

PROCEEDINGS OF SPIE



SPIE—The International Society for Optical Engineering

Multimedia Networks: Security, Displays, Terminals, and Gateways

V. Michael Bove, Jr.
Barbara Derryberry
Clifford R. Holliday
Louis S. Lome
Vishal Markandey
Andrew G. Tescher
Bhaskaran Vasudev
Chairs/Editors

4–5 November 1997
Dallas, Texas

Sponsored and Published by
SPIE—The International Society for Optical Engineering



Volume 3228

SPIE is an international technical society dedicated to advancing engineering and scientific applications of optical, photonic, imaging, electronic, and optoelectronic technologies.

QA76.575
.M845
1997



The papers appearing in this book comprise the proceedings of the meeting mentioned on the cover and title page. They reflect the authors' opinions and are published as presented and without change, in the interests of timely dissemination. Their inclusion in this publication does not necessarily constitute endorsement by the editors or by SPIE.

Please use the following format to cite material from this book:

Author(s), "Title of paper," in *Multimedia Networks: Security, Displays, Terminals, and Gateways*, V. Michael Bove, Jr., Barbara Derryberry, Clifford R. Holliday, Louis S. Lome, Vishal Markandey, Andrew G. Tescher, Bhaskaran Vasudev, Editors, Proceedings of SPIE Vol. 3228, page numbers (1998).

ISSN 0277-786X
ISBN 0-8194-2661-X

Published by
SPIE—The International Society for Optical Engineering
P.O. Box 10, Bellingham, Washington 98227-0010 USA
Telephone 360/676-3290 (Pacific Time) • Fax 360/647-1445

Copyright ©1998, The Society of Photo-Optical Instrumentation Engineers.

Copying of material in this book for internal or personal use, or for the internal or personal use of specific clients, beyond the fair use provisions granted by the U.S. Copyright Law is authorized by SPIE subject to payment of copying fees. The Transactional Reporting Service base fee for this volume is \$10.00 per article (or portion thereof), which should be paid directly to the Copyright Clearance Center (CCC), 222 Rosewood Drive, Danvers, MA 01923. Payment may also be made electronically through CCC Online at <http://www.directory.net/copyright/>. Other copying for republication, resale, advertising or promotion, or any form of systematic or multiple reproduction of any material in this book is prohibited except with permission in writing from the publisher. The CCC fee code is 0277-786X/98/\$10.00.

Printed in the United States of America.

2-8-98 612

Contents

vii Conference Committees

Part A Multimedia Networks: Security, Displays, Terminals, and Gateways

SESSION 1 COMPRESSION

- 4 **Software open system for MPEG video and audio transmission [3228-01]**
G. Cabitza, M. G. Setzu, G. Fregonese, Ctr. for Advanced Studies, Research and Development in Sardinia (Italy)
- 13 **Adaptive region-based multiscaled motion-compensated video coding for error-prone communication channels [3228-04]**
P. Bahl, Microsoft Corp.; W.-L. Hsu, Digital Equipment Corp.
- 26 **Streaming CBR transmission of VBR stored video [3228-06]**
S. Sen, J. K. Dey, J. F. Kurose, Univ. of Massachusetts/Amherst; J. A. Stankovic, Univ. of Virginia; D. Towsley, Univ. of Massachusetts/Amherst
- 37 **Comparison of VBR video smoothing and transmission schemes under a unified abstraction model [3228-07]**
J. Zhang, NEC USA
- 49 **Buffer control for variable-bit-rate video over wireless ATM LAN [3228-08]**
L. Nguyen, Carleton Univ. (Canada); S. Panchanathan, Arizona State Univ.

SESSION 2 GATEWAYS AND HOME NETWORKS

- 58 **Networked embedded systems: a look at the issues, approaches, and trade-offs [3228-10]**
C. A. Schwaderer, Microware Systems Corp.
- 84 **Single-chip multimedia network access using installed in-home cable and that can be remotely managed [3228-11]**
M. Gandar, MediaFlow, Inc.; D. Mlynek, Swiss Federal Institute of Technology; A. Osseiran, Engineering School of Geneva (Switzerland)
- 92 **Access network evolution and residential gateways [3228-12]**
R. McLellan, A. Berenbaum, J. Ordille, G. Ryan, Lucent Technologies/Bell Labs.
- 98 **Internet gateway for long-distance 1394 home network [3228-13]**
X. Lu, P.-G. Zhang, W. Y. Chen, Texas Instruments Inc.
- 105 **Breaking open the set top box [3228-16]**
D. Banks, A. Wiley, N. Catania, A. N. Coles, D. Smith, S. Baynham, E. Deliot, R. Chidzey, Hewlett-Packard Labs. (UK)
- 117 **JTV: Java-enabled television [3228-15]**
P. Thrift, T. Killian, Texas Instruments Inc.

LINDA HALL LIBRARY
KANSAS CITY, MISSOURI

iii

Contents

vii Conference Committees

Part A Multimedia Networks: Security, Displays, Terminals, and Gateways

SESSION 1 COMPRESSION

- 4 **Software open system for MPEG video and audio transmission [3228-01]**
G. Cabitza, M. G. Setzu, G. Fregonese, Ctr. for Advanced Studies, Research and Development in Sardinia (Italy)
- 13 **Adaptive region-based multiscaled motion-compensated video coding for error-prone communication channels [3228-04]**
P. Bahl, Microsoft Corp.; W.-L. Hsu, Digital Equipment Corp.
- 26 **Streaming CBR transmission of VBR stored video [3228-06]**
S. Sen, J. K. Dey, J. F. Kurose, Univ. of Massachusetts/Amherst; J. A. Stankovic, Univ. of Virginia; D. Towsley, Univ. of Massachusetts/Amherst
- 37 **Comparison of VBR video smoothing and transmission schemes under a unified abstraction model [3228-07]**
J. Zhang, NEC USA
- 49 **Buffer control for variable-bit-rate video over wireless ATM LAN [3228-08]**
L. Nguyen, Carleton Univ. (Canada); S. Panchanathan, Arizona State Univ.

SESSION 2 GATEWAYS AND HOME NETWORKS

- 58 **Networked embedded systems: a look at the issues, approaches, and trade-offs [3228-10]**
C. A. Schwaderer, Microware Systems Corp.
- 84 **Single-chip multimedia network access using installed in-home cable and that can be remotely managed [3228-11]**
M. Gandar, MediaFlow, Inc.; D. Mlynek, Swiss Federal Institute of Technology; A. Osseiran, Engineering School of Geneva (Switzerland)
- 92 **Access network evolution and residential gateways [3228-12]**
R. McLellan, A. Berenbaum, J. Ordille, G. Ryan, Lucent Technologies/Bell Labs.
- 98 **Internet gateway for long-distance 1394 home network [3228-13]**
X. Lu, P.-G. Zhang, W. Y. Chen, Texas Instruments Inc.
- 105 **Breaking open the set top box [3228-16]**
D. Banks, A. Wiley, N. Catania, A. N. Coles, D. Smith, S. Baynham, E. Deliot, R. Chidzey, Hewlett-Packard Labs. (UK)
- 117 **JTV: Java-enabled television [3228-15]**
P. Thrift, T. Killian, Texas Instruments Inc.

LINDA HALL LIBRARY
KANSAS CITY, MISSOURI

iii

- 123 **Evaluation of medium-access mechanisms in HFC systems with ATM cell transport** [3228-17]
C. Sierens, K. J. Venken, R. Hoebeke, Alcatel Telecom (Belgium); J. D. Angelopoulos,
T. G. Orphanoudakis, National Technical Univ. of Athens (Greece); M. S. Nunes, Instituto
de Engenharia de Sistemas e Computadores/IST (Portugal)
- 135 **Novel dynamic caching for hierarchically distributed video-on-demand systems** [3228-18]
K. Ogo, C. Matsuda, K. Nishimura, NTT Human Interface Labs. (Japan)
- 147 **Distributed video server using a new striping technique** [3228-19]
K. Fukazawa, H. Suzuki, C. Sasaki, NTT Human Interface Labs. (Japan)

SESSION 3 COLLABORATIVE TOOLS AND APPLICATIONS

- 160 **Application-sharing object for multimedia conferencing** [3228-22]
J. B. Oh, J. S. Kim, S. B. Cho, Electronics and Telecommunication Research Institute (Korea)
- 169 **New model for distributed multimedia databases and its application to networking of museums** [3228-23]
K. Kuroda, N. Komatsu, K. Komiya, H. Ikeda, Telecommunications Advancement Organization
of Japan
- 174 **Reflection of presence: toward more natural and responsive telecollaboration** [3228-24]
S. Agamanolis, A. Westner, V. M. Bove, Jr., MIT Media Lab.
- 183 **Multimedia medical consulting system** [3228-25]
K. Komiya, K. Momose, Kanagawa Institute of Technology (Japan); S. Ueno, Tokai Univ. (Japan)
- 190 **ITU T.120 series of standards as a basis for conference applications** [3228-26]
D. Trossen, Technical Univ. of Aachen (FRG); T. K. Helbig, Philips GmbH Research Labs. (FRG)
- 202 **Remote troubleshooting and monitoring of process control systems by the Internet** [3228-27]
D. Yin, Graz Univ. of Technology (Austria)
- 211 **Generation of a 3D hair model from hair volume** [3228-28]
W. M. Kong, H. Takahashi, M. Nakajima, Tokyo Institute of Technology (Japan)
- 219 **Interactive multimedia service system on ATM networks** [3228-09]
T. Nam, B. Kim, T. Jeong, Electronics and Telecommunications Research Institute (Korea)

SESSION 4 DISPLAYS AND HARDWARE ISSUES

- 232 **Color calibration on electronic display systems** [3228-30]
R. Vohsbeck-Petermann, Fraunhofer Institute for Computer Graphics (FRG)
- 238 **User interface using a 3D model for video surveillance** [3228-31]
T. Hata, S. Boh, A. Tsukada, M. Ozaki, Mitsubishi Electric Corp. (Japan)
- 247 **Low-cost high-resolution non-shadow-mask hybrid LCD-CRT display** [3228-32]
K. L. Landel, Media Technology Group; J. L. Johnson, U.S. Army Missile Command

Part B Computer and Network Security

SESSION 5 SECURITY OF OPTICAL NETWORKS

- 260 **Physical security considerations in all-optical networks (Invited Paper) [3228-35]**
D. Marquis, M. Médard, R. A. Barry, S. G. Finn, MIT Lincoln Lab.
- 272 **Smart photonic networks and computer security for image data [3228-36]**
J. Campello, J. T. Gill, M. Morf, M. J. Flynn, Stanford Univ.
- 280 **Survivable all-optical cross-connect meshed-ring communications networks [3228-37]**
I. Rubin, J. Ling, Univ. of California/Los Angeles
- 292 **Secure authentication protocol for Internet applications over CATV network [3228-38]**
L.-P. Chin, Shih Chien Univ. (Taiwan)
- 297 **Overview of image security techniques with applications in multimedia systems [3228-39]**
R. B. Wolfgang, E. J. Delp, Purdue Univ.

SESSION 6 QUANTUM CRYPTOGRAPHY

- 310 **Security of quantum cryptography against individual attacks [3228-40]**
B. A. Slutsky, R. R. Rao, P.-C. Sun, Y. Fainman, Univ. of California/San Diego
- 322 **Evaluation of the cost of error-correction protocol in quantum cryptographic transmission [3228-41]**
L. Tančevski, Laval Univ. (Canada); B. A. Slutsky, R. R. Rao, Y. Fainman, Univ. of California/San Diego
- 332 **Practical aspects of quantum cryptography [3228-42]**
H. Zbinden, G. Ribordy, N. Gisin, Univ. of Geneva (Switzerland)

SESSION 7 OPTOELECTRONIC DEVICES AND ALGORITHMS FOR SECURITY

- 342 **Ultrafast cryptography using optical logic in reconfigurable feedback shift registers [3228-43]**
M. Médard, MIT Lincoln Lab.; A. H. Chan, Northeastern Univ.; J. D. Moores, K. L. Hall, K. A. Rauschenbach, S. A. Parikh, MIT Lincoln Lab.
- 354 **All-optical signal encoder-decoder for secure communications [3228-44]**
W. R. Babbitt, Montana State Univ.; A. Mohan, J. A. Ritcey, Univ. of Washington
- 366 **Low-latency photonic network using ultrashort-pulse transmission and time-domain processing [3228-45]**
A. G. Nowatzky, Digital Equipment Corp.
- 377 **Block-cipher scheme based on back iteration of the 2D cellular automata [3228-46]**
M. Madjarova, M. Kakuta, T. Obi, M. Yamaguchi, N. Ohyama, Tokyo Institute of Technology (Japan)

SESSION 8 OPTOELECTRONIC CDMA SYSTEMS AND DEVICES

- 390 **Fast-wavelength-hopped CDMA system for secure optical communications [3228-47]**
C. F. Lam, E. Yablonovitch, Univ. of California/Los Angeles
- 399 **Spectrally encoded CDMA system using Mach-Zehnder encoder chains [3228-48]**
C. F. Lam, R. B. Vrijen, D. T. K. Tong, M. C. Wu, E. Yablonovitch, Univ. of California/
Los Angeles
- 408 **One-dimensional photonic crystals from CDMA [3228-49]**
S. Wang, H. Erlig, H. R. Fetterman, Univ. of California/Los Angeles; V. Grubsky, J. Feinberg,
Univ. of Southern California
- 418 **Two-dimensional CDMA systems for optical networks with increased security demands
[3228-50]**
L. Tančevski, L. A. Rusch, Laval Univ. (Canada)
- 429 *Addendum*
- 431 *Author Index*

Reflection of Presence: Toward more natural and responsive telecollaboration

Stefan Agamanolis, Alex Westner, and V. Michael Bove, Jr.

{stefan, westner, vmb}@media.mit.edu

MIT Media Laboratory

20 Ames Street, Cambridge, Massachusetts 02139 USA

ABSTRACT

The purpose of *Reflection of Presence* is to create a framework for a telepresence environment that intelligently responds and adapts itself to its inhabitants in order to enhance interpersonal communication or collaboration. The metaphor is that of a “magic mirror” in which one not only sees a reflection of oneself, but also reflections of the other remotely-located participants, just as if everyone is standing in the same room looking at each other through a real mirror. Using visual and auditory cues, segmented images of participants are dynamically layered into a single display using varying transparency, position and scale to reflect center of attention and degree of presence. Wireless tangible interfaces allow participants to customize their shared space and collaboratively manipulate and annotate media objects in the background. The system is novel in that it is implemented totally as a set of cooperating scripts instead of through a low-level programming language, enabling rapid experimental changes in the behavior of the prototype.

Keywords: telecollaboration, telepresence, multimedia, videoconferencing

1 Introduction

The Object-Based Media group at the MIT Media Laboratory concentrates on harnessing the power of digital processing and analysis to move from traditional waveform-like representations of visual and auditory media toward more semantically and physically meaningful “object-based” representations.[1] We are also interested in examining and understanding what kind of impact these alternative models might have on how we entertain and communicate using electronic media.

This work first led us to embark on several projects exploring how object-based thinking might provide new artistic opportunities in the area of interactive cinema. For example, in one of our productions, *The Wallpaper*, various elements of a visual and aural scene (humans, chairs, desks, ambient sounds) are recorded and stored separately from each other in a computer and then recombined in real-time during playback according to instructions in a script. Representing the scene in this way allows the creator of this piece to experiment with changing camera angles and shot selection, swapping characters, and modifying acoustical characteristics based on audience interaction in order to express different subjective points of view of the same story. The piece is also responsive to passive circumstances of its delivery, by showing more or fewer cuts and close-ups depending on the size and shape of the viewing window.

In the past, we concentrated on applications involving mostly pre-recorded and heavily pre-processed objects, but the arrival of more powerful processing hardware has allowed us to apply our research to real-time applications involving “live” objects, which led to our interest in teleconferencing and telepresence.



Figure 1: *Several frames from a live demonstration of Reflection of Presence with three participants. Each sees himself as though looking into a mirror, but a mirror also occupied by reflections of the others. Documents, images, and other media objects may be placed into the background.*

Reflection of Presence is an ongoing project that explores the application of object-based techniques to enhance remote synchronous interpersonal communication and collaboration. Our intent with this project is not to recreate an environment that imitates the feeling of physical face-to-face reality. Rather, we wish to use digital image and audio processing and analysis to create a natural shared communication space that responds intuitively and intelligently to its inhabitants, providing features that enhance and extend natural human communicative abilities.[2] In designing *Reflection of Presence*, we thought less about building a system that would operate within severe bandwidth or processing constraints and instead focused on issues of how people might ideally communicate electronically if these technological restrictions were lessened.

The metaphor we have chosen to work with in *Reflection of Presence* is that of a “magic mirror”—a mirror in which you see not only a reflection of yourself, but also the reflections of the other participants in the virtual space just as if they were standing next to you, looking at you through a real mirror, even though they are all in separate remote locations. However, instead of seeing a reflection of your surroundings in this “mirror,” you may see images, documents, movies or other pieces of media in the background, placed there by any one of the participants (Figure 1).

Instead of consisting of simple static reflections, the composition of this scene is dynamically changing, utilizing varying transparency, position or scale in order to emphasize or de-emphasize the presence of each participant as determined from visual and auditory cues. In a way, the system tracks not only physical presence but also “stage presence,” and uses this information to more intelligently compose the visual display to enhance communication among the participants. The center of attention is made the most prominent figure in the frame while other less active participants fade off into the background. Audio is spatialized

such that if you see someone on your left, then his voice emanates from your left.

We are experimenting with the use of special physical objects that can be tracked by the system to allow participants to interact with and customize the space in simple and intuitive ways—to manipulate documents in the background, control the playback of movies, or annotate the scene as if drawing on the surface of a mirror. Our goal is for the participants to “meet each other, not the system.” [3]

2 Related work

Reflection of Presence combines and builds on many concepts introduced in the work of several other research bodies. The idea of identifying the “center of attention” in a multipoint videoconference can be traced back at least 25 years to a system that uses “voice voting” to automatically switch video from one camera to another in the same conference location so that the current speaker is always seen in close-up to remote viewers.[4] When no one is speaking, a wide-angle camera view of all the participants in a single location is transmitted. A “self view” feature is also provided in which participants have the ability to see the outgoing transmission. The system was mainly designed to eliminate the need for a human technician to track the speaker by manually switching video sources.

Other researchers have presented videoconferences in which all participants appear in the same window. The *Personal Presence System* demonstrates a videoconference in which each participant has complete and private control over the size and positioning of the other participants’ video streams, similar to a window-based display on a personal computer.[5] People who are considered more or less important can be made larger or smaller by manual manipulation. Keying based on color or brightness is used to mask out the background portions of a video stream. Centralized video bridging hardware was specially designed for this system in order to address processing, bandwidth, and video cross-coding concerns.

CU-SeeMe VR presents a videoconference in which participants appear as live video windows in a three-dimensional virtual space.[6] Again, simple keying may be used to mask out portions of the background of each stream. In contrast to the *Personal Presence System* which allows one to move other participants around on the screen, the *CU-SeeMe* environment permits each participant to control only his own navigation through the virtual environment. An extensive point-and-click interface includes, among other things, a map view of the space and a text chat box. Spatialized audio reflects the relative position and orientation of each participant.

The *Clearboard* system pioneers the notion of seamlessly integrating shared work and interpersonal spaces into a common environment.[7] One participant draws on a large glass board or table and at the same time is able to see an image of the other participant and his drawings “behind” the glass. Polarizing filters and half-mirrored drawing surfaces enable the separation of one participant and his drawings from the projection of that of the other. In the first mockup of *Clearboard*, participants drew directly on the glass using colored markers, and in a later version, a shared drawing application, *TeamPaint*, provided the ability for sessions to be stored and recalled. A key concept of the *Clearboard* system is “gaze awareness” and the ability to track the focus of attention of the other participant.

3 Implementation

Reflection of Presence combines many of the concepts introduced separately in previous works with some of our own innovations to create a shared space with unique social characteristics. Our system is also novel in that it is implemented totally as a set of cooperating scripts, enabling rapid experimental changes in the behavior of and relationship between all of the pieces that make up the prototype.

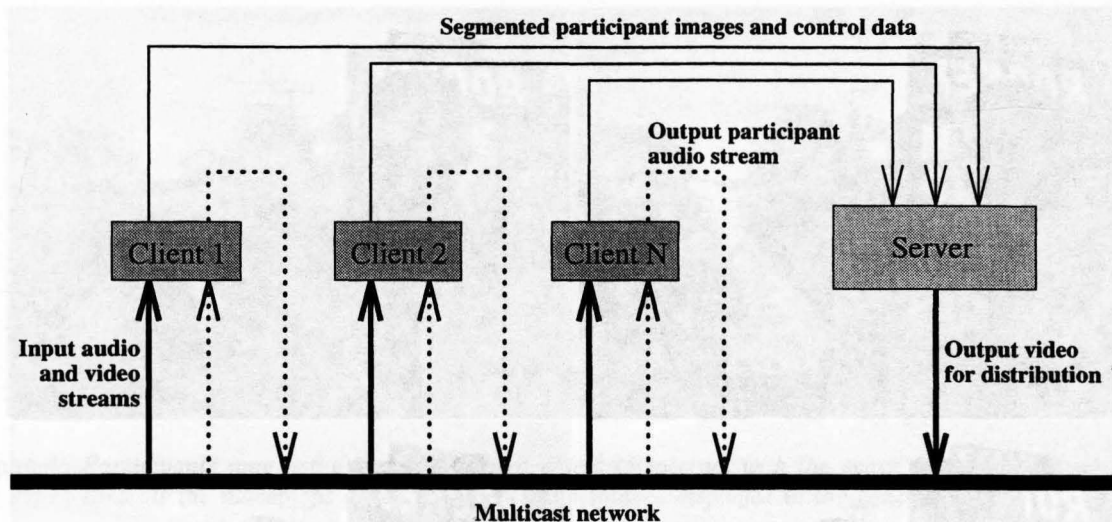


Figure 2: Overview of logical data connections in Reflection of Presence.

3.1 Topology

The system employs a “star” topology in which several clients, all running on separate machines in remote locations, connect to a common central server via TCP and UDP. Each client workstation captures audio and video from a single participant. The image of the participant is segmented from his background and transmitted to the server whose job is to combine all of the participants’ images into a single frame for distribution back to the clients. This basic cycle needs to occur in real-time, many times per second, with relatively low lag, and needs to be synchronized with the distributed audio as well. Figure 2 graphically depicts the data connections employed by the system. In our early experiments, a video network was used to distribute the output from the server back to the clients. Our latest version allows the use of multicast motion-JPEG streams.

3.2 Segmentation

In the current prototype of Reflection of Presence, a special background-adaptive luminance/chrominance difference algorithm is utilized to extract the participant from his surroundings. The output of this segmentation process, performed on the client workstation, is a full 8-bit “alpha” channel which enables the server to create very clean-looking compositions. A bounding box operation is also applied on the client in order to avoid sending full-size video frames to the server, and also to track the position of the participant.

Modifying parameters independently on each color channel allows for more flexibility in perfecting the segmentation under different background and lighting conditions. The algorithm is fast and has served us well in our early experiments, but it is imperfect and still requires capturing a background image prior to beginning a conference session and sometimes at several points during a session if the background conditions change radically for some reason. Interrupting an in-progress session in this manner is obviously highly undesirable.

An updated version of this algorithm accounts for small lighting changes by slowly adapting regions of the captured background image, and it enables automatic calibration of the segmentation while the system is in operation. However, since the client workstation is relatively unencumbered compared to the server, we are able to experiment with more complicated segmentation systems. A future version of Reflection of Presence may use a multidimensional analysis of image attributes such as motion, color, and texture coupled with a



Figure 3: *The center of attention, as selected by the system through a variety of heuristics, occupies the front layer and appears opaque, while others appear partially transparent in the background.*

region-growing scheme, which would not require capturing static background images.

3.3 Audio

After some processing and analysis, the audio stream captured from each client is transmitted directly to each of the other clients in a multicast fashion, bypassing the server. Incoming audio from other clients is mixed and output to speakers or headphones. This approach may not be ideal for perfect audio-video synchronization, but it does prevent each participant from having to listen to a delayed copy of his own voice when speaking. Audio is only transmitted if a participant is speaking or otherwise makes audible noises over a certain threshold, so only a small number of streams need to be mixed at any one point in time. Documents in the background, such as movies, may also generate audio streams that are heard by all participants.

Before transmission, the client audio streams are also spatialized to reflect the position of each participant's image in the output frame, thereby enhancing a listener's ability to track the position of a speaker. For example, if one participant physically moves his body to the left, the other participants will hear his voice moving to the left as well. A participant may even move all the way off the area covered by his camera, and his voice will still emanate from the side of the space that he exited from.

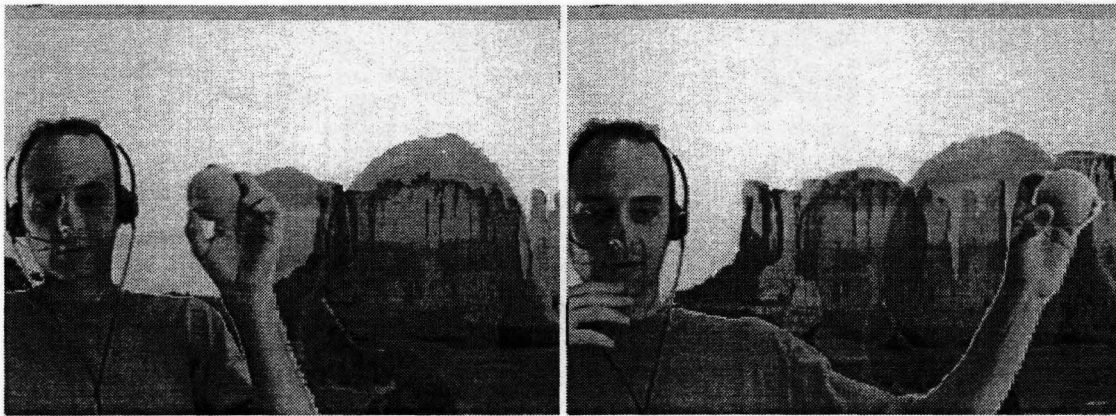


Figure 4: *Participants may use a specially-colored object to interact with the space in various ways. Here, the object controls the movement of the documents and images displayed in the background.*

3.4 Image composition

The client process listens to how loudly each participant is talking and watches how vigorously each is moving in order to determine whether the participant “wants” to become the center of attention. The server uses this information to decide how to render all of the participants in the output frame (Figure 3). For example, if a participant speaks above a certain loudness or raises a hand, he is brought to the front layer and becomes opaque. If the same participant stops talking and remains still for a few moments, he relinquishes the front layer to the new center of attention and slowly fades into the back layers, becoming partly transparent. If more than one participant wants to be the center of attention, a “first-come-first-served” strategy is employed in which they all appear opaque and are composited in the front layers of the frame, but the first participant who seized the center of attention will retain the very front layer until relinquishing it as described before.

3.5 Interaction

In order to communicate and collaborate more effectively, participants in Reflection of Presence need a way to customize and manipulate objects within the shared space they inhabit. We are currently using a specially-colored hand-held object, usually a small ball, which provides a wireless and tangible means of interacting with the virtual environment. The client workstation watches for the presence of this object and sends the server information about its position relative to the “mirror.”

The colored object serves different purposes depending on the context in which it is used. For example, in one mode of the system, participants may place images or documents originating from a local file system or the World Wide Web into the background of the environment. The colored objects act as “scrolling knobs,” allowing participants to navigate in any direction over a seemingly infinite background plane (Figure 4). Here, the positions of the objects relative to the center of the display window act as vectors controlling the movement of the background. If several objects are visible simultaneously, the vectors obtained from each are added together to determine the motion of the background, creating a situation where participants can either cooperate or compete with each other to accomplish a certain task. With this interface, it is also possible to “grab” a particular document and move it to another location in the background.

Another mode allows participants to make annotations by “painting” onto the background (Figure 5). The colored objects serve as paint brushes, and the proximity of the object to the “mirror” can control either the size of the brush or the opacity of the paint. Different colors are selected by “dipping” the object into a palette displayed at the bottom of the screen. In a variation of this mode, participants may simply use their bodies to paint images into the background.

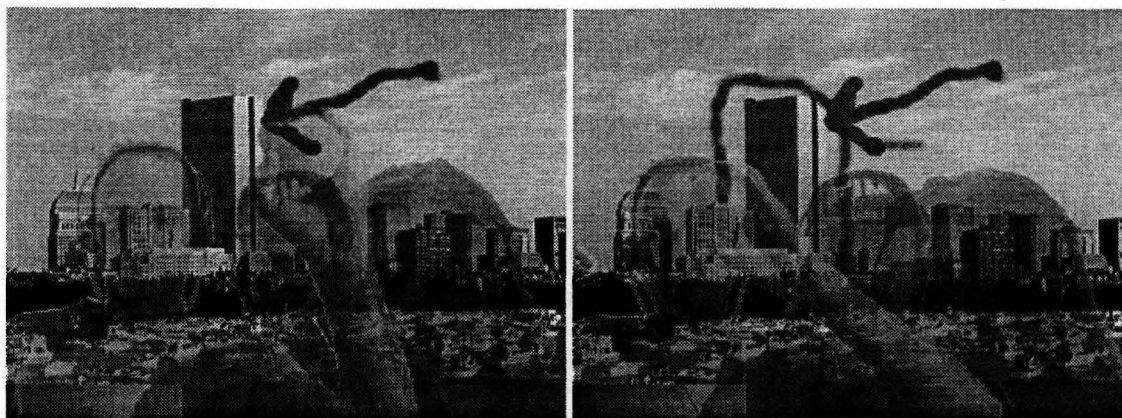


Figure 5: *In a different mode, the colored objects serve as paint brushes, allowing participants to annotate documents or collaborate on a drawing.*

The system also allows for the playback of pre-recorded video sequences in the background with synchronized audio. In this mode, the colored objects serve as a “shuttle” control. When an object is brought into view, the video pauses. Moving the object left or right rewinds or fast-forwards the video at different rates. Live video sources may also be displayed (Figure 6).

Participants switch between modes by selecting options from an overlay menu that can be invoked at any time by holding the object extremely close to the camera. The entries are highlighted in red as an object is moved around the menu. Selections are made by holding an object steady on a single menu entry for a predetermined amount of time. Before an entry is actually activated, it is highlighted in yellow to signify that the time to activation is near.

3.6 Performance

With 3 participants and full-color 320 by 240 resolution output, our prototype runs reliably between 10 and 15 frames per second with a latency that hovers at approximately 200 milliseconds. However, the performance of the system is highly dependent upon the speed of the processing hardware and network involved and whether or not data compression is performed. Due to the availability of an ATM network in our laboratory, our implementation is more limited by processing power than network bandwidth, and therefore we decided in the prototype not to use compression on most data connections. Clearly an implementation over a more limited network, or given dedicated compression/decompression hardware, would compress the video and audio objects.

4 Design environment

One unique aspect of Reflection of Presence is that the entire system is implemented as a set of scripts written in the *Isis* scripting language.[8] *Isis* is, at its core, a full programming language that is tailored in a number of ways, both in syntax and internal operation, for the development of computationally-demanding responsive multimedia applications. The environment is designed to support various levels of both authoring abstraction and user expertise. Within the same language base, creators may, on one hand, script complex image compositions and behaviors using high-level constructs or, on the other hand, write drivers for in-house sensors and hack low-level system operations. The small yet complete syntax lessens the burden on novices and still allows experienced programmers to take full advantage of their skills. In this respect, the language is well-suited for large collaborative projects.



Figure 6: *Live video feeds or pre-recorded motion-JPEG sequences with synchronized audio may be displayed in the background.*

Having Reflection of Presence scripted entirely in Isis enables a great deal of freedom to experiment with the behavior of the system and see results quickly. Once, just a few hours before an important demonstration, we wanted to add text “labels” that would appear below participants when they spoke to serve as an aid in recalling names. We were able to implement this behavior in a short amount of time simply by adding one additional object to the scene that follows each participant and appears only when the participant is the center of attention. Isis also provides an efficient means for incorporating functions written in C into the core of functionality available in the language. For example, new segmentation algorithms can be prototyped and tested using Isis’s installed image processing routines as building blocks and, once perfected, written directly in C to improve their performance.

5 Future directions

Since we have a high-bandwidth ATM network in our laboratory, we commonly demonstrate Reflection of Presence using full-motion full-color video for all participants, but this is not a necessity. If, for instance, one client lacks video capability due to hardware or bandwidth limitations, then a still image “avatar” of that participant could be shown to the video-equipped sites. This strategy could also be used if a participant needs to use a telephone to enter the space from a remote location where no processing hardware is available. If a client also lacks audio capability, then text or synthesized speech could be rendered. In addition, in the current version, all participants view the same video stream—there is no personalization of the video output for each client. However, experiments are underway to tailor each participant’s video and audio output to correspond to local processing and display capabilities and user preferences.

We are also exploring alternative scenarios in which the system could be applied besides a simple tele-meeting environment. For example, how might the system be used for artistic collaboration by creating shared painting, dancing, or theater spaces? What new forms of expression or storytelling would be enabled? How could the environment be used to collaboratively navigate large databases of information or catalogs of merchandise? How might children play and learn in the space? What new interactive gaming opportunities are afforded by the system? Our focus has been on collaborative environments, but how would the space be altered if the context was competitive, as in a betting or stock trading situation?

Having all participants appear in the same frame instead of as “talking heads” in separate video windows may have powerful psychological effects that we are only beginning to discover and understand. What effect does inhabiting a space together as overlapping image layers have on a sense of community or cohesiveness? What effect does the ability to see yourself in reflection have on the way you communicate with the others?

In what other ways could the relations between individuals and groups be visually portrayed in such a space? These questions will be the subject of further study as the development of the system progresses.

6 Acknowledgments

The authors wish to thank the many individuals who have contributed to this project including John Watlington, Christian Baekkelund, and Ann Bui. This research has been supported by the Digital Life Consortium at the MIT Media Laboratory.

References

- [1] V. M. Bove, Jr., "Object-Oriented Television," *SMPTE Journal*, 104, Dec. 1995, pp. 803-807.
- [2] J. Hollan and S. Stornetta, "Beyond Being There," *Proc. CHI*, 1992, pp. 119-125.
- [3] W. A. S. Buxton, "Telepresence: Integrating Shared Task and Person Spaces," *Proc. Graphics Interface*, 1992, pp. 123-129.
- [4] R. C. Edson, *et al.*, "Video Conference System using Voice-Switched Cameras," U. S. Patent 3,601,530, 1971.
- [5] D. G. Boyer and M. E. Lukacs, "The Personal Presence System: a Wide-Area Network Resource for the Real-Time Composition of Multipoint Multimedia Communications," *Multimedia Systems*, 4, 1996, pp. 122-130.
- [6] J. Han and B. Smith, "CU-SeeMe VR Immersive Desktop Teleconferencing," *Proc. ACM Multimedia '96*, ACM, New York, 1996, pp. 199-207.
- [7] H. Ishii, M. Kobayashi, and K. Arita, "Iterative Design of Seamless Collaboration Media," *Communications of the ACM*, 37, August 1994, pp. 83-97.
- [8] S. Agamanolis and V. M. Bove, Jr., "Experiences with Multi-Level Scripting for Responsive Multimedia," *IEEE Multimedia*, Winter 1997, (in press).