

## Supporting Group-to-Group Collaboration in Videoconferences

Andreas Girgensohn, John Boreczky, Patrick Chiu, Jonathan Foote, Lynn Wilcox, Stephen Smoliar  
*FX Palo Alto Laboratory*  
*{andreasg, johnb, chiu, foote, wilcox, smoliar}@pal.xerox.com*

### Abstract

*This paper describes our experiences in an ongoing collaboration between our research group in Palo Alto and a similar group in Japan. The collaboration is based on semimonthly videoconferences. Due to cultural and language barriers, we found it important to use a variety of communication techniques, ranging from audio and video to text-based chats. During the collaboration, we encountered a number of technical issues that caused communication breakdowns, and have taken various steps to improve the videoconference experience. In addition to improving the sound quality by using better microphone placements, we developed a number of systems that support different aspects of a videoconference. One of the new components is a system for panoramic two-way video communication that digitally combines images from an array of inexpensive video cameras. We also provided a means for the remote display of presentations, made design changes to our conference room, and added support for taking meeting minutes and for reviewing captured videoconferences. These changes have resulted in an improved videoconference experience and in a more effective collaboration with our Japanese colleagues.*

### 1. Introduction

Providing proper support for group-to-group videoconferences is difficult. We have been engaged in a series of videoconferences between our research group at FXPAL and a Fuji Xerox group in Japan once or twice a month for the last nine months. In these conferences, we have discussed common projects and the productization of FXPAL research. The conferences consisted of both structured presentations and free-form discussions. Because of cultural and language differences, even face-to-face conferences can be difficult. Videoconferencing makes subtle facial and verbal cues much harder to detect, so it was important to use many different forms of communication. We used audio and video channels, presentation material, shared drawing, and chat.

Effectively using these multiple communication streams proved to be a complex problem.

Videoconferences were held in a conference room designed to support local presentations and meetings as well as simple person-to-person videoconferences. It seemed natural to use the meeting capture technology already present in the room to support group-to-group videoconferencing with Fuji Xerox. Over the course of the nine months, we made many adjustments to the room and experimented with new technologies to address problems we encountered. The new technologies included a panoramic video system, means for the remote display of presentations, and support for taking meeting minutes and for reviewing captured videoconferences.

In the next section, we discuss the problems that we encountered in adapting our meeting capture technologies to videoconferences. The following sections describe the components of our collaborative videoconference support. We continue with comparing our approach to the related work. In conclusion, we describe the benefits of our approach and directions for future research.

### 2. Meeting Capture and Videoconferences

We have conducted and captured local meetings in the Kumo conference room at FXPAL for several years. That room was equipped for meeting capture with cameras, microphones, and a large rear-projected display. The cameras were controlled by a skilled operator, who selected the camera and view to be recorded. Microphones were placed in the ceiling to keep an informal feel to the room.

Audio and video were recorded and stored in the MBase [8] video management system. In addition, images of presentation material displayed on the large display were captured, time stamped, and stored with the video to allow review of presentation material after the meeting.

We found that simply transmitting the audio and video being recorded locally in Kumo did not work well for the videoconferencing scenario. Because of language

differences, a high quality audio signal was essential for understanding. A large part of the problem was the sound quality resulting from having a conference between two large rooms with almost a dozen people in each room. The microphone placement that sufficed for making recordings resulted in audio that was inaudible when sent across international phone lines. In order to improve sound quality, better microphone placement was required, as was a higher quality phone line.

We found that an audio-only connection without a video link led to an awkward, disembodied feeling during the conference. Others have reported that video is useful when meeting participants have cultural and language differences [23]. However, transmitting the video being recorded with the multiple cameras in Kumo was not effective. It required a person to operate the cameras, and the view that was transmitted was often not the view the remote participants wanted.

We then tried allowing the steerable cameras to be controlled from the remote side. While that allowed the remote participants to focus on areas of interest, this approach turned out to be too cumbersome to be truly useful. We finally added a panoramic video system in the hope to replicate the affordances of “face-to-face” meetings at a distance. Instead of physically moving a steerable camera, the equivalent effect can be achieved by extracting an interesting portion of a larger image. When combined with person tracking, this can automatically create a good video image of the speaker without the need for a human operator.

We also encountered problems with sharing presentation material. Due to limited image resolution, the video of the presentation material that was transmitted was not legible. We tried using NetMeeting [16] for the remote control of a presentation but could not make it work through the two firewalls between our locations. An early solution of copying the presentation material before a meeting and giving “next slide” instructions led to several communication breakdowns. We finally expanded our screen capture service that was already used for meeting capture to automatically provide updated screen images.

For better note taking during meetings we integrated our meeting minute and meeting capture tools. These tools enable several people to take time-stamped minutes during the meeting and make them available on Web pages after the meeting. The tools also provide a means to capture presentation material during the meeting. Both the presentation material and the notes are linked to the video on the Web page for the meeting. By making these tools available to participants on both sides of the connection, we are able to provide a comprehensive overview of the videoconference.



Figure 1. Initial meeting setup

### 3. Video Transmission

As discussed in the previous section, we first attempted to give the remote participants more control of what they saw by providing several steerable mmEye cameras [3] that transmit video streams via HTTP and can be controlled from the remote side. Figure 1 illustrates the video streams of that setup. Much of the screen real estate was taken up by images from the local and remote cameras. Even in that situation, we already utilized a second display to show one of the camera images.

For improving the communication, we started with the assumption that higher picture resolution and a wider field of view can help communication via video [19]. Though there remain significant problems such as gaze direction and audio, we have been experimenting with using panoramic video to enhance the videoconferencing experience. We feel that panoramic video has two major value propositions. The first is to replace a human camera operator. Instead of physically moving a steerable camera, the equivalent effect can be achieved by extracting an interesting portion of a larger image. This can automatically create a good video image of the speaker without the need for a human operator. The second advantage of panoramic video is that it can show the group reaction to a speaker by showing a wide-angle view of the entire room. This reaction is highly critical during meetings, and is difficult to convey to a remote location.

We developed FlyCam, a system that generates a seamless panoramic video image from multiple adjacent cameras [7, 8]. The name alludes to the compound eyes of insects that form sophisticated images from an array of sensors. FlyCam component cameras are mounted on a rigid substrate such that each camera’s field of view overlaps that of its neighbor. The resulting images are aligned and corrected using digital warping, and combined to form a large composite image. The result is a seamless high-resolution video image that combines the views of all cameras.



**Figure 2.** FlyCam webcam application

The FlyCam server application functions as a virtual steerable video camera, allowing each client to request an individual view from the panoramic image. Figure 2 shows the client; the virtual camera is steered by clicking on the panoramic image or the left/right arrows, while the “+” and “-” zoom controls have the obvious functionality. Unlike other webcams that use a steerable camera, every client can choose its own unique combination of pan, tilt, and zoom

Though a user can easily select a desired image, we can eliminate human input entirely for a truly automatic camera control system. To this end, we have implemented algorithms that automatically select an appropriate virtual camera view. Because the FlyCam is fixed with respect to the background, motion analysis does an excellent job of detecting interesting foreground objects, such as the speaker. Thus the system can serve as an automatic camera operator, by steering a virtual camera at the most likely subjects. For example, in a videoconference, the camera can be automatically steered to capture the person speaking.

#### 4. Redesigning the Room for Effective Videoconferences

Our early person-to-person videoconferences in Kumo used the one main display to view the remote participants. The camera was located at the edge of the display. When a local participant wanted to share presentation materials such as PowerPoint slides, the materials were placed on the main display, replacing the view of the remote participants.

As we began to conduct group-to-group videoconferences, we added a camera at the side of the room that was controllable by the remote participants. We added a second display to allow simultaneous display of presentations and remote cameras (see Figure 1). However, this camera placement did not allow good eye contact.

While trying to improve the videoconference experience, we found that we had to pay closer attention to the room design and the placement of cameras and displays. It is clear that camera views and presentation materials should be viewable on any display to allow flexible arrangements. Each display needs a camera so that eye contact can be maintained.

Figure 3 presents a room design that was used at a recent videoconference. The local participants sit around a table in front of the room. The large display shows the slides of the remote presenter. The smaller displays to the sides show views of the remote presenter, the remote audience, and the local audience. A FlyCam is placed under the left display that shows the remote audience, so that the proper gaze direction is maintained. In this configuration, the remote presenter and the remote audience still share a display. That makes it impossible to establish a semantically consistent space for gaze communication [21] in which the remote participants know when a local participant looks at them.



**Figure 3.** Videoconference using multiple video panoramas.

Figure 4 shows the rendering of a 3D model for one envisioned redesign of the room. With the remote audience display on the side wall, this design gives the impression that the local and remote rooms are located side-by-side. The display for the remote speaker placed at an angle on the front wall on the same side as the remote

display. For a local presenter standing at the front of the room, it might be better to place the remote audience display on the back wall so that the presenter can easily see both the local and the remote audience and thus ensure proper gaze direction. We plan to mount both side and rear displays to allow for flexible room configurations.



Figure 4. 3D model of conference room changes.

## 5. Slide Image Capture

In many of our videoconferences, we gave presentations to each other using PowerPoint slides. In early videoconferences, we tried using NetMeeting [16] for the remote control of a PowerPoint presentation but could not make that work through the two firewalls between our locations. The alternative of copying the PowerPoint file before a meeting and giving “next slide” instructions led to several communication breakdowns when the instructions were missed and different slides were shown at the videoconference sites. We finally expanded our screen capture service that was already used for meeting capture to automatically provide updated screen images. Slides are captured by a screen-capture component on the PC workstation used to drive the meeting room main display. Images from the display are captured in equal intervals. Captured images are compared to the previous image and saved if a change occurred. We found that capturing images once every two seconds provides sufficient information without interfering too much in the normal operation of the PC or requiring too much bandwidth for transferring the captured images to the remote site. Although some image quality is sacrificed by using screen captured image rather than the original PowerPoint, this technique supports other types of presentations such as those based on Web pages.

The captured slide images are made available from a Web server. During videoconferences, the *LiteSlideShow* applet periodically checks for new images from the slide

image server and displays them if a change occurred. In Figure 3, the presentation screen shows a Web browser with the *LiteSlideShow* applet that displays a slide from the remote presentation.

The slide images captured during a videoconference are also saved as part of the meeting recording. The time that a particular slide is displayed by the presenter is recorded along with the slide image. That allows us to use the slide images as entry points into the recorded video and to combine them with the meeting minutes.

## 6. Video Management

We record the videoconferences to provide later access to their content. This provides the opportunities for conference participants to later review passages during which important topics were discussed. Other users might have missed the conference and want to review it without having to spend a whole hour watching the entire video. The FXPAL MBase system [9] was developed to address such needs for local meetings and we adapted it to be used for videoconferences as well. During both local meetings and videoconferences, the video can be shot by a camera operator or automatically. A camera operator has the ability to direct multiple cameras and follow the speakers in close focus. However, that requires much additional effort so that we have started recording the video produced by the FlyCam view selected by motion tracking. This also provides the same view that the conference participants have.

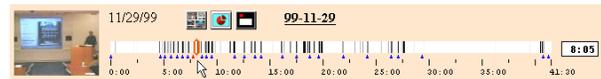


Figure 5. MBase entry for a video.

The video is recorded directly in MPEG, stored in the MBase system, and can be played back right after a videoconference. Information about captured meetings can be accessed from the MBase system. This Web-based system provides access to video collections that are organized into directories grouped by topic (e.g. project reviews, seminars). A number of video analysis techniques are used to give users summaries of and access points into the videos. Figure 5 shows an entry for a video. A timeline visualizes different video features such as camera shots. Users can move the mouse over the timeline to see the corresponding keyframes that are marked with blue triangles along the timeline.

MBase offers several different means to access a video. A comic-book-like visual summary, Manga, provides a quick summary of the salient events and good entry points into the video. The captured presentation

materials are another set of entry points into the video. To time-align these different media, the clocks of the workstation recording the video, the PC used for the presentation, and other systems producing meeting-related material such as meeting minutes are kept synchronized with the Network Time Protocol. This protocol provides sufficient accuracy to correlate events to the right frame in the video.

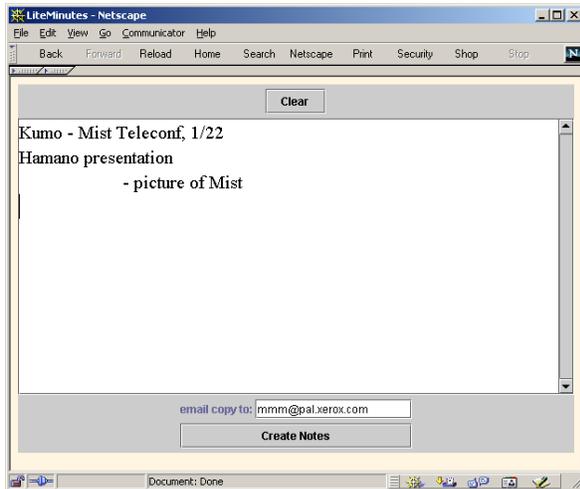


Figure 6. LiteMinutes note-taking applet running inside Web browser.

## 7. Meeting Minutes

From our experience with local meetings, we found that meeting minutes integrated with the captured video and the presentation material make it easier to review the meetings. In our conference room, the heavier weight media such as video, audio, and slide images are captured. These are then combined with the more lightweight and interactive medium of notes taken during a meeting.

From our experience with multimedia note taking systems (see [6, 27]), we find that it is important for a note taking application to support rapid interaction. Taking notes during a live event requires the user to pay close attention and sometimes the note taker has to participate in the meeting in addition to formulating notes. This makes it difficult for novice users to fiddle with user interface widgets and perform tasks such as labeling or organizing information. After some experiments with various hardware devices and software using HTML forms and applets, we arrived at a minimalist Java applet for taking text notes on laptops.

We found laptop computers with a wireless network connection to be well suited for taking meeting minutes. Laptops are familiar devices that require no training for

someone already familiar with a PC workstation and its keyboard interface. Hooking it up to a wireless network allows the note taker to sit anywhere in the room and provides an unobtrusive form factor. While text is limited in expressiveness compared to ink (which supports both writing and drawing), having the text entered while note taking during a meeting saves the scribe from the time consuming task of transcribing the notes after a meeting.

### 7.1. Creating Meeting Minutes

Creating multimedia meeting minutes with LiteMinutes [5] is very easy: meeting participants or a designated scribe walk into a meeting with their laptops or use the wireless laptops supplied in the room. The system can support more than one note taker simultaneously. When several people take notes, all of their notes can be made accessible from our meeting minute Web server and they all point into the same video recording and captured slide images. We currently do not merge individual notes into a single document. Normally, a single scribe takes notes in our videoconferences.

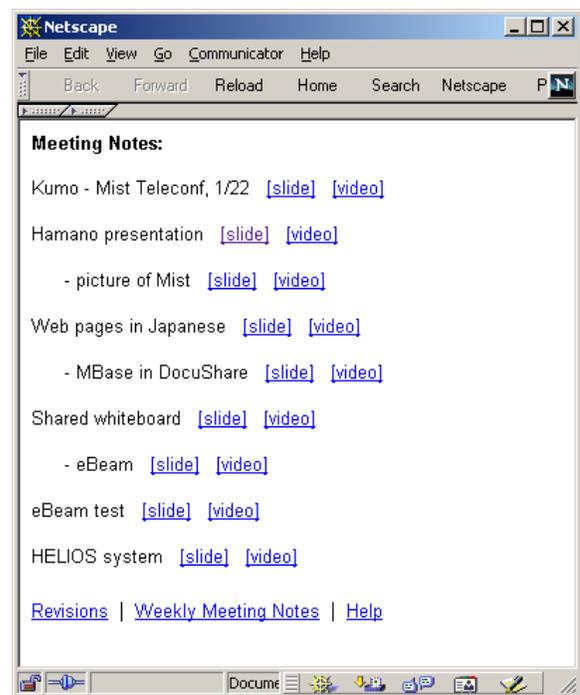


Figure 7. LiteMinutes Web page.

A screen shot of this applet is shown in Figure 6. The user simply takes notes in a text window. In the mean time, the meeting is recorded on video and slide images shown by the speaker are captured. At the end of the meeting, the user submits the notes to a CGI script on the

notes server for processing and distribution. The notes are parsed by the server and formatted in HTML, with a small addition at the end of each note item, where video and/or slide links are appended (see Figure 7).

For viewing of slide images, an applet called *LiteSlideViewer* is used. Like the *LiteSlideShow* applet used for remotely viewing slide images, this applet displays a single slide image. It also provides controls to change slides, to see thumbnails of all slides presented during a videoconference, and to play the captured video from the time corresponding to the shown slide. The slide link of a note item on the page shown in Figure 7 is a CGI script that fetches the *LiteSlideViewer* applet and shows the target slide image.

## 8. Related Work

Other researchers have investigated the limitations of using conventional videoconferencing for presentations and meetings and methods for overcoming these limitations. Microsoft Research has been investigating making video of distributed remote viewers visible to lecturers on a side display [14]. They found that the video images were distracting to the local audience, and that placement of cameras and displays is important. Having remote audiences gathered in central locations, as we do, can cut down on attendance, but adds to the feeling of a shared experience. Systems that do not show the remote audience [12] result in a lower level of interaction. The lack of feedback makes it difficult to gauge how well the remote audience comprehends the communication, which is vital with our language and cultural issues. Even reducing the quality of the video channel seems to have a negative impact on the perceived understanding of the remote audience [13].

There has been considerable prior work in the areas of panoramic video. Many approaches have been to compose existing still images into a panorama that can be dynamically viewed [4, 11], or to compose successive video frames into a still panorama [22]. Because these techniques involve computationally expensive image registration, none of them can be done practically at video rates. In contrast, the system used for our videoconferences uses cameras with fixed, known alignments, so that displacements need not be calculated at all. A group at UNC uses an approach similar to ours, in which 12 video cameras arranged in two hexagons, along with mirrors to form a common center of projection. They compose the panoramic image using the texture mapping hardware of a SGI O2 [15]. For person tracking, systems based on steerable cameras must compensate for camera motion as well as a narrow field of view [26]. This is much less of a problem for a

panoramic camera such as ours with a motionless and much wider field of view.

For systems combining meeting minutes with video, most multimedia notetaking systems are not as easy to use as *LiteMinutes*. The *FXPAL NoteLook* system [6] allows users to incorporate images from the video sources of the room activity and presentation material into the notes, and users can take free-form notes with digital ink. The images and ink strokes are indexed to the video recording for retrieval. It requires training to use and is not designed for novices. Other multimedia pen-based systems also require a certain amount of training; examples are *Audio Notebook* [21], *Classroom 2000* [1], *FXPAL Dynamite* [27], *Filochat* [25], and *Marquee* [24]. *WEmacs* [18] is a text note taking application based on the GNU Emacs editor. Its user interface is more complicated than the *LiteMinutes* text box, and it assigns functions to special characters (e.g., a Tab is used to separate note items). The *Where Were We* system (W3) [17], which is related to *WEmacs*, supports making annotations and video recording during a live event. Each note is created in a separate user interface widget, which makes it difficult to use in a live meeting. There are a number of related video annotation systems (e.g., see [10] for an overview). A more recent system is *Microsoft MRAS* [2]. It is designed for asynchronous video annotation, and supports text and voice annotations. Each text annotation is created in a separate user interface widget. In contrast, *LiteMinutes* has a single text box for arbitrarily many note items to support rapid interaction during a live meeting.

## 9. Opportunities for Consumer Services

The facilities that we have discussed above have contributed significantly towards enhancing the quality of the videoconferences that take place between our laboratory in Palo Alto and our colleagues' laboratory in Japan. However, our solutions are highly technology-intensive, in addition to being media-rich; and their implementation has definitely benefitted from leveraging the infrastructure of a large organization. This has led us to consider the extent to which smaller-scale operations, which are just as likely to require videoconferences to support their sales and production operations, can also benefit from our contributions.

The usual approach to making expensive technology more accessible to a broader customer base is through rental. Using "videoconference" and "rental" as input to a Web search engine, one can now find several companies that have taken this approach. The problem is that the effectiveness of the equipment depends heavily on how well it is configured at the customer site; and

that configuration rarely, if ever, involves a connection to the customer's local area network to support the archiving of a video recording or time-synchronized notes in the manner described above. Furthermore, even without a connection to the customer's network, setting up equipment for a videoconference can be time-consuming and is likely to discourage the customer from scheduling videoconferences very often.



**Figure 8.** The Here's Looking at You conference room.

Another approach is to make a fully-configured site available at some easily accessible public location. Here's Looking at You (<http://www.hereslookingatyou.com/>) is an example of a company that is taking this service-oriented approach for businesses in a large metropolitan area (Houston). They provide a media-rich room with all the necessary telecommunications infrastructure and abundant support for video and audio input. Technical support is also provided at an additional charge. Additional fees also support scan converters, document cameras, and VCRs that can be used for both recording and playback. However, the infrastructure does not appear to support recording quality of the type described above.

We are already used to thinking about service-based approaches to document production, as the success of Kinko's has demonstrated. We believe the technology described above constitutes an integration of services for videoconferencing with those of document production (notes taken during the meeting and video recording of the meeting) and document management (such as is supported by MBase). Our experiments with panoramic video have also demonstrated that, with appropriate user

interfaces, much of the need for additional technical support may be eliminated. The technology described above thus provides the basis for a new service-based approach to supporting videoconferencing that can benefit businesses of all sizes and may even open the videoconferencing market up to more general public use.

## 10. Conclusions

We discussed technical problems that we encountered during the collaboration with our Japanese colleagues in a series of videoconferences. We developed a number of different systems that address different aspects of those problems. First, to improve the remote participants' view, we initially introduced remotely steerable cameras and then replaced them with a panoramic video system. Our panoramic video system, FlyCam, provides a better view of the remote participants and automatically tracks a presenter. Second, a redesign of our conference room placed cameras and displays such that the remote participants were more visible and the gaze direction was correct. Third, in response to problems with remotely showing presentation slides and keeping presentations in sync, we started using the LiteSlideShow system that provides a simple means to present PowerPoint slides and other presentation materials remotely while requiring only a small amount of network bandwidth and working with HTTP proxy servers through firewalls. Fourth, to allow the participants to later review a videoconference and to provide access to other interested parties, we recorded the video and provided different user interfaces to it through our MBase system. Finally, as a further enhancement to the captured videoconference, we used the LiteMinutes system that allowed one or more of the participants to take time-stamped meeting minutes. The observations of the meeting minutes process indicated that a simple text applet on a wireless laptop provided a good way to take notes in a meeting that are later easily accessible on the Web and link to the presented slides and the captured video.

In putting these components together, our remote collaboration has become more effective by staying better on topic and resolving issues more quickly and thoroughly. There are still problems caused by the language and cultural differences, but the wide variety of communication modes has made it possible to conduct efficient videoconferences. There are still a few areas that we want to improve. For example, we are investigating audio source location from a microphone array to augment the automatic tracking [27]. By automatically pointing a virtual camera towards an audio source, the FlyCam system could better capture audience questions. We expect that this and other improvements will make

the videoconferences feel more natural and have them come closer to the experience of a face-to-face meeting.

## 10. Acknowledgements

We thank our colleagues at Fuji Xerox CRC, especially Toshihisa Hamano and Yoshihiro Masuda, and our colleagues at FXPAL, especially Shingo Uchihashi, Tohru Fuse, Don Kimber, Matthew Cooper, Elizabeth Churchill, and David McDonald, for participating in the videoconferences and for many suggestions for improvements. We also thank Surapong Lertsithichai for creating the 3D models for the conference room changes.

## 11. References

- [1] Abowd, G.D., Atkeson, C.G., Feinstein, A., Hmelo, C., Kooper, R., Long, S., Sawhney, N., and Tani, M. "Teaching and learning as multimedia authoring: the classroom 2000 project," *Proceedings of the ACM Multimedia '96 Conference*. ACM Press, pp. 187-198.
- [2] Barger, D., Gupta, A., Grudin, J. and Sanocki, E. "Annotations for streaming video on the Web: system design and usage studies," *Computer Networks*, Elsevier Science, Vol. 31, No. 11-16, pp.1139-1153, 1999.
- [3] Brains Corporation, mmEye Multi Media Server, <http://www.brains.co.jp/>
- [4] Chen, S. and Williams, L. "View interpolation for image synthesis," *Computer Graphics (SIGGRAPH '93)*, pp. 279-288, August 1993.
- [5] Chiu, P., Boreczky, J., Girgensohn, A., and Kimber, D. "LiteMinutes: An Internet-Based System for Multimedia Minutes," *Proceedings of Tenth World Wide Web Conference (2001)*. To appear.
- [6] Chiu, P., Kapuskar, A., Reitmeier, S., and Wilcox, L. "NoteLook: Taking notes in meetings with digital video and ink," *Proceedings of ACM Multimedia '99*. ACM Press, pp. 149-158.
- [7] Foote, J., and Kimber, D., "Enhancing Distance Learning with Panoramic Video," *Proceedings of the 34<sup>th</sup> Annual Hawaii International Conference on System Sciences*, Jan 2001.
- [8] Foote, J., and Kimber, D., "FlyCam: Practical Panoramic Video," *Proceedings of IEEE International Conference on Multimedia and Expo*, vol. III, pp. 1419-1422, 2000.
- [9] Girgensohn, A., Boreczky, J., Wilcox, L., and Foote, J. "Facilitating video access by visualizing automatic analysis," *Proceedings of INTERACT '99*. IOS Press, pp. 205-212.
- [10] Harrison, B., Baecker, R. M. "Designing Video Annotation and Analysis Systems," *Graphics Interface '92*, pp. 157-166.
- [11] IPIX, the Interactive Pictures Corporation, <http://www.ipix.com/>
- [12] Isaacs, E., Morris, T., Rodriguez, T., and Tang, J., "A Comparison of Face-To-Face and Distributed Presentations," *Proceedings of CHI '95*, ACM Press, pp. 354-361.
- [13] Jackson, M., Anderson, A., McEwan, R., and Mullin, J., "Impact of Video Frame Rate on Communicative Behavior in Two and Four Part Groups," *Proceedings of CSCW 2000*, ACM Press, pp. 11-20.
- [14] Jancke, G., Grudin, J., and Gupta, A. "Presenting to Local and Remote Audiences: Design and Use of the TELEP System," *Proceedings of CHI 2000*, ACM Press, pp. 384-391.
- [15] Majumder, A., Seales, W.B., Gopi, M., and Fuchs, H. "Immersive teleconferencing: a new algorithm to generate seamless panoramic video imagery," *Proc. ACM Multimedia 99*, Orlando, FL, pp. 169-178, 1999.
- [16] Microsoft.NetMeeting. <http://www.microsoft.com/windows/netmeeting/>
- [17] Minneman, S., Harrison, S. "Where Were We: making and using near-synchronous, pre-narrative video," *Proceedings of the ACM Multimedia '93 Conference*. ACM, New York, pp. 207-214.
- [18] Minneman, S., Harrison, S., Janssen, B., Kurtenbach, G., Moran, T., Smith, I., and van Melle, B. "A confederation of tools for capturing and accessing collaborative activity," *Proceedings of the ACM Multimedia '95 Conference*. ACM Press, pp. 523-534.
- [19] Moore, G., "Sharing Faces, Places, and Spaces: the Ontario Telepresence Project Field Studies," Chap 14 in *Video-Mediated Communication*, ed. Finn, K., Sellen, A., and Wilbur, S. Lawrence Erlbaum Associates, Mahwah, New Jersey, 1997.
- [20] Stifelman, L. *The Audio Notebook: Paper and Pen Interaction with Structured Speech*. Ph.D. Thesis. MIT Media Lab, 1997.
- [21] Taylor, M.J. and Rowe, S.M. "Gaze Communication using Semantically Consistent Spaces," *Proceedings of CHI 2000*, ACM Press, pp. 400-407.
- [22] Teodosio, L., and Bender, W., "Salient Video Stills: content and context preserved," *Proc. ACM Multimedia 93*, Anaheim, CA, pp. 39-46, 1993.
- [23] Veinott, E., Olson, J., Olson, G., and Fu, X., "Video Helps Remote Work: Speakers Who Need to Negotiate Common Ground Benefit from Seeing Each Other," *Proceedings of CHI '99*. ACM Press, pp. 302-309.

[24] Weber, K. and Poon, A. "Marquee: a tool for real-time video logging," *Proceedings of CHI '94*. ACM Press, pp. 58-64.

[25] Whittaker, S., Hyland, P., and Wiley, M. "Filochat: handwritten notes provide access to recorded conversations," *Proceedings of CHI '94*. ACM Press, pp. 271-276.

[26] Wang, C., and Brandstein, M., "A hybrid real-time face tracking system," *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) '98*, IEEE.

[27] Wilcox, L.D., Schilit, B.N., and Sawhney, N. "Dynamite: A Dynamically Organized Ink and Audio Notebook," *Proceedings of CHI '97*. ACM Press, pp. 186-193.