A Vision of Telepresence
for Medical Consultation and Other Applications

Henry Fuchs and Ulrich Neumann

Department of Computer Science
University of North Carolina
Chapel Hill

fuchs@cs.unc.edu
neumann@cs.unc.edu

Abstract

A future telepresence system may significantly enhance the assistance that a remote expert can render to a distant physician during time-critical medical encounters such as the initial medical intervention in serious injury cases. The preliminary design of such a telepresence system consists of two major components: 1) a set of cameras and other image acquisition devices at the patient and local physician's site (Fig. 1a) that can acquire, continuously and in real time, a usefully detailed 3D description of the relevant parts of the environment such as the patient's regions of injury, 2) a virtual environment facility at the remote specialist's site (Fig. 1b) which allows that specialist to walk about the virtual environment defined by the continuously updating 3D information from the patient's site. These two components, connected by a high speed communication network, will form a system that is unobtrusive to the local physician and patient. It also provides for the remote specialist a comprehension of the patient's condition that is so natural as to require zero training time.

While the greatest technical challenge is the realization of component-one - extracting the useful detail, recent results in robotic vision, both in extracting such information from a multiplicity of cameras [Koch 93] and from light striping systems [Gruss, Tada, Kanade 92], are encouraging us to examine whether this extraction of 3D structure can be done in real time. Should one of these systems be successfully built and used, it may have implications for reducing the cost of medical care by improving the treatment in the first minutes of an emergency situation and by reducing the cost and delay in non-critical situations in which a second opinion from a specialist is desired. Other applications in non-medical areas can be contemplated should the cost of these systems become sufficiently low. One can images that systems with such capabilities may become a preferred mode of teleconferencing in the 21st century.

1. Introduction

Today's teleconferencing systems can be though of as simplified telepresence systems in that they do not allow, with their fixed cameras and microphones, any convenient navigation for the users in the remote environments. While many teleconferences are still useful, these restrictions are onerous for many complex tasks such as a medical diagnosis or consultation on surgical procedures. Movable boom-mounted cameras may be used to provide control of the view, but the camera switching and direction-controlling mechanisms (e.g.: joysticks and buttons) are likely to require training or a human operator to manipulate. A future telepresence systems, in which the model of the remote environment is registered one-to-one with the users physical environment, promises to provides a natural way of comprehending a complex 3D environment that is physically remote by using natural body and head motions to control the viewing position. In addition to viewing control, it is also important for the remote consultant to communicate information back to the patient site. Specifically, the consultant should be able to point to areas on the virtual patient while relating instructions to the local physician. Rather than using joysticks or some similarly indirect metaphor for picking an area of interest, the consultant can simply indicate the location of interest with a 3D mouse. The corresponding point is highlighted by a computer-steered laser-pointer at the patient site, thereby communicating directly to the local physician the consultant's point of interest.

pp 565-571

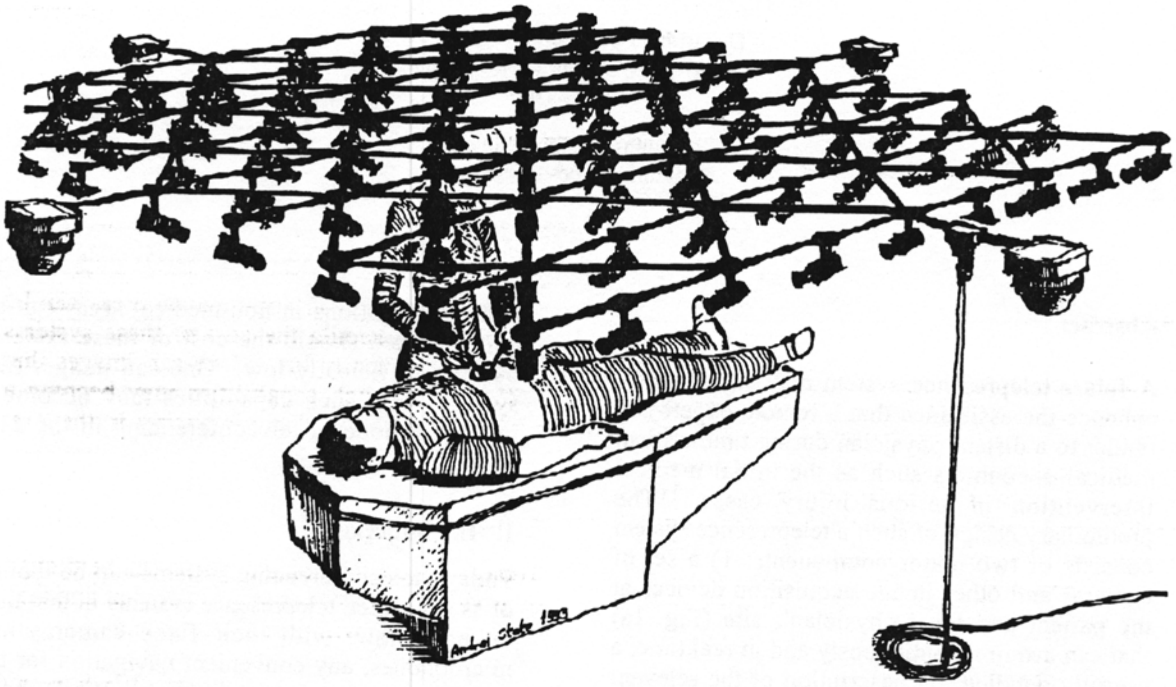

Fig. 1a - Local patient and physician under light-stripe and sensor array

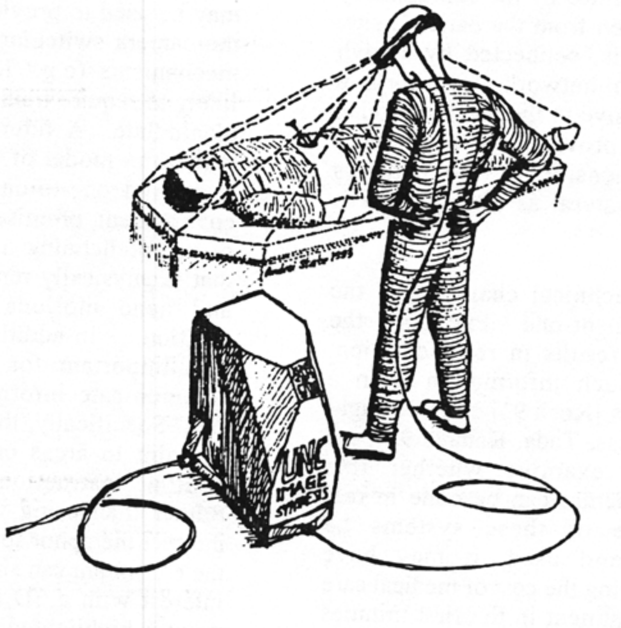

Fig. 1b - Remote physician viewing virtual patient

2. Virtual Environments

A telepresence system such as the one of interest here involves aspects of virtual environments and augmented reality. It may be helpful to understand the relations between these systems by briefly studying the taxonomy presented in [Bishop, Fuchs, et.al 1992]. In particular, the possibility should be considered that, in such a visualization-rich situation, the local and perhaps also the remote physicians may wish to visualize certain medical data such as ultrasound images registered in-place within the patient (e.g., for such tasks as determining the presence and precise location of foreign bodies lodged within an accident victim). The taxonomy described below outlines the structural relationship among various such possibilities.

2.1. Simulated Environments

Simulated environments are the most common class of virtual environment applications. In a typical simulated environment (Fig. 2) a computer system maintains a world model and allows the person to perceive and interact with it. The person in the simulated environment wears an HMD which displays images generated by the graphics system. As the person moves about and turns his or her head, the system tracks the change in the HMD position and computes new images for display. The person is presented with the illusion of being immersed in a modeled world. In addition to the head-tracker information, further control can be provided by 6 DOF cursors, hand gestures, menus selections, speech, or any input that can be sensed by the computer. These inputs can be used to alter the world model or move within it. A person's perception of the modeled world is not limited to visual images only, but may include other senses as well.

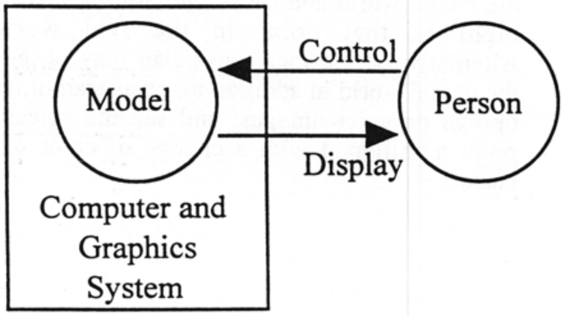

Fig. 2 - Simulated environment

2.2. Augmented Reality

An augmented reality (Fig. 3) combines a simulated environment with direct perception and action. The real world and computer-generated graphics are viewed together, often spatially registered for the purpose of annotation or enhancement. The computer model must accurately reflect the real world to enable registration of the real world and computer graphics images. Real world images and computer graphics may be combined optically or electronically for display in the see-through HMD. Optical merging can be achieved with half-silvered mirrors or beam splitters [Sutherland 68]. If video cameras are mounted on the HMD near the wearer's eye positions, electronic merging of the computer graphics and the video images can be performed on a pixel-by-pixel basis. When depth information is known for the real world images, electronic merging facilitates computing the proper obscuration for the combined scene. An example of an augmented reality (Fig. 4) shows ultrasound data overlaid on a video image of a patient [Bajura, Fuchs, Ohbuchi 92]. In this case obscuration is not performed so that the ultrasound data remains visible. An augmented reality is useful for annotating the real world with data from the world model or enhancing the sensory perceptions of the person using the system with computer graphic visualizations.

2.3. Medical Telepresence

Medical telepresence (Fig. 5) combines elements of a simulated environment and an augmented reality. The local physician experiences an augmented reality with either a video or optical see-through HMD. For example, an on-line scanner (MR or ultrasound) provides a real-time view of the patient's internal anatomy and any surgical instruments inserted in the area. The scanner data is registered and displayed in the correct location of the real world image. The local physician interactively controls the location and parameters of the scanning. Augmented reality provides a natural way for the physician to understand the positions of the scan data and surgical instruments within the patient. This perception of spatial locations is superior to that obtained from a physically separate scanner display.

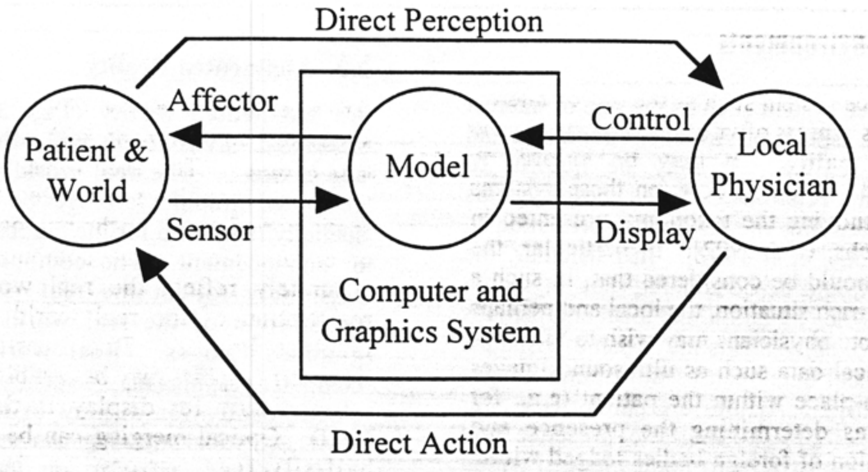

Fig. 3 - Augmented reality

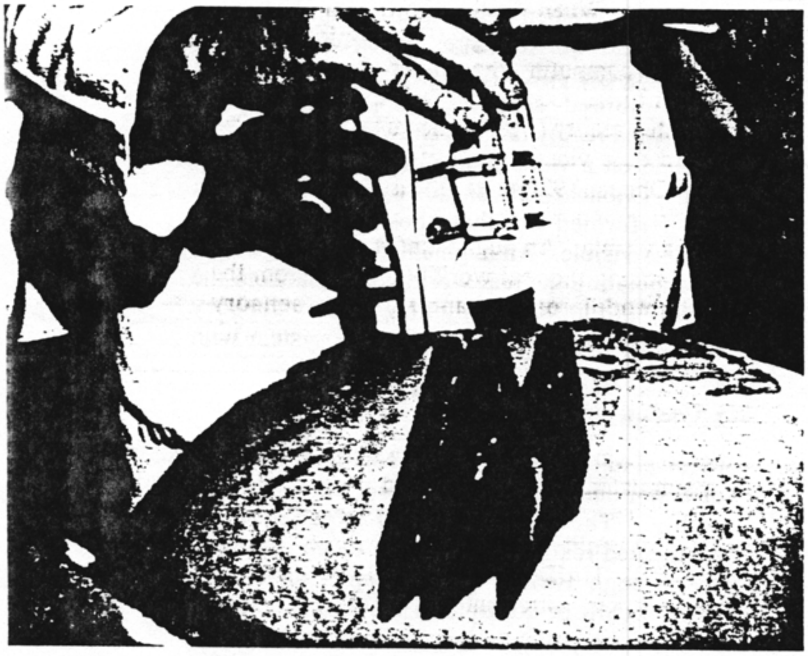

Fig. 4 - Ultrasound visualization in augmented reality

The same augmented reality perceived by the local physician is approximated for the remote specialist. To do this, a model of the patient, the patient's immediate surroundings, and all pertinent medical data from associated equipment is acquired by the computer system. (The details of how this may be done are described in section three.) The model is transmitted to the remote location over a high-bandwidth communication link, and the remote system renders a simulated environment for the specialist there. Since the distant specialist only perceives the model world, it is essential to make it as accurate a representation of the real world as possible. One of the research challenges in this area is capturing as much fidelity in the model as possible and doing it at interactive rates. Section three describes our approach to this problem.

Due to the critical and experimental nature of this application, it is not appropriate to give the remote specialist control over any affectors that touch the patient in any. However, the remote specialist should be able to communicate effectively to the local physician and in some cases the patient. To provide visual communication from the remote specialist, in addition to audio communication, a *virtual-pointer* is provided by several laser pointers mounted in the environment about the patient. The remote specialist can "touch" a portion of the model world and cause the laser pointers to highlight that point in the real world. Alternatively, the local physician may observe the model world in addition to, or instead of, the optical or video images, and see the selected point highlighted with a change of color or a cursor.

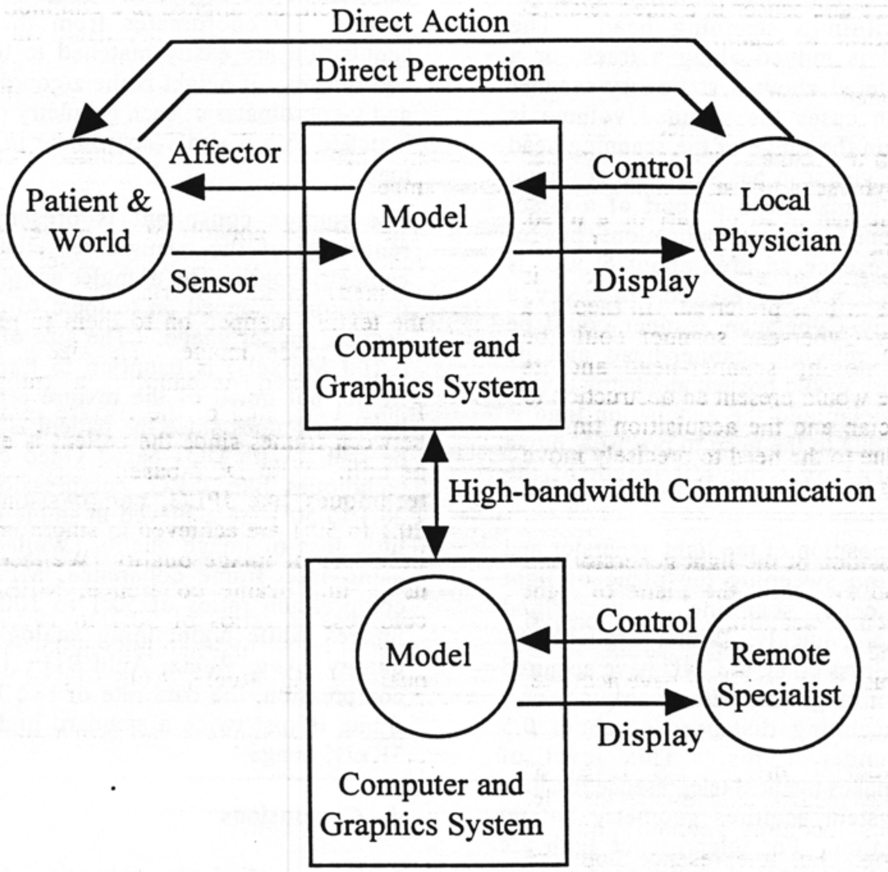

Fig. 5 - Medical telepresence combining simulated environment and augmented reality

3. Model Acquisition

A major research question is how to acquire a model of the patient and surrounding space. Ideally, the model is geometrically accurate and the modeled volume includes the patient, the surrounding instrumentation, and the local physician when he or she is within reasonable proximity of the patient. In addition to representing the physical geometry, the model must also contain the scanner data used to augment the local physician's images so that similar visualizations can be computed for the remote specialist's vantage point. Any instrumentation (e.g.: ECG, blood pressure monitor) connected to the patient should also have its data and display position included in the model so that the remote specialist can view that data on a simulated model of the instrument located in a corresponding position of the model world. Lastly, the time to detect and update any changes in the model must be as short as possible to facilitate interactive guidance from the remote specialist.

The model acquisition problem is not the only issue that needs addressing for medical telepresence systems to become a reality, but it is the issue that presents the greatest challenge for the computer graphics community. A possible solution is supplied by computer vision research that has focused on the geometry acquisition problem for many years. One approach that appears particularly promising for medical telepresence is light-striping. Light-striping methods generate infrared stripes in the scene by shining a plane of light into the environment. Sensors view the stripe as it moves over the profile of objects in the scene. If the stripe and sensor orientations are known, the 3D coordinates of any illuminated point may be determined by triangulation [Gruss, Tada, Kanade 92].

Light-striping is the basic technology behind the Cyberscan, a commercial scanner from Cyberware Corp. The system incorporates a light-plane source and sensor in a fixed relative-orientation within a scanning head. The scanning head is moved along a track, or a platform is rotated about a stationary scanner head. In both cases the scanned volume is contained within the range of the scanning head motion. The Cyberscan has sufficient scanning volume for a human head or part of a torso. This is sufficient for many applications, but for medical telepresence an area about 6' x 3' is necessary, and 8' x 8' is preferred. In theory, a larger full-body Cyberscan scanner could be built, but the moving scanner-head and its support structure would present an obstruction to the local physician and the acquisition time is rather lengthy due to the need to precisely move the scanner-head over the full length of the patient.

By fixing the position of the light generator and the sensors, and sweeping the plane of light across the scene, scanning is performed unobtrusively and quickly. With a parallel VLSI sensor array, Gruss et. al. at CMU have acquired 28x32 points of range data with an accuracy of < 0.1%, at a scanning distances of about 0.5 meters, in under 1 ms. This level of performance makes medical telepresence feasible. The CMU system acquires geometry but no color information. For telepresence both are necessary, but the patient geometry need only be captured to a relatively coarse grid. A sample-grid spacing of <0.5cm over an 8'x8' scan area requires about 512x512 scan points which may be achieved with a 16x19 array of the current CMU sensors. It may be necessary to use several light-stripe sources to adequately cover such a large area, but the speed of these sensors allows over thirty sweeps to be performed within a single 30 Hz video frame.

In addition to geometry, color must be acquired over a large scan area. Color should be acquired at a high resolution so that as much detail as possible is presented to the remote consultant. Fortunately, color cameras with 640x512 resolution are common and suitable for the task. If the 16x19 array of light-strip sensors is interleaved with a 16x19 array of color cameras, the resolution over an 8'x8 area is 10240x9728, or equivalently, over 100 dpi. Allowing for some overlap and range of depth, 64 dpi is readily obtained.

Camera and lens system calibration may be performed off-line for each camera to compensate for radial lens-distortion and obtain its position, field of view, and viewing direction [illegible] calibrations are aligned in a common coordinate system, 3D coordinates from the geometry acquisition are easily matched to texture map coordinates; if height is the z-coordinate, the x and y-coordinates of each geometry point can be matched to a u and v-coordinate in the texture map.

The remote consultant is presented with a rendering of the triangle tessellation of the geometric mesh. The triangles are rendered with the texture mapped on to them to producing a detailed model image. The size of the texture (~100 Mpixels) is daunting to transmit every frame, but much of the texture remains fixed between frames since the patient is expected to be still. With DCT-based video compression techniques like JPEG, compression ratios of 20:1 to 50:1 are achieved in single images with little loss of image quality [Wallace 91]. By using inter-frame coherence, MPEG obtains compression ratios of 50:1 to 100:1 in video images while maintaining analog video tape quality [Ang, Ruetz, Auld 91]. Using JPEG compression, the data rate of two Mpixels per frame is just twice a standard high resolution 1Kx1K image.

4. Conclusions

From our preliminary exploration of this system, a plethora of questions remain, both technical and sociological:

- Will systems such as this be beneficial or will they be perceived as high tech gimmicks of very limited utility?
- Will such a facility be used frequently enough to be cost effective? Would a rural facility ever spend its scarce resources on one of these systems rather than purchasing another ambulance -- or two.
- Can the real time extraction of a 3D geometric description of the local facility be performed with sufficient accuracy with a reasonable amount of acquisition and computing equipment?
- Won't most isolated rural physicians be resistant to such a facility since they have traditionally prided themselves on their self-reliance?
- Will rural health center administrators embrace

- Why is a whole ceiling of cameras required rather than a movable panel of cameras that can be optimally positioned to "view" the area of greatest interest?

Conclusive answers for most of these questions await a working model of the system.

5. References

[Ang, Ruetz, Auld 91] Ang, Ruetz, Auld, IEEE Computer 1991

[Bajura 93] Michael Bajura, "Camera Calibration for Video See-Through Head-Mounted Display." *TR93-048 Tech Report* UNC Chapel Hill, July 1993

[Bajura, Fuchs, Ohbuchi 92] Michael Bajura, Henry Fuchs, Ryutarou Ohbuchi, "Merging Virtual Objects with the Real World." *Computer Graphics* Vol 26, no. 2, pp. 203-210, July 1992

[Bishop, Fuchs et.al. 92] Gary Bishop, Henry Fuchs, et. al. "Research Directions in Virtual Environments." *Computer Graphics* Vol 26, no 3, pp. 153-177., August 1992

[Gruss, Tada, Kanade 92] Andrew Gruss, Shigeyuki Tada, Takeo Kanade, "A VLSI Smart Sensor for Fast Range Imaging." *Proceedings of IROS*, 1992

[Kanade, Gruss, Carley 91] Takeo Kanade, Andrew Gruss, L. Richard Carley, "A Very Fast VLSI Rangefinder." *IEEE International Conference on Robotics and Automation*, pp. 1322-1329, April 1991

[Koch 93] Reinhard Koch, "Automatic Reconstruction of Buildings from Stereoscopic Image Sequences." *Proceedings of Eurographics '93*. Computer Graphics Forum. Vol 12, no 3, pp. c339-c350, September 1993

[Sutherland 68] Ivan Sutherland, "A Head-mounted Three Dimensional Display." Fall Joint Computer Conference, 1968. Thompson Books Wash. DC., pp.757-764.

[Wallace 91] Wallace, Communications of the ACM, July 1991