Again, we are trying to tap into the rich set of conventions people have now for using physical doors.) For an open door, one could see a much-reduced image of what was in the vroom.

The focus of Polyscope and Vrooms as they are currently implemented is somewhat different: Polyscope emphasizes awareness and monitoring, Vrooms casual communication. On the balance we prefer the Vrooms approach, since it is more solidly grounded in people's spatial intuition. Therefore, we are interested in augmenting the Vrooms design to encompass more of Polyscope's awareness functions. For example, we might make the capability for immediate video connections be a property of some vrooms but not others. Then, rather than a Polyscope observer window, one could have a common vroom without immediate video connection capability in which most people routinely hung out, with minimal commitment. Again, whether users find the Vrooms approach more intuitive, and whether they would in fact find it useful for this awareness function, needs to be investigated.

The full benefit of these systems would become apparent only after they are used to connect physically remote sites. Recently, Paul Dourish and others have implemented the lower-level parts of a system that spans multiple sites (EuroPARC in England and PARC in California), allowing frame-grabbed images from cameras in both labs to be viewed on a single display. We hope that Polyscope or Vrooms-like systems will be implemented on top of these lower-level protocols.

ACKNOWLEDGEMENTS

Many thanks to all the EuroPARC researchers who provided a rich environment for doing research on interactive video, and who used and commented on these systems, and in particular to Paul Dourish and Lennart Lövstrand for technical help. Thanks also to Michael Sannella for analyzing the Polyscope data, and to Sara Bly, Marc Davis, Paul Dourish, Bjorn Freeman-Benson, Ed Grossman, Jiarong Li, John Maloney, Tom Moran, Michael Sannella, and Molly Wilson for helpful comments on a draft version of this paper. This work was supported by Rank Xerox EuroPARC.

REFERENCES

[1] D. Applegate, J. Aspnes, T. Freeman, and B. Yee. TinyMUD, 1990. Unpublished computer software.

[2] W. Buxton and T. Moran. EuroPARC's Integrated Interactive Intermedia Facility (iiiF): Early Experiences. In *Proceedings of the IFIP WG8.4 Conference on Multi-user Interfaces and Applications*, Heraklion, Crete, Sept. 1990.

[3] D. A. Henderson and S. K. Card. Rooms: The Use of Multiple Virtual Workspaces to Reduce Space Contention in a Window-Based Graphical User Interface. *ACM Transactions on Graphics*, 5(3):211–243, 1986.

[4] R. Kraut, C. Egido, and J. Galegher. Patterns of Contact and Communication in Scientific Research Collaboration. In *Proceedings of the Conference on Computer-Supported Cooperative Work*, pages 1–12, Portland, Oregon, Sept. 1988. ACM.

[5] W. Mackay. Personal communication, July 1990.

[6] A. MacLean, K. Carter, L. Lovstrand, and T. Moran. User-Tailorable Systems: Pressing the Issues with Buttons. In *CHI '90 Conference Proceedings*, pages 175–182, Seattle, Apr. 1990. ACM.

[7] R. W. Root. Design of a Multi-Media Vehicle for Social Browsing. In *Proceedings of the Conference on Computer-Supported Cooperative Work*, pages 25–38, Portland, Oregon, sept 1988. ACM.

[8] M. Stefik, G. Foster, D. G. Bobrow, K. Kahn, S. Lamming, and L. Suchman. Beyond the Chalkboard: Computer Support for Collaboration and Problem Solving in Meetings. *Communications of the ACM*, 30(1):32–47, 1987.

[9] J. C. Tang and S. L. Minneman. VideoDraw: A Video Interface for Collaborative Drawing. In *CHI 90 Proceedings*, pages 313–320, 1990.