HAVE'2006 - IEEE International Workshop on
Haptic Audio Visual Environments and their Applications
Ottawa, Canada 4-5 November 2006

Hapto-Audio-Visual Environments for Collaborative Training of Ophthalmic Surgery Over Optical Network

P. Boulanger, G. Wu, W. F. Bischof, and X. D. Yang
Department of Computing Science University of Alberta,
Edmonton, Alberta, T6G 2E8, Canada

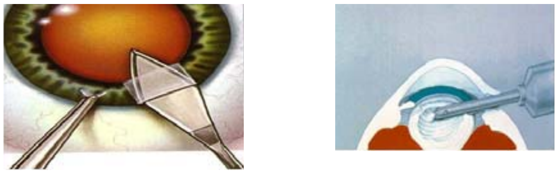
Abstract *- This paper presents the results of a two-year project to develop a shared hapto-visual-audio-virtual environment (HAVE) with advanced multi-point video conferencing, new display and interface technologies, and distributed latency-compensated haptic technologies for collaborative medical research and training in ophthalmology. One of the goals of this project is to create collaborative training environment, in which residents can remotely learn, in real-time, cataract operations from real operations performed by teaching surgeons. The assumption of this work is that a trainee surgeon can learn the complex hand-eye coordination necessary for becoming a good ophthalmic surgeon by feeling and seeing every move the expert surgeon makes, through a complex haptic, auditory, and visual playback interface. Experimental results are presented.*

Keywords *- Medical Trainer, Haptic Interface, Stereo Display, Remote Training, Tele-immersion*

I. INTRODUCTION

This paper presents the results of a two-year project to develop a shared hapto-visual-audio-virtual environment (HAVE) with advanced multi-point video conferencing, new display and interface technologies, and distributed latency-compensated haptic technologies for collaborative medical research and training in ophthalmology. The need to train ophthalmic surgeons comes from the fact that there is a large demand for such operations. Cataracts are a leading cause of blindness in the world. More than 1.3 million cataract procedures are performed each year in the U.S. alone, and visual disabilities associated with cataracts account for 8 million of physician visits each year.

Cataract surgery is typically an outpatient procedure that takes less than an hour. Most people are awake and need only local anesthesia. On rare occasions, people may need general anesthesia, for example, if they have difficulties laying flat or if they have claustrophobia. Two things happen during cataract surgery, first, the clouded lens is removed, and second, a clear artificial lens is implanted. More specifically the surgeon starts by making a small incision where the cornea meets the conjunctiva (see Figure 1a). The surgeon then uses a probe, which vibrates with ultra-sound waves, to break up (emulsify) the cataract and suck out the fragments (see Figure 1b). Once the cataract is removed, a clear artificial lens is implanted to replace the original clouded lens (see Figure 1c). This lens implant is made of plastic, acrylic or silicone and becomes a permanent part of the eye. Sometimes a suture point is added (see Figure 1d).


(a) First incision (b) Removal of the cataract

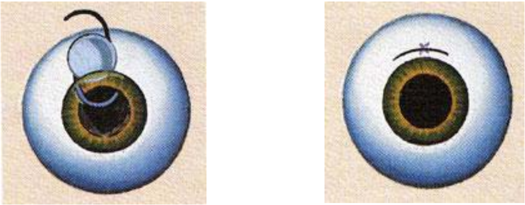

(c) Insertion of the lens (d) Suture

Figure 1: Basic steps of cataract surgery.

Although cataract surgery appears to be a simple procedure, it is in fact a complex one that requires extensive training to avoid complications such as:

- Vitreous Hemorrhage (0.3% of procedures);
- Uveitis (1.8%);
- Increased Eye Pressure (1.2%);
- Retinal Detachment (0.7%);
- Endophthalmitis (0.13%).

One of the problems with surgeons located in remote regions is that, similar to musicians, the lack of day-to-day practice may reduce the ability to perform these tasks well, hence many surgeons are reluctant to move away from large urban centers.

One of the goals of this project is to create a collaborative training environment in which residents can remotely learn, in real time, cataract operations from real operations performed by training surgeons. Up to now, many trainee surgeon had to practice on an analog replacement, such as a pig eye or a virtual simulator, where a program tries to recreate a similar experience through haptic and visual interfaces. In both cases, it is very hard to capture the subtle moves that a training surgeon makes during an operation.

1-4244-0761-3/06/$20.00 ©2006 IEEE

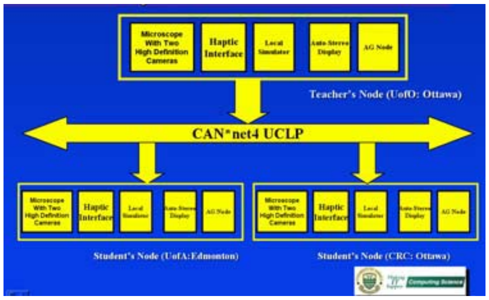
These subtle moves often make the difference between a good surgeon and a novice. By recording every subtle move, voice intonations, and visuals of well-trained surgeon during an operation, it is possible for junior residents to be trained by remote experts in real-time over a high-speed network, or after-the-fact in playback mode through a multi-modal interface capable of replicating those moves. It has been shown in numerous psychophysical experiments that the brain can lean complex hand-eye tasks by imitating experts. The thesis of this work is that a trainee surgeon can learn the complex hand-eye coordination necessary for becoming a good ophthalmic surgeon by feeling and seeing every move the expert surgeon makes, through a complex haptic, auditory, and visual playback interface. This creates in effect a new tele-immersive training system that not only allows the trainee to feel and see the training surgeon, but also permits the trainer surgeon to feel and see how the trainee performs during an operation, allowing him to give suggestions along the way. Figure 2 shows two block diagrams describing how the system works. Figure 2a illustrates the so-called one-way teaching mode, where every move of the teaching doctors is transmitted via a high-speed optical link to two students. In this mode, the students only require a multi-modal playback interface to participate. The only mode of interaction with the teacher is through questioning using video conferencing such as Access Grid. If there is also a digitizing station at the student station, as illustrated in Figure 2b, it is possible for the student to start operating and to transmit to the other students as well as to the training surgeon every moves he makes. The trainer can then feel and see those moves and offer corrections and suggestions by voice or through the haptic device, giving the trainee in essence a tele-immersive nudge. In section II, we review the current state of the art of medical trainer in the light of the proposed system. In section III, we describe the digitizing station and in Section IV the multi-modal playback interface. We conclude by describing the current state of our system and our plan for usability testing.

II. PREVIOUS WORK

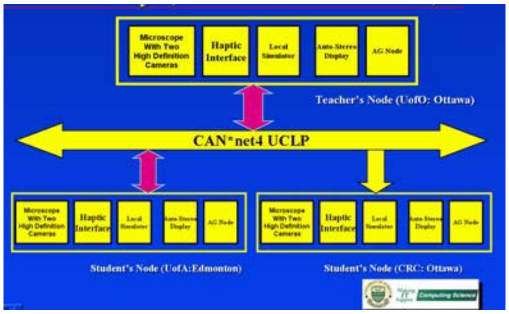
The literature on surgical trainers is vast and covers many fields ranging from engineering over medicine to psychophysics. In this section, we cover first the current state of the-art by reviewing the newest multi-modal trainers. We then review the latest literature on how humans learn fine motor activities and how those skills can be transferred. Finally, we discuss how the proposed system compares to other remote surgical trainers.

A. Multi-modal Surgical Trainers

HapticIO [1] is a Minimally Invasive Surgery (MIS) multi-modal trainer. In this system an interface toolbox provides to the trainee with two (up to three) haptic devices with a shape of scissors handles and a virtual "video-endoscope" device. The trainee uses the haptic devices to control the virtual tools in the virtual environment. The system creates an approximate force feedback, allowing the trainee to interact with the virtual organs in the scene. The virtual video-endoscope device allows the trainees to move their view in the virtual environment as if they were controlling a real endoscope camera. As in real life, the corresponding scene changes can be seen from an attached LCD monitor. The system uses medical image data (CT, MRI) to create realistic 3D models of the organs. The data is visualized using real-time volume rendering techniques. In addition to MIS training, HapticIO can also be configured for the training of ear-nose-throat (ENT) surgery. In this case, a simple PHANTOM device from Sensable Tech. Inc., USA is used. HapticIO is now a commercially available product.



(a) One way teaching mode



(b) Monitoring mode

Figure 2: Two mode of operation of tele-immersive training system.

Webster and his colleagues developed a suturing simulator [2] based on haptic force-feedback and a stereo display to improve user performance in terms of accuracy and speed. The system has a PHANTOM device serving as the main haptic user interface. Hemostats are placed at the PHANTOM end effectors to provide a realistic feeling of holding the instruments. Manipulation of 3D virtual needles and needle holders is performed through the hemostats. The user can feel the force feedback when suturing the soft tissue in the virtual environment. A "Reaching display" is used as a display, where the user holds the hemostats under the mirror, which reflects a stereo image of the mounted monitor. This configuration is believed to give users a better haptic feel. In order to measure user performance, a scoring test is used to evaluate user satisfaction. Task completion time is also recorded. Unfortunately, no user study was conducted to show that skill transfer actually takes place with this system.

Abolmaesumi and his colleagues developed a haptic-based medical image examination system [3] for diagnosis training. Here the trainee controls a virtual probe through a PHANTOM device. A section of a human body created by CT and MRI is displayed, using a regular monitor, with the probe in a virtual environment. The virtual probe creates a cutting plane that can be moved through the body. The volumetric data image is then sliced along the plane and displayed. Force feedback is provided once the virtual probe touches the surface of the body. In this work, no results were reported to demonstrate the performance of the system.

The vitreous surgery trainer [18] from Asahikawas Medical College, Japan, is an ophthalmology surgery simulator. The system is equipped with virtual high-resolution color stereo binoculars, two 6DOF haptic positioning devices and foot switches. Several virtual patient eyes with various retinal diseases were created to illustrate realistic conditions. The virtual binoculars cameras provide two images of a virtual eye simulating the stereo display surgeons use in real operating conditions. Foot pedals are used to switch instruments and control the view of the virtual environment. The system simulates the removal of proliferative tissues on the retina. No evidence has been provided that the system is actually used by any institute, and no usability study was performed.

B. Haptic Training for Motor Skill Transfer

Human learns novel motor skills through observation and practice. Recent studies have shown that haptic assistance can help human improve motor skill learning process [5]. Govindarajan [6] developed a haptic handwriting training system where an expert's handwriting was recorded in terms of temporal position, velocity and force information and play back to a trainee. A PHANTOM device was used as the haptic interface to provide force feedback and position information. Trainees were trained under different conditions, no assistance, position control, and haptic profile control. After the training, trainees were asked to reproduce the characters they had learned. Results show that the subjects under haptic training produced better shape characters compared to those who trained without haptic assistance. Wang [7] used haptic feedback to teach people how to write Chinese characters. A similar "record-and-play" method was used in his system. The results show that subjects trained under haptic assistance could write Chinese better and faster than those who did not experience haptic training.

Feygin [8] investigated the role of haptics in the training of human motor skills. He introduced the haptic guidance paradigm, in which the trainee is physically guided through a desired motion by a haptic device. Using this system, the trainee, was able to rapidly gain understanding of kinesthetic requirements. In this study, subjects learned a novel 3D motion under three training conditions, vision-only, haptic-only, and haptic + vision. After the training, the subjects were asked to reproduce the movement under two recall conditions – vision only and haptic only. Task performance was measured in terms of temporal position, shape, completion time and drift. Results show that haptic guidance mainly helps to improve the temporal aspect of the task, while visual information is better for learning the trajectory shape. Following Feygin's study, Liu [9] found that haptic guidance had only a short-term effect. He also stated that subjects tended to quickly forget what had been learned once haptic guidance was withheld, hence there was a need for frequent repetitions.

C. Remote Surgical Training

Most of the medical simulators provide skill transfer through the trainee's practice on the systems. As a result, much more attention has been paid to the interaction between human and virtual world. Haptic has been used to enhance this interaction. However, training by practice can only be called self-teaching. There is no doubt that teachers play an important role in the training process. Incorporating the interaction between instructor and student into surgical simulators is expected to have a significant impact on existing systems. Surgical training systems integrated with human-human interaction are usually running in a Shared Virtual Environment (SVM). In SVM, participants perform tasks in a collaborative fashion, and haptic feedback is used to preserve fidelity and to improve task performance [10][11].

Gunn developed a networked surgical training system [12][13] for human-human interaction during surgical training. The system runs on two CSIRO haptic + stereo workbenches. The instructor and the student communicate verbally through head sets. They can see, touch and interact with the same virtual organ in the scene. In the haptic guidance mode, the instructor can hold the end of the student's tool and physically guide him through an ideal motion. During a training session both of them can feel each other's resistance. The system also allows them to perform a task collaboratively, such as dissecting soft tissue together. Video clips of real operations can be played to give the student the appearance of a real organ. Gunn conducted numerous tests on this system [14]. He placed an instructor in the U.S. to use the system to teach students, all located in Australia, for a gall bladder removal operation. The system was running over a broadband Internet connection. A group of 70 medical professionals audited the demonstration in Australia. After the test, the audience was asked to vote on the usability of the system. Positive feedbacks were given for the system performance in terms of interaction between participants, remote teaching, haptic interaction and haptic guidance.

Matthew described the design and the implementation of a similar system [15][16]. He also conducted an analysis on the verbal communication between instructor and student during a training session. The system was equipped with two PHANTOM devices to support ambidextrous operations. A haptic guidance tool was also supported, CT scans and real pictures of organs were used to create realistic models. The communication mechanism also provided the participants with the ability to draw anywhere in the scene as well as on a shared whiteboard. The system was used to teach bone surgery. Two workbenches were placed together so that the participants could talk to each other without the aid of microphone. The verbal conversation was recorded for analysis. Unfortunately, the instructor did not use the haptic guidance tool during the entire trial. The author stated that this might be due to the efficiency of the other communication modes. Other trials [17] were conducted on the same system to measure the acceptability of the system. The system was demonstrated to a group of professionals. Questionnaires on the demonstration provided very positive feedback.

To our knowledge, there are no systems reported in the literature that do not use some sort of virtual simulation to represent the real world. In contrast, our system has no need for a simulated world because all stimuli are recorded, and they are accurate since they represent real-life operations.

III. TRAINER DIGITIZING STATION

In order to digitize a surgery, the following basic information needs to be recorded

- Sound
- High-magnification stereo views of the eye
- Two large-scale views of body posture
- Location and orientation of the tools
- Forces applied to the tools by the surgeon

Figure 2 shows an illustration of the sensing modalities in a normal operating room. Because of the type of application, the following constraints must be respected:

- The digitized modalities must be perfectly synchronized
- Compression should be minimal to avoid misinterpretation
- The digitizing equipment must be operating-room compatible

In this section, we describe the basic elements of the prototype digitizing station (see Figure 3). It is composed of five basic building blocks:

- A calibrated stereo digitizing video server
- A calibrated optical tracking system
- An auto-stereo display system
- A video conferencing system based on Access Grid
- A high-speed network with minimal latency based on Canarie User Controlled Light Paths (UCLP).

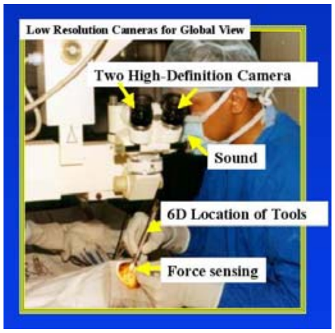


Figure 2: Sensing Modalities

Figure 3: Trainer Station

A. Calibrated Stereo Digitizing Video System

The main goal of this system is to digitize the binocular microscope used by the surgeon. The system is composed of two Basler Bayer cameras A301fcVGA operating at 30Hz and synchronized by hardware. They are mounted over the operating area using a high-precision positioning device illustrated in Figure 3. This positioning devise allows to adjust the relative position of the stereo cameras for optimal stereo viewing conditions on an auto-stereo display (see Figure 3). One of the key conditions for the use of video signals in a medical application is that the quality of the stereo replication received by the clients should be as good as looking directly into the binocular of the microscope. This means that we are not allowed to use compression since compression algorithms could hide visual structures that are important for a surgical procedure. In special cases, one could use the JPEG2000 lossless compression algorithm if bandwidth is an issue. In addition, we also had to develop a color calibration procedure for each camera since excellent color replication is key to identifying veins and other subtle structures in the eye. In order to do so, we used a simple algorithm based on a linear least square color correction algorithm. A complete description of the algorithm can be found in Boulanger [19].

In addition, we also had to correct for lens distortions since one of the requirements for the system was the ability to perform 3D measurement on the eye, and to add to the image the ability to perform augmentation. Using the same Macbeth color calibration chart used in color calibration, we were able to develop a fully automatic calibration algorithm based on Tsai's calibration code. For more details on those algorithms refer to the paper by Boulanger [19]. This new procedure makes color calibration and geometric calibration fully automatic and robust to ambient light.

In the project, a video server was developed to deliver to the video information over a high-speed network with the high quality that is required for this application. Figure 4 shows a block diagram of this server. The video client/server is composed of a JPEG2000 compressor that transmits to the client lossless Bayer images that are reconstructed on the client side. The server also sends color calibration information as well as the geometric correction information to the client. On the client side, all this information is combined to create distortion-free color-calibrated stereo images. Figure 5 shows a block diagram of the calibrated video client.

B. 3D Multiple Tool Tracking

In order to measure the position of the tip of the medical instruments (see Figure 3), we used a Phoenix system Visualeyez 4000 real-time active 3D targets tracking system that determines the positions of the LEDs mounted on a target assembly (see Figure 6a). The Phoenix system can measure in real-time (60 Hz) 120 active LEDs at a precision of 0.015 mm. The developed target assembly is very versatile and can be mounted on various surgical instruments. It is also designed to avoid optical occlusions, one of the key problems with optical tracking. Based on a geometric calibration algorithm, we are able to determine the position and orientation of the tip of each instrument at 60 Hz. Using the Visualeyez temporal targets coding scheme, we are also able to design the tracking system to compute the positions of two tools at the same time in the field of view allowing for ambidextrous manipulations.

C. Video Conferencing Using Access Grid

One of the key elements for any medical remote teaching system is the ability to have teacher and students discuss the various surgical procedures being performed. In our prototype, we decided to us Access Grid (AG) from Argon National Laboratory [20] as our main video conferencing technology. One of the key advantages of AG, besides being free, is its ability to connect numerous participants in a meeting with minimal load on the host computer. In AG, it is also possible to have more than one camera on the participating nodes allowing for multiple view of the operating site. In our prototype, we decided to locate two cameras near the teaching surgeon. One camera is used to teach general posture with an overall view of the surgeon. The second camera is used to give a macro view of the surgeon's hands and the tabletop manipulations. AG also allows for whiteboard display and the sharing of presentations. Because of the real-time requirement of the optical tracking system the AG node on the teaching side was located on a second computer. Since we used a high-bandwidth optical network in this project, we had no issues associated with the synchronization of voice, video, and haptic manipulations.

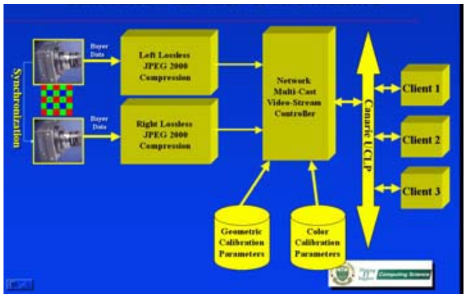


Figure 4: Calibrated Video Server.

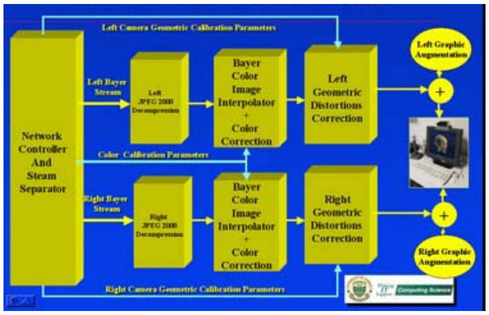
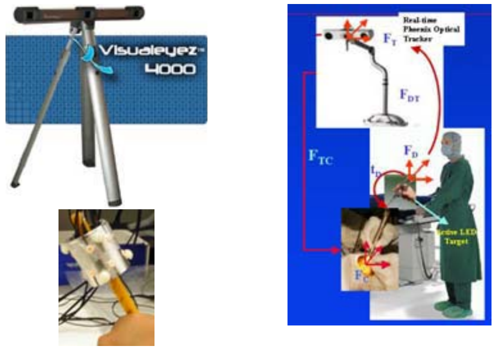


Figure 5: Calibrated Video Client.



(a) Visualeyez tracking system with target assembly

(b) Toll tracking coordinate system

Figure 6: Tool tracking system

D. Network Issues

The network performance requirements for this project are challenging. The trainer network data is mostly streaming media, namely calibrated stereo video (148 Mb/s), Access Grid (max. 10 Mb/s), and haptic signals (max. 5 Mb/s).

The end points of our system are about 4000 km, given that we need to demonstrate our system between Ottawa and Edmonton. This creates large delays. In the case of voice and video, this is not a problem, but for haptic feedback, this delay is critical especially if the delay fluctuate. If not compensated by proper delay compensation algorithms and networking this could have cause severe problem for the project. Fortunately, Canarie Inc., the main provider of optical network connectivity in Canada gave us access to a User Controlled Light Path between the two cities. Using this dedicated optical connection, we are able to achieve a transmission speed of 620Mb/s with an approximate constant delay of 55ms, which is acceptable for our application.

IV. MULTIMODAL PLAYBACK STATION

On the playback side, a multi-modal interface was built to display the various sensing modalities. Figure 7 shows the multimodal playback station. It is composed of a laptop as the main computer for controlling various rendering hardware such as two haptic interfaces and an auto-stereo display. The auto-stereo display allows for the display of the two cameras digitizing at high magnification the eye and the surgical tools. For the haptic rendering, we are using two Omni haptic interfaces to allow ambidextrous manipulations. The haptic rendering was developed using OpenHaptic. The laptop is also in charge of the tele-conferencing and the control and synchronization between the trainer data server.

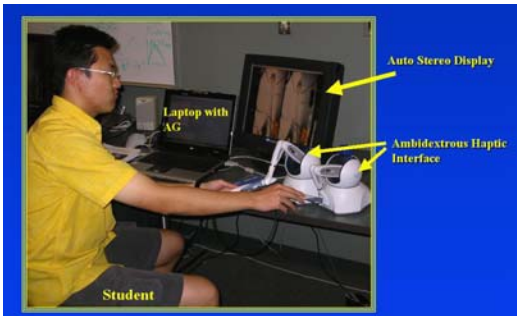


Figure 7: Trainee Multi-Modal Interface

V. CONCLUSIONS

The expected result of this project is a demonstration and evaluation of a realistic HAVE immersive collaborative environment application for the training of ophthalmic residents in cataract surgery between the University of Ottawa and the University of Alberta through regional high-speed networks and CA*net4. The first prototype is working very well. We are able to remotely capture subtle moves from the training surgeon, and we are also able to demonstrate that the network delay does not influence the performance of the system. The next step is to perform a formal usability study of this interface to demonstrate that one can perform true surgical skill transfer over very large distances.

This project was financed in part by Canarie Advanced Research Project ARP-20.

REFERENCES

[1] U. Kuhnapfel, H. K. Cakmak, B. Chantier, H. Maass, G. Strauss, C. Trantakis, E. Novatius, J. Meixensberger, K. Lehmann, H. J. Buhr, M. Lawo, and G. Bretthauer. "HapticIO: Haptic Interface-Systems for Virtual-Reality Training in Minimally-Invasive Surgery". Proceedings of the Internationale Statustagung "Virtuelle und Erweiterte Realität", Leipzig, 19.-20. Feb. 2004

[2] R. W. Webster, D. I. Zimmerman, B. J. Mohler, M. G. Melkonian. A prototype haptic suturing simulator. Medicine Meets Virtual Reality, IOS Press

[3] Y. Shon and S. McMains. "Evaluation of Drawing on 3D Surfaces with Haptics". IEEE Computer Graphics and Applications, 24(6), 40-50, 2004

[4] P. Abolmaesumi, K. Hashtrudi-Zaad, D. Thompson, A. Tahmasebi, "A Haptic-based System for Medical Image Examination". Proceedings of IEEE 26th Annual International Conference of the Engineering in Medicine and Biology Society, Vol. 3, pp. 1853-1856, San Francisco, CA, September 2004.

[5] F. A. Mussa-Ivaldi and J. L. Patton. "Robots can teach people how to move their arm". IEEE International Conference on Robotics and Automation (ICRA), San Francisco, CA, 2000.

[6] S. Govindarajan and T. Kesavadas. "Motor Skill Training Assistance Using Haptic Attributes". Proceedings of the First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, 452-457, 2005.

[7] D. Wang, Y. Zhang, and C. Yao. "Stroke-based modeling and haptic skill display for Chinese calligraphy". Virtual Reality, 9, 118-132, 2006.

[8] D. Feygin, M. Keehnder and F. Tendick. "Haptic Guidance: Experimental Evaluation of a Haptic Training Method for a Perceptual Motor Skill". Proceedings of the 10th Symposium On Haptic Interfaces For Virtual Environments and Teleoperator Systems, 2002.

[9] J. Liu, J. L. Emken, S. C. Cramer, and D. J. Reinkensmeyer. "Learning to Perform a Novel Movement Pattern using Haptic". Proceedings of 9th International Conference on Rehabilitation Robotics, 37-40, 2005.

[10] J. Kim, H. Kim, B. Tay, M. Muniyandi, and M. A. Srinivasan. "Transatlantic Touch: A Study of Haptic Collaboration over Long Distance". Teleoperators and Virtual Environments , 13, 328-337, 2004.

[11] C. Basdogan , C.-H. Ho, M. A. Srinivasan, and M. Slater, "An experimental study on the role of touch in shared virtual environments". ACM Transactions on Computer-Human Interaction (TOCHI), 7(4), 443-460, 2000.

[12] C. Gunn, M. Hutchins, M. Adcock, and R. Hawkins. "Trans-world haptic collaboration". Proceedings of SIGGRAPH 2003 (Sketches and Applications), San Diego, USA, July 2003.

[13] C. Gunn, M. Hutchins, M. Adcock, and R. Hawkins, "Surgical training using haptics over long internet distances," Proceedings of Medicine Meets Virtual Reality 12, USA, 2004.

[14] C. Gunn, M. Hutchins, D. Stevenson, M. Adcock, and P. Youngblood, "Using collaborative haptics in remote surgical training," Proceedings of WorldHaptics, Pisa, Italy, 2005.

[15] A. Matthew, M. Hutchins, D. Stevenson, C. Gunn, A. Krumpholz, T. Adriaansen, B. Pyman, and S. O'Leary, "Communication in a networked haptic virtual environment for temporal bone surgery training," Virtual Reality, 1 - 11, 2005

[16] M. Hutchins, M. Adcock, D. Stevenson, C. Gunn, and A. Krumpholz. "The design of perceptual representations for practical networked multimodal virtual training environments," Proceedings of the 11th International Conference on Human-Computer Interaction, Las Vegas, 2005.

[17] M. Hutchins, S. O'Leary, D. Stevenson, C. Gunn, and A.Krumpholz, "A networked haptic virtual environment for teaching temporal bone surgery," Proceedings of Medicine Meets Virtual Reality 13, IOS Press, 2005

[18] T.Hikichi, A. Yoshida, S. Igarashi, N. Mukai, M. Harada, K. Muroi, and T. Terada. "Vitreous Surgery Simulator," Archives of Ophthalmology, 118, 1679, 2000.

[19] P. Boulanger. "From High Precision Color 3D Scanning of Cultural Artifacts to its Secure Delivery Over the WEB: A Continuum of Technologies", Workshop on "Recording, Modeling and Visualization of Cultural Heritage in Ascona Switzerland, May 22-26, CD Proceedings, 2005.

[20] http://www.accessgrid.org