

AccessBot: an Enabling Technology for Telepresence

Jason Leigh (spiff@evl.uic.edu)
Maggie Rawlings, Javier Girado, Greg Dawe, Ray Fang,
Alan Verlo, Muhammad-Ali Khan, Alan Cruz,
Dana Plepys, Daniel J. Sandin, Thomas A. DeFanti

Electronic Visualization Laboratory,
University of Illinois at Chicago
USA

Abstract

The goal of the AccessBot™ project is to provide a new form of access for the disabled that integrates teleconferencing with life-sized display screens, robotics and high speed networking, to create a virtual presence for the handicapped participant at meetings. The use of a life-sized display and high-fidelity video and audio conferencing rather than existing conference room meeting systems is to ensure that the handicapped participant commands as equal a presence as the regular participants in the meeting. The use of a zoom/pan/tilt camera empowers the handicapped participant with capabilities beyond what is "humanly" possible- giving them a "bionic" eye with which they can see far greater distances.

This paper will describe the implementation of the AccessBot and the lessons learned from its deployment at the Supercomputing 1998 conference in Orlando, Florida; and at the National Center for Supercomputing Applications' ACCESS Center in Washington D.C.

Introduction

The problem of providing accessibility for the disabled goes far beyond merely providing ramps and elevators in buildings, cars, airplanes, and buses. There are in fact people, such as the aged, or those afflicted by Multiple Sclerosis (M.S.), who simply do not have the strength or health to travel even when physical access is available. The goal of the AccessBot project is to provide a new form of access by utilizing high-fidelity video conferencing and remote control to provide a telegraphed presence (or telepresence) that is as commanding as actual physical presence.

A number of difficult problems arise when attempting to achieve this seemingly simple goal. From a human-factors standpoint, the user-interface must be readable and easy to use. Since hand-to-eye coordination can often be a problem for M.S. sufferers or the aged, the interface should not require fine motor control to operate. The quality of the audio and video must be high enough so that motion artifacts due to data loss or rapid gesturing of the user, are minimized. Latency of the audio and video must be low enough so that delays between the user and the audience are imperceptible. Finally the system must be easily deployable so that it can be shipped to a remote site and activated with a minimum amount of configuration and expert intervention.

This paper describes a prototype of the AccessBot that we have built in an attempt to address some of these issues.

Implementation

The AccessBot system as a whole is comprised of two units- the remote station and the home station. Figure 1 shows the remote station. The remote station is the *actual* AccessBot. It is the device that is actually transported and deployed at distant meetings. The home station (shown in Figure 1) is the endpoint that the user uses to control the AccessBot in order to remotely participate in the meeting.

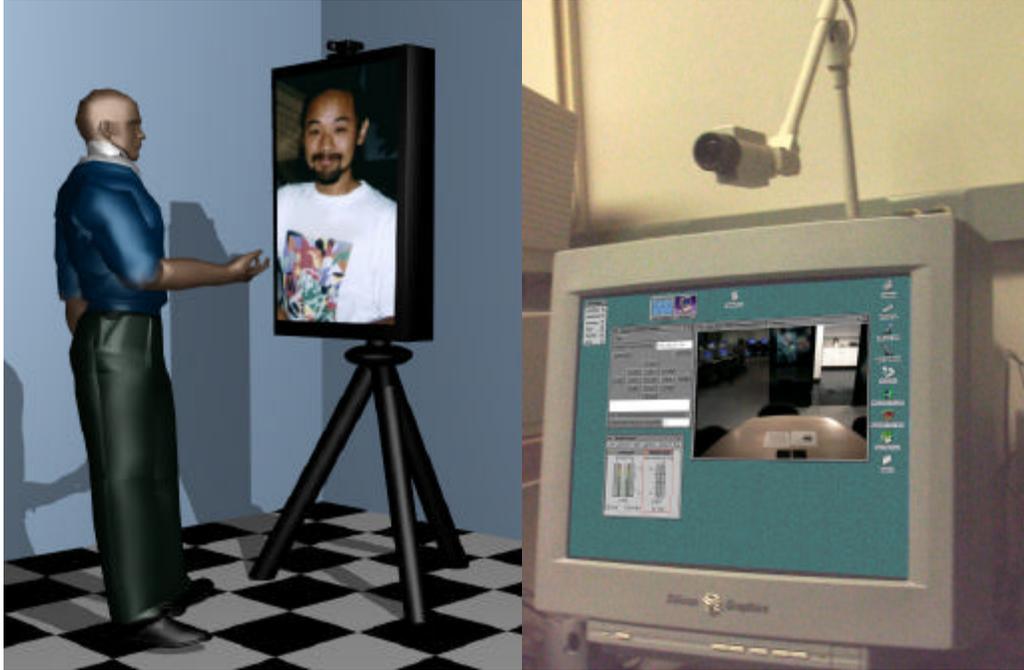


Figure 1: The AccessBot (left) and its home station (right.) The video image on the home station can itself be routed to a large display for convenient viewing. The camera at the home station can be tilted at 90 degrees to match the orientation of the AccessBot's plasma panel.

To exemplify presence on the AccessBot, a large 40-inch plasma screen is tilted at 90 degrees and mounted on a tripod so as to depict the remote participant in a life-sized scale. A remotely controlled auto focus, pan-and-tilt camera is mounted on top of the plasma screen. This camera serves two functions; firstly it allows the viewer at the home station to look around the room under his/her-control. Secondly the zoom lens on the camera allows the user to read notices on the walls of the meeting room or focus-in on a document on a table. The zoom lens hence empowers the user with capabilities beyond what is "humanly" possible- giving them a "bionic" eye with which they can see far greater distances.

Pairs of Silicon Graphics O2s provide audio and video streaming on the AccessBot and the home station. PC-based systems were evaluated however none were able to deliver non-interlaced NTSC (640x480) resolution video at 30 frames per second while also incurring a minimum latency in the video encoding and decoding process. PC-based systems typically encoded video at 320x200 resolution at 15 frames per second. The O2s were ideally suited to the task because they possessed JPEG compression hardware that was fast enough to encode and decode JPEG images at 30 frames per second. MPEG encoding schemes were not chosen because of the need for MPEG to accumulate a sequence of frames before efficient encoding can be performed.

Since we are displaying video at 30 frames per second, each video is captured at 33ms intervals. Hence the end-to-end delay between capturing a frame, delivering it, and displaying it is approximately: $33 + \text{network_delay} + 33$ ms. Previous experiments in video conferencing have shown that the threshold at which a delay is detectable over a long distance phone call is approximately 100ms. Since the encoding and decoding delays already total 66ms there is little time left to accumulate additional frames as required by MPEG without suffering a perceivable lag in the conferencing experience.

In the majority of existing video conferencing systems only the regions of the screens that change are updated. Furthermore the corresponding regions of the screen are updated with whatever data packets that arrive. This scheme is used because it can significantly reduce the bandwidth requirements of the teleconference so that video telephony can be brought over the congested Internet. This scheme however suffers a major drawback. There is no real regard for the overall coherence of the image; when the subject in front of the camera moves too rapidly or if the image composition between successive frames is vastly

Leigh, J. et al. **AccessBot: an Enabling Technology for Video Conferencing**, in CDROM proc. INET 2000, The 10th Annual Internet Society Conference, 18-21 July 2000, Pacifico Yokohama Conference Center Yokohama, Japan.

different, the image will appear to disintegrate into a cloud of square particles. We have found this image disintegration to be distracting and hence in our approach we display all of a frame or none of it. That is, if packets are lost then the entire frame is discarded hence preserving the clarity of each image. Since the focus of our work is video telephony over *high-speed* networks we were not compelled to develop techniques for congested networks.

Figure 2 illustrates the connections between the O2s (Bonnie and Clyde) at the AccessBot and at the home station (Butch and Sundance). At each endpoint one O2 is dedicated to video and audio encoding and the other is dedicated to decoding. All O2s are connected either via 100Mbit Ethernet or an ATM adapter. The user at the home station would operate a point and click graphical interface on Butch to control the AccessBot camera. This interface would then send camera commands as UDP packets to Bonnie. Bonnie would then route these commands via its serial port to the camera. Figure 1 shows a snapshot of the home site. Notice the camera at the home site is oriented at 90 degrees to allow the image of the user's head and torso to fill the vertical length of the plasma display's screen. The camera on the AccessBot however was kept in its standard horizontal orientation to give the viewer at the home site a regular landscape view of the conference room.

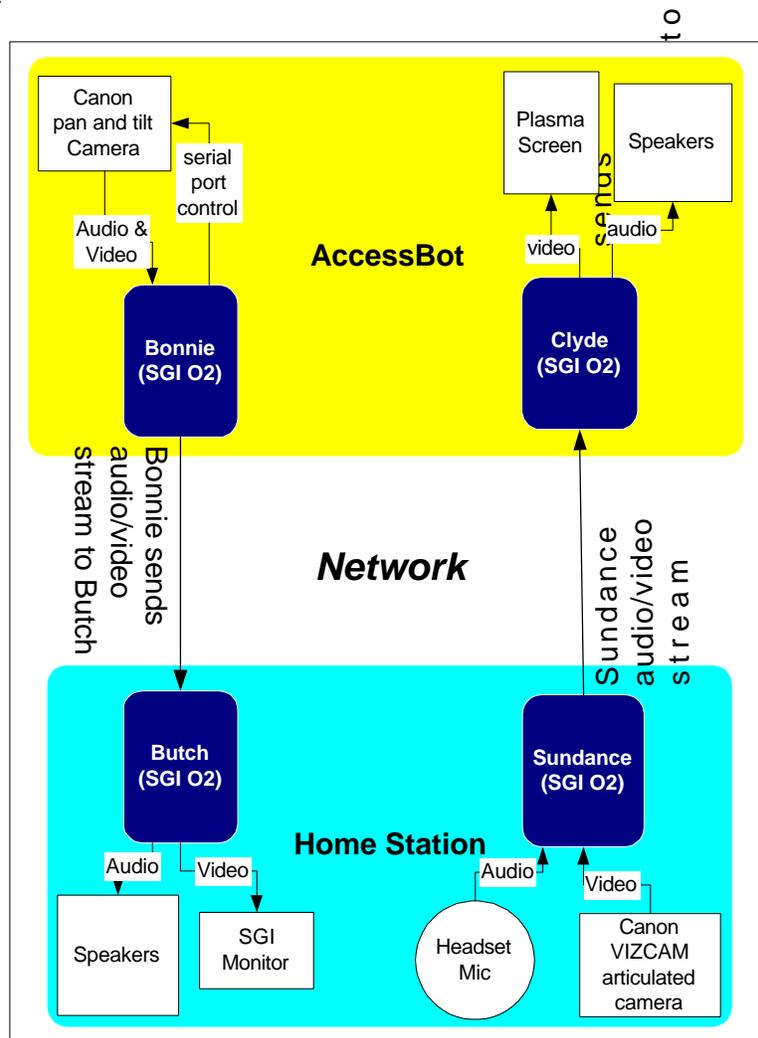


Figure 2: AccessBot connection schematic.

The AccessBot's custom video streaming software (vc2way) is designed to take advantage of the O2s hardware JPEG compression/decompression capabilities, and where possible, minimize the latency between image capture and image delivery. Vc2way was originally designed to create video avatars in teleimmersive environments (3D collaborative, immersive environments)[1,2]. The camera control

Leigh, J. et al. **AccessBot: an Enabling Technology for Video Conferencing**, in CDROM proc. INET 2000, The 10th Annual Internet Society Conference, 18-21 July 2000, Pacifico Yokohama Conference Center Yokohama, Japan.

software was written in Java so that it can easily be ported to PCs, should PC video conferencing technology improve in the future. The camera control software allowed the user to pan, tilt and zoom the camera at incremental angles or directly to specific orientations. The rate at which the panning, tilting or zooming occurred could be controlled by a speed slider. Figure 5 shows the home station user-interface.

Deployment and Lessons Learned

The first AccessBot was demonstrated at the Supercomputing 1998 conference in Orlando, Florida (Figure 3). This prototype was driven by only a single O2, which performed both the video encoding and decoding task. We found that a single O2 loaded the system too heavily for the separate camera-controlling interface to work; hence a separate PC-based laptop was used for remote camera control. At the conference audio and video were streamed between Chicago and Orlando over the vBNS (a formerly National Science Foundation funded, OC12 network). We noticed at the conference that the large vertical plasma screen caught the attention of many passers-by. This is typically not the case for regular teleconferencing systems.



Figure 3: Photograph of the AccessBot at Supercomputing 1998 in Orlando, Florida.



Figure 4: Photograph of the AccessBot at the National Center for Supercomputing Applications' ACCESS Center in Washington D.C. (Photograph courtesy of Tom Coffin- NCSA)

Currently an AccessBot has been deployed at the National Center for Supercomputing Applications' (NCSA) ACCESS Center in Washington D.C (Figure 4). This allows us to evaluate the technology in terms of bandwidth utilization and usability.

Bandwidth Utilization

Table 1 shows the theoretical bandwidth requirements as a function of Jpeg compression quality. Through trial and error we found that a Jpeg quality of Q=75 yielded the best results. At Q=75 we found that on average a unidirectional video stream consumed approximately 4Mbps of bandwidth. We did not experiment with bandwidth requirements at low Q levels since at those levels image quality was too poor for useful video conferencing. At higher Q levels (Q=100) the O2's internal networking buffers overflowed. Although this problem could be solved by allocating more buffer space, this was intentionally not done to minimize the latency between video encoding and transmission.

	Q=20	Jpeg Quality Q=75	Q=100
Compressed Frame Size	10 K Bytes	12-15 K Bytes	25 K Bytes
Theoretical 1-way Bandwidth at 30fps	2.5 Mbps	2.8-3.6 Mbps	6 Mbps
Experimental 1-way Bandwidth at 30fps		4 Mbps	
Total Recommended Bandwidth for 2-way (LAN)		8Mbps	
Total Recommended Bandwidth for 2-way (WAN)		14Mbps	

Table 1: Theoretical and experimental bandwidth utilization of the AccessBot vc2way software.

Usability Issues

What is immediately apparent when someone appears on the AccessBot screen is that their life-sized image does indeed command a presence that does not usually occur from viewing small video conferencing windows. This is important because although the AccessBot's technology was initially targeted for the disabled, it is clear that presence is extremely important in a business meeting context regardless of whether one is disabled or not.

Leigh, J. et al. **AccessBot: an Enabling Technology for Video Conferencing**, in CDROM proc. INET 2000, The 10th Annual Internet Society Conference, 18-21 July 2000, Pacifico Yokohama Conference Center Yokohama, Japan.

The graphical interface (Figure 5) for controlling the remote camera was tested by one of the members of our team who is currently suffering from M.S. She found it unusable because the point and click windowing interface required fine motor control to operate. As she gradually lost control in the later stages of her M.S. it became increasingly difficult to use the interface. We are now investigating the possibility of either using a large joystick or a touch screen to control the camera.

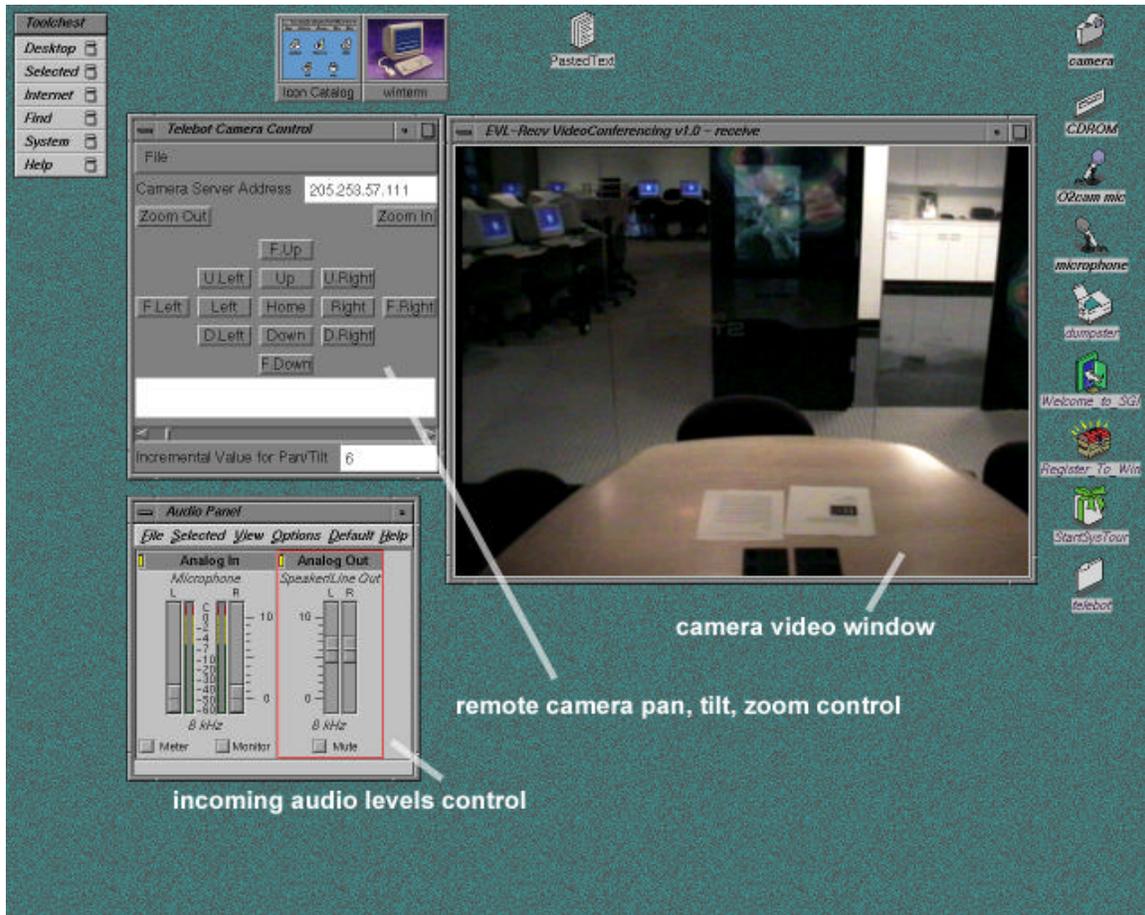


Figure 5: Graphical user interface for the home station.

As noted earlier the end-to-end lag of the system is in excess of $66\text{ms} + \text{network_delay}$. When controlling a remote camera, an additional network delay is introduced because there is also the time needed to propagate the command to the remote camera. In practice we noticed that this lag made operation of the remote camera difficult. Past work in telerobotics has shown that latency in the feedback for such tasks should ideally be less than 125ms [3,4]. In our home station interface, users adopted a *move-and-wait* strategy to control the camera. They tended to make smaller camera movements to prevent themselves from overshooting their target. However they were frustrated when it took them a long time to move the camera over larger distances. This problem may be minimized if the interface were re-designed to allow them to employ a *point-and-go* strategy instead. That is, the user could point at the area of the video image where they wanted to look and then the system would automatically compute and adjust the orientation of the camera accordingly.

Leigh, J. et al. **AccessBot: an Enabling Technology for Video Conferencing**, in CDROM proc. INET 2000, The 10th Annual Internet Society Conference, 18-21 July 2000, Pacifico Yokohama Conference Center Yokohama, Japan.



Figure 6: Alternative sketch of the home station interface.

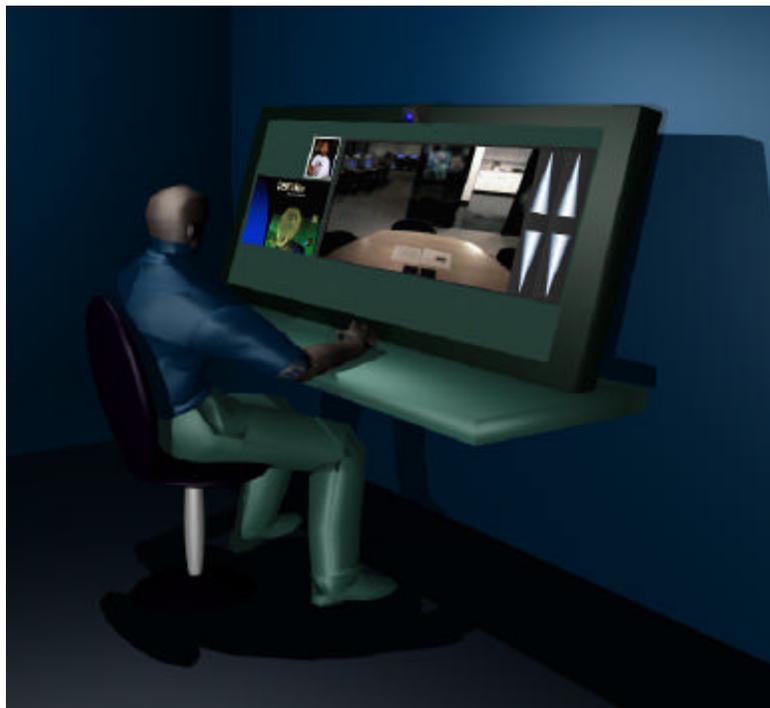


Figure 7: Concept sketch of an alternative home station design using a large 50'' plasma screen with a touch membrane mounted over the display surface.

Figures 6 and 7 combine an alternative user interface design with a touch-screen mounted on a 50'' plasma display. The user can physically touch the video window to indicate the location to direct the camera. The large zoom controls on the side can be used to control absolute zoom. The small video portrait inset at the top left hand corner of the screen is used to give the user feedback of what is being sent to the remote AccessBot's screen. One problem with this new design is that the camera on the home station would be

Leigh, J. et al. **AccessBot: an Enabling Technology for Video Conferencing**, in CDROM proc. INET 2000, The 10th Annual Internet Society Conference, 18-21 July 2000, Pacifico Yokohama Conference Center Yokohama, Japan.

mounted too high (because of the size of the plasma display.) Users would no longer be able to maintain proper eye contact. Acker and Levitt [5] have shown that improved eye contact increases satisfaction with videoconferencing as a medium for negotiation. It is not clear whether past solutions using beam-splitters [6] would be viable in this situation. Alternatively one could mount cameras at each side of the AccessBot and use image-based rendering algorithms to interpolate the correct front view of the subject- but currently such algorithms are not able to operate without significant lag.

In our first prototype of the AccessBot an echo-canceller was not used. The headset microphone that the user at the home site used provided effective echo-cancellation so that the audience at the remote site was able to hear the user clearly. However the echo that occurred from the audio at the remote site feeding back into the remote control camera was distracting. The solution to this would be to build-in an echo canceller into the AccessBot.

Finally, we are far from having reached a point where the AccessBot can be easily deployed and configured. Currently the AccessBot is shipped in a "road case" and requires two people to lift and set it on its tripod. Furthermore all the connections between the O2 and the various peripherals (plasma screen, pan-and-tilt camera, etc) and networking have to be configured manually. This meant that an experienced crew had to be there to configure it. What is desired is a self-deploying structure in which all the hardware connections are maintained; and where the only software configuration would be the networking IP addresses. The remainder (such as JPEG compression rate, image transfer rate etc) should be configurable remotely from the home site. Ideally these parameters should be configured automatically based on the AccessBot's sensing of the underlying network connectivity between the two sites.

Conclusions and Future Enhancements

The AccessBot provides much in the way of projecting a remote user's presence into a distant meeting. However much work needs to be done to provide the user with a sense of being an intimate part of the meeting. Hence there are many possible improvements that can be made in the future. The following enumerates just a few of them:

1. Provide a large plasma display screen at the home station to improve the sense of presence in the meeting for the user.
2. Explore single computer solutions- preferably a multiprocessor PC with a high-end video conferencing board capable of streaming video at NTSC resolution and 30 frames per second with a minimum of latency.
3. Include a small projector from which remote Powerpoint presentations can be controlled.
4. Use lighter weight LCD displays rather than heavy plasma screens; or use projectors to rear-project onto a translucent screen. This may require mirrors to shorten the throw distance of the projectors.
5. Improve the home station user-interface to allow direct tracking of camera based on the region of video selected. This can help greatly reduce the problem of trying to control a camera when the network latency is high.
6. Explore emerging high-bandwidth wireless technologies to further enable ubiquitous deployment.
7. Explore forward error correction schemes to compensate for packet losses during video transmission and hence improving video quality.
8. Improve the manner in which the AccessBot is deployed so that it is able to configure itself for optimum frame rate and video quality.
9. Build a deployment structure that can house all the AccessBot equipment so that it is transported with all its connections in tact. Figure 8 shows an early conceptual sketch of a self-deploying structure.



Figure 8: Concept sketch for a self-deploying AccessBot (Image courtesy of Alan Cruz and Mohammad-Ali Khan).

Acknowledgments

We would like to thank Tom Coffin and J. J. Jamison for assisting in the deployment of the AccessBot at the ACCESS Center in Washington D.C. We would like to thank Maggie Rawlings who pioneered this project while also combating Multiple Sclerosis.

The research, collaborations, and outreach programs at the Electronic Visualization Laboratory (EVL) at the University of Illinois at Chicago are made possible by major funding from the National Science Foundation (NSF), awards EIA 9720351, EIA-9802090, EIA-9871058, ANI-9712283, ANI-9730202, and ACI-9418068, as well as NSF Partnerships for Advanced Computational Infrastructure (PACI) cooperative agreement ACI-9619019 to the National Computational Science Alliance. EVL also receives major funding from the US Department of Energy (DOE), awards 99ER25388 and 99ER25405, as well as support from the DOE's Accelerated Strategic Computing Initiative (ASCI) Data and Visualization Corridor program. In addition, EVL receives funding from Pacific Interface on behalf of NTT Optical Network Systems Laboratory in Japan.

The AccessBot™ is a trademark of the Board of Trustees of the University of Illinois.

References

- [1] Wang, F., Video Avatars in Collaborative Virtual Environments, M.S. Thesis, University of Illinois at Chicago, 1998.
- [2] Leigh, J., Johnson, A. E., Brown, M., Sandin, D., DeFanti, T. A., Visualization in Teleimmersive Environments, *IEEE Computer*, pp. 66-73, Dec. 1999.
- [3] Sheridan, T. B., Space Teleoperation through Time Delay: Review and Prognosis, *IEEE Transactions on Robotics and Automation*, vol. 9, no. 5, October 1993, pp. 592-606.

Leigh, J. et al. **AccessBot: an Enabling Technology for Video Conferencing**, in CDROM proc. INET 2000, The 10th Annual Internet Society Conference, 18-21 July 2000, Pacifico Yokohama Conference Center Yokohama, Japan.

[4] Hannaford, B., Ground Experiments Toward Space Teleoperation with Time Delay, Teleoperation and Robotics in Space, Chapter 4, 1994, pp. 87-106.

[5] Acker, S., and Levitt, S. Designing videoconference facilities for improved eye contact. *Journal of Broadcasting and Electronic Media* 31, #2, pp.181-191, 1987.

[6] Abel, M., et al. Telecollaboration Research Project, in Computer Augmented Teamwork by Robert Bostrom, Richard Watson, Susan Kinney, pp. 126-138, Van Nostrand Reinhold, New York, 1992.