Improving, Expanding and Extending 3D Telepresence

Greg Welch, Henry Fuchs, Bruce Cairns, M.D.[†], Ketan Mayer-Patel, Diane H. Sonnenwald[§], Ruigang Yang[‡], Andrei State, Herman Towles, Adrian Ilie, Michael Noland, Vincent Noel, and Hua Yang

The University *of* North Carolina *at* Chapel Hill
The Department of Computer Science
Chapel Hill, NC 27599-3175
Email: {welch, fuchs, kmp, andrei, herman}@cs.unc.edu

[†] The University *of* North Carolina *at* Chapel Hill
Department of Surgery
Chapel Hill, NC 27599
Email: bruce_cairns@med.unc.edu

[§] The Swedish School of Library and Information Science
Göteborg University and the University College of Borås
501 90 Borås, Sweden
Email: diane.sonnenwald@hb.se

[‡] The University *of* Kentucky
Department of Computer Science
Lexington, KY 40507-1464
Email: ryang@cs.uky.edu

Abstract

For over a decade we have been pursuing a vision of three-dimensional (3D) telepresence. Our primary driving problem is 3D medical consultation. Our basic goal is to enhance and expand medical diagnosis and treatment for life-critical trauma scenarios by providing an *advising* health care provider and a distant medical *advisee* a "live" high-fidelity sense of 3D presence with each other.

Here we provide some background on our current capabilities, and describe plans and some strategies to improve the 3D resolution and quality, to increase the working volume, and to extend the user paradigms for 3D telepresence.

Keywords: Telepresence, tele-existence, Virtual Reality, medical consultation, computer vision, user interfaces.

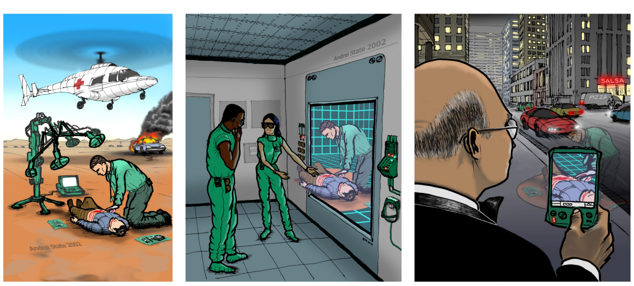

Figure 1: Conceptual sketches of some three-dimensional (3D) medical telepresence paradigms. Left: a portable multi-camera unit. Center: hospital-based 3D viewing. Right: mobile 3D viewing. (Sketches by Andrei State.)

1. Introduction

Medical trauma is frequently referred to as the "hidden epidemic of modern society ." It is responsible for more lost productive years than heart disease, cancer, and stroke combined [9, 4]. Because trauma can occur at anywhere and any time, and early, appropriate intervention saves lives, trauma management has long been proposed as an ideal telemedicine application [13]. While there have been reports of success in limited trials using 2D video teleconferencing [12], issues such as viewpoint control and depth perception continue to limit the application and acceptance of the currently available systems [18, 19, 20, 17]. Tachakra reports in [16] that the "impaired depth perception (with 2D conferencing) is a significant problem in telemedicine." 3D *telepresence* [15] has been described as the "ultimate development in telemedicine" but a concept that "remains in its developmental infancy" [13].

Our vision of 3D medical consultation has a corresponding overriding technology goal: we want to create a physically natural and intuitive visual and aural sense of presence in a remote place, while minimizing a participant's awareness of the technology-related external factors—they should be able to ignore the technology and concentrate on the medical task. Ideally we want an advisor and remote advisee to feel as if they are standing or kneeling next to each other. We want them to see and hear each other almost as if they were together, to see the same things on and around the patient, and to be able to naturally point and gesture with their hands.

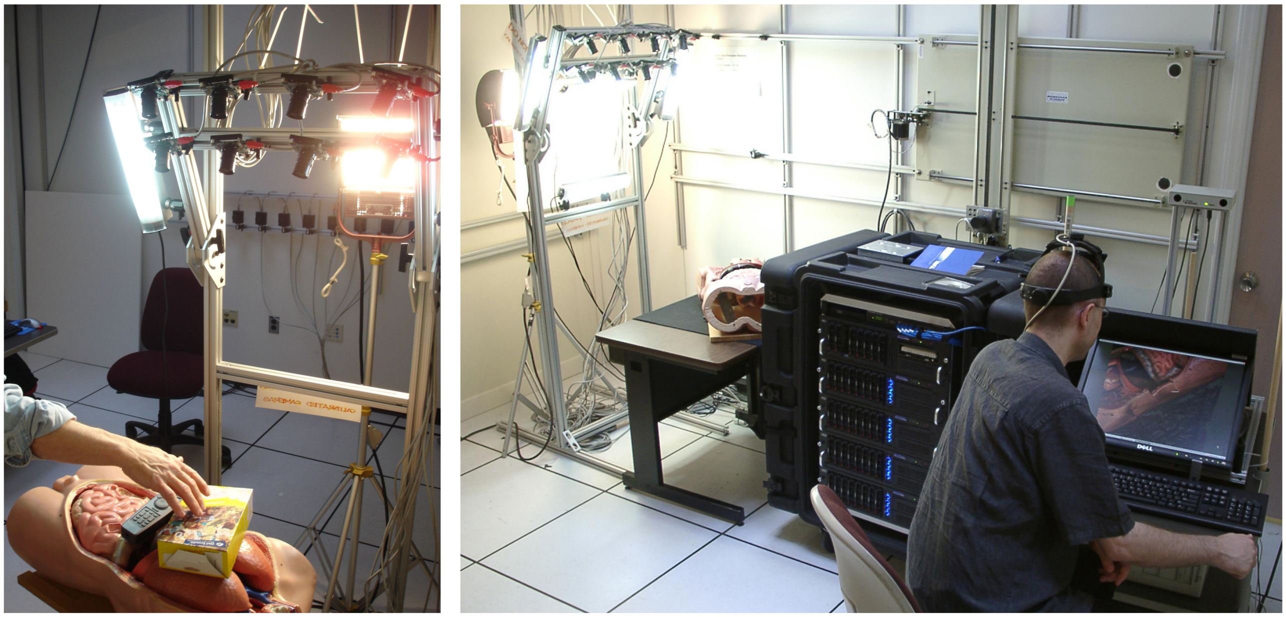

Figure 2: The current 3DMC prototype. Left: a camera-lighting array with eight Firewire cameras and high-frequency area lights. Right: co-author Andrei State uses the transportable rack with five high-performance servers and a head-tracking interface to view live 3D imagery from the eight cameras.

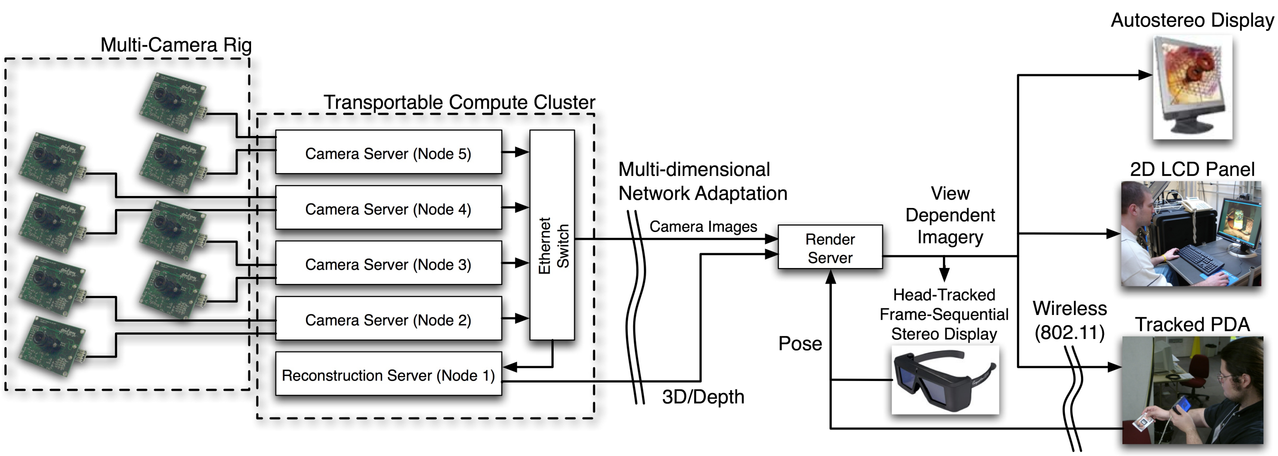

Figure 3: An overall prototype system diagram. The equipment to the left of the "Multi-dimensional Network Adaptation" boundary (middle) would reside with the patient and advisee. This includes most of the equipment shown in Figure 2. The portion to the right would reside near or with the remote consultant. This includes the display and head tracking shown on the right in Figure 2. Note that the *Multi-dimensional Network Adaptation* (shown in the middle) has not yet been integrated into the prototype.

We are a little more than mid-way through a three-year United States National Library of Medicine funded effort aimed at "3D Telepresence for Medical Consultation" (3DMC). Our research is focused on the key technological barriers to 3D telepresence including real-time acquisition, networking, tracking, and displays, and undertaking a parallel systems integration effort to focus our research and development aimed at various usage scenarios.

In the following sections we provide background on our current capabilities, and describe plans and some strategies to tackle the next technological barriers to improving the 3D resolution and quality, increasing the working volume, and extending the user paradigms for 3D telepresence.

2. Current Prototype and Methods

Our current 3DMC prototype consists of a camera-lighting array, a transportable compute/rendering cluster, and a head-tracked display. See Figures 2 and 3. The camera-lighting array (left in Figures 2 and 3) consists of eight 640×480 resolution digital (IEEE 1394a) color cameras from Point Grey Research [1]. The cameras are currently mounted in two horizontal rows of four on a portable stand that can be positioned next to a patient. The cameras are positioned so their visual fields overlap the same region of interest on the patient. Mounted around the cameras are multiple high-frequency fluorescent fixtures.

The compute/rendering cluster (right in Figure 2, middle-left in Figure 3) consists of five dual-processor PCs mounted in a transportable, shock-resistent rack case. Four PCs function as camera servers, JPEG compressing the raw Bayer pattern images and forwarding the compressed video streams via dedicated gigabit Ethernet to the 5^{th} PC. The 5^{th} PC then decompresses the video streams, loading the color images into texture memory of the graphics card for view-dependent 3D reconstruction as described next.

2.1. 3D Telepresence

To achieve a visual sense of 3D telepresence we use a novel approach called *View-dependent Pixel Coloring* (VDPC) [22]. VDPC is a hybrid image-based and geometric approach that estimates the *most likely color* for every pixel of an image that would be seen from some *desired viewpoint*, while simultaneously estimating a 3D model of the scene. By taking into account object occlusions, surface geometry and materials, and lighting effects, VDPC can produce results where other methods fail—in the presence of textureless regions and specular highlights—conditions that are common in medical scenes.

We use the graphics hardware on the 5^{th} PC (see above) to perform the 3D reconstruction very quickly as the images arrive from the camera server PCs [23]. The basic idea is to use the graphics hardware to rapidly render the camera images onto a series of virtual (computer graphics) planes swept through the scene, searching in parallel for the best color matches (least variance) at a dense set of points on the planes. The result is a relatively dense depth map that we can then render, again using the graphics hardware.

In some early experiments we constructed a one-meter-cubed rig with eight downward looking cameras. Figure 4 shows a sequence of images reconstructed from a recording of Dr. Bruce Cairns, M.D. performing a mock surgical procedure on a physical patient model. Those views (images) were reconstructed off line after the procedure. Figure 5 shows some results from our current prototype (Figure 2). Those views were reconstructed on line, in real time.

2.2. Displays

When a medical advisor is on duty in a hospital, it is reasonable to expect that they might have access to facilities for stereoscopic, head-tracked viewing of dynamic 3D reconstructions of the remote patient and advisee. Our current prototype addresses this scenario with a high-resolution LCD panel and an Origin Instruments opto-electronic tracker with software for head-tracked visualization. As shown on the right in Figure 2, the user wears a head band with three infrared LEDs that are tracked in real time by a small sensor unit. From this we compute the location of the user's dominant eye and render the reconstructed imagery from that point of view, providing natural monoscopic head-motion parallax. We plan to add time-multiplexed (shuttered) stereo visualization soon, and have even been investigating *autostereoscopic* displays.

Tracked PDA Viewing. We also want to provide the best possible 3D experience when the medical advisor is away from the hospital. For a remote display we are looking at *personal digital assistants* (PDAs). Our goal is to develop or adapt tracking technology and user interface paradigms that will allow a remote medical advisor to use a PDA as a "magic lens" for interactively looking at the remote scene from different perspectives [5, 6, 10, 11].

We are currently investigating a two-handed patient "prop" paradigm [7]. The basic idea is that the advisor would have a physical prop that serves as a surrogate for the patient, and a PDA that is tracked *relative to the prop.* For example, the PDA cover could serve as the prop. The advisor would then hold the prop in one hand and the PDA in the other, moving them around with respect to each other as needed to obtain the desired view. Figure 6 shows two different prototypes.

From a user interface perspective, this is not a new idea. Hinckley, et al. experimented with using a dolls head or rubber ball and various tools as 'props' for neurosurgeons visualizing patient data [7]. While each object was tracked relative to the world, the interaction was handled with respect to the individual objects. Hinckley found that users could easily position their hands relative to one another quickly—a task we all do frequently.

For our specific application the prop paradigm provides the user with an instant visual target to aim their "magic lens" at. The approach also affords new ways of looking at the data. For example, an advisor can rotate the prop to quickly get a different view, rather than spending time and energy walking around to the other side. As a bonus, tracking a PDA relative to another object is a much more tractable problem than tracking a PDA relative to the world in general, opening up a number of potential tracking solutions that were otherwise not feasible.

3. Challenges and Future Plans

At the end of our current project funding (see Section 1) we will still be far from our long-term vision, however we will have completed the first stages of fundamental research and prototype development and will have the results of a formal set of experiments comparing 2D televideo with 3D telepresence. We are encouraged by our results so far, and are ready to push the work forward into new areas related to higher-fidelity 3D reconstruction and display, larger-scale prototypes, full-duplex telepresence, new and improved advisor/advisee interfaces, and new usage paradigms. We discuss some of the issues and plans here.

3.1. Improving 3D Telepresence

Any approaches we develop for dynamic 3D scene reconstruction will be based on image features seen in 2D imagery from cameras mounted in the room, near the patient (on a stand or the surgical lamp), or even on the surgeon. The success of any such approaches fundamentally depends on the visibility and quality of the observed features. No algorithm can succeed with cameras that are blocked by instruments or moving humans, or features that are indiscernible, much less absent.

While we have many concrete research plans for such things as heterogeneous active camera networks and static pre-operative laser scanning, there are two particular areas of research that we believe offer the greatest hope for addressing feature visibility and quality: *Imperceptible Active Reconstruction* and *Optimal Camera Placement.* We describe them briefly here.

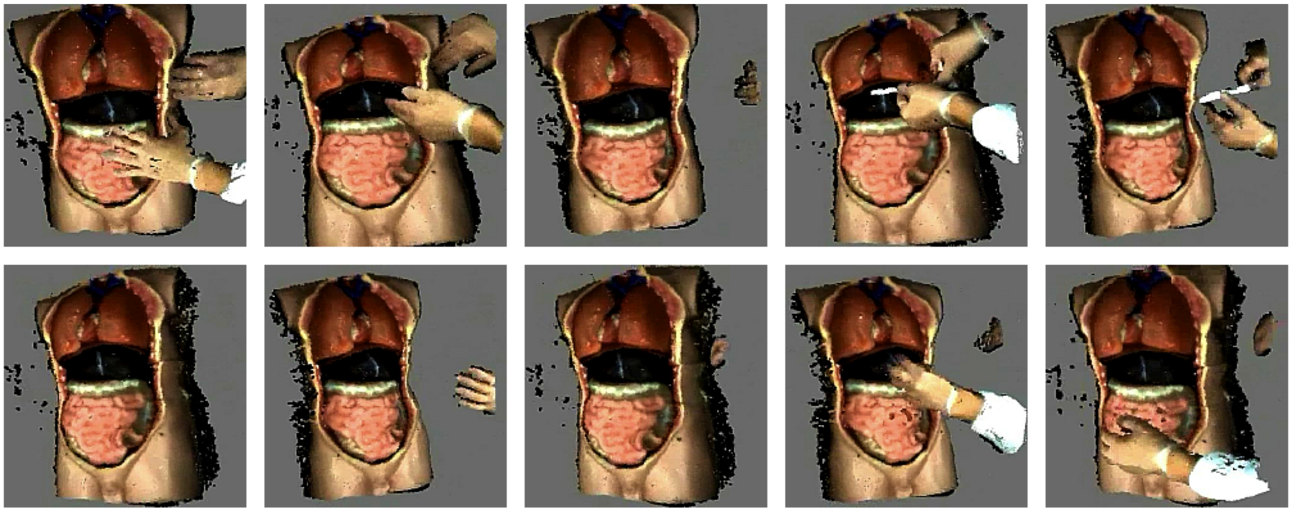

Figure 4: Novel view images from an early mock surgical procedure on a training torso. Each image is from a different time during the session, and a completely novel viewpoint (none the same as the cameras).

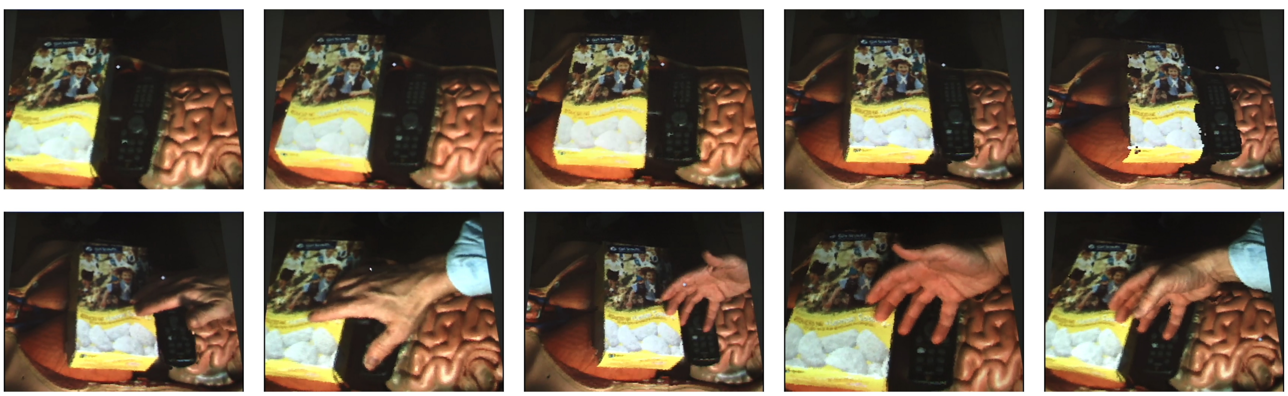

Figure 5: A more recent sequence of novel view images reconstructed using the system shown in Figure 2. We set a box of Girl Scout cookies on top of the training torso to provide more obvious scene geometry. As in Figure 4, each image is from a different point in time and a completely novel viewpoint.

Imperceptible Active Reconstruction. The first area of work involves active control of room lights, and something we call *imperceptible structured light* (ISL). The basic idea is to embed carefully chosen texture patterns in digital projector imagery in a way that renders the patterns invisible to human viewers. We intend to use digital projectors as "smart surgical lights" that illuminate the patient and the room with what appears to be white light, but in fact contains imperceptible patterns. Injecting such ISL patterns into a scene will improve feature quality, significantly improving the quality and robustness of 3D reconstruction results. The greatest new challenge in this area is to develop methods that allow multiple ISL projectors to project into the same space, without interfering with each other.

Optimal Camera Placement. To address the visibility problems we plan to mount cameras in the room, on equipment (where possible, e.g., surgical lamps), and even on the surgeon's head. However attempting to put cameras "everywhere" would be impractical if not impossible. To help with this problem we are developing new simulation and visualization tools to help us evaluate candidate designs, e.g., camera quantity and arrangement, resolution and feature signal strength, coverage, and occlusions. Specifically we are developing a software tool that will allow a designer to interactively change camera (and projector) configurations and parameters, while visualizing the expected steady-state 3D uncertainty throughout the acquisition space [3]. Like fluid or air flow visualization, the idea is to allow the designer to see the "flow" of sensor information throughout the space, looking for hot and cold spots, etc.

In addition we have been working on a geometric simulator designed to help us plan camera positions and orientations in a scene where we want to perform 3D reconstruction [14]. This tool is *geometric* in that it only takes into account a camera's field of visibility, resolution, and occlusions, as opposed to the *stochastic* tool described above and in [3], which also takes into account the expected scene dynamics, the camera frame rate, measurement noise, etc.

Figure 6: Left: Our first tracked PDA prototype used a HiBall-3000 ^{TM} tracking system [2], with sensors mounted on the PDA (Toshiba e800, left hand) and the surrogate (right hand). Right: Our current prototype uses a PointGrey DragonFly camera [1] mounted on the PDA (left hand). The prop (right hand) has a printed image of our training torso on it, along with a grayscale pattern. We use the ARToolkit [8] to track the surrogate with respect to the PDA (camera).

3.2. Expanding 3D Telepresence

Our results thus far have been limited to a cubic volume approximately *40 centimeters on a side* with a volumetric resolution of approximately *three 3D points per centimeter*. We can display these (small-scale) results in an immersive Virtual Reality display system or on the web as stereo movies or dynamic 3D models.

Expanding the working volume presents monumental challenges in terms of 2D video acquisition, semi/fully-automated dynamic 3D scene reconstruction, and presentation. For example, with respect to acquisition and 3D reconstruction we aim to *scale up the working volume* while simultaneously improving the *resolution* and achieving sufficient *accuracy and visibility*. Doubling the spatial resolution and increasing the acquisition volume to even just two meters on a side (eight cubic meters) would result in an approximately 500-fold increase in the number of 3D points to be estimated. Achieving one millimeter spatial resolution over a typical hospital room would result in an approximately 20,000-fold increase in the number of 3D points. Furthermore the 3D points would have to be estimated 15–30 times per second for dynamic scenes.

The following are three example areas of research that we are working on aimed at expanding the effective working volume: *Model-Based Reconstruction and Tracking*, *Differential Camera Setups*, and *Immersive 2D Panoramas*.

Model-Based Reconstruction and Tracking. While the dense dynamic depth information provided by VDPC is valuable, it is noisy, and as we scale up the number of cameras and the working volume, it will result in massive amounts of data with needs for massive bandwidth. To address these and other issues, we are also investigating methods for *model-based reconstruction and tracking* of objects in a scene, and *continual refinement* of their models. The idea is that if one has a model of a scene object (e.g., polygonal or analytical) then instead of repeatedly performing dense depth maps corresponding to the object, one could simply track the object's dynamic pose (position and orientation). This offers the possibility of dramatically reducing the real-time bandwidth needs, and of offering higher-quality results from filtering and refinement. We have only been able to touch on this approach with our current funding, but believe that it could hold the key to high fidelity and low bandwidth 3D telepresence.

Differential Camera Setups. While we remain interested in synthesizing novel views from cameras in general position, we are also exploring the possibility of using cameras that are geometrically configured to support direct measures of the derivatives (rates of change) of image intensities or feature positions in a small neighborhood, with respect to object pose (position and orientation). This camera setup could be used to produce image feature Jacobian measurements that are used in a recursive predictor-corrector approach to object tracking, and perhaps to estimate object pose changes which could then be integrated to estimate the pose.

Immersive 2D Panoramas. With funding from the United States Department of Energy we have achieved encouraging preliminary results for a method we call "2D Group Telepresence." Our methods, which were targeted at DOE's tele-collaboration needs, allow for relatively high-fidelity and *wide-area* viewing of a remote scene with limited depth complexity. We believe we can extend this approach to increase the allowable depth complexity, and to support outward looking cylindrical or spherical panoramas to provide an advisor with better situational awareness.

3.3. Extending 3D Telepresence

Our current prototype operates only in a setting where the acquisition and display are co-located, it operates only in one direction (half-duplex), offers limited user interfaces, and our work has been primarily focused on medical or surgical telepresence. However we interested in extending the reach, the functionality, and the applications. Here we discuss *Operating Over Conventional Networks*, *Full-Duplex 3D Telepresence*, *New Advisor/Advisee Interfaces*, and *New Usage Paradigms*.

Operating Over Conventional Networks. Our current prototype operates using a dedicated (completely isolated) gigabit ethernet network to communicate between the processors shown to the right in Figure 2 and the middle-left in Figure 3. To extend the system over longer distances we want to be able operate over conventional networks, ranging from the Internet, to cellular networks, to local-area wireless networks. To do so however our system needs to be able to adapt to variations in bandwidth and latency throughout the network.

In [21] we describe a new adaption framework that attempts to carefully manage the network resources to ensure that at all times we transmit the data (images or 3D) that is most useful to the overall application and the goals of the user at any moment. The use of this *Multi-dimensional Network Adaptation* framework is indicated in the middle of Figure 3. This framework is not currently integrated into our prototype, however we are planning complete integration (as indicated in Figure 3) within a year.

Full-Duplex 3D Telepresence. By the end of our current funding we expect some success in giving an advisor a sense of presence with a small portion of a remote advisee's working area. But an equally critical component of the collaboration is the advisee's sense of presence with the advisor. We believe that giving the advisee some sense that the advisor is standing by their side, and can see what they see, is crucial in terms of technology acceptance, advisee trust in the system, and the overall effectiveness.

With current funding we have been able to achieve modest success in one direction ("half-duplex" telepresence), and believe we will be ready to tackle bi-directional or "full duplex" telepresence in a follow-on effort. We believe that achieving this will be more difficult than simply replicating the results of our preliminary work, in particular for an advisee in a portable scenario.

For example, should we attempt to give the remote advisee a 2D or a 3D version of the advisor? Do they need a stereo and/or head-tracked display? Is audio alone sufficient in some cases? Ideally the advisee would see a representation of the advisor in the local scene with them. The advisor should appear to look at the advisee and the patient when the advisor is actually doing so, and if the advisor points with a finger or a virtual laser pointer the advisee should see the finger/laser at the appropriate place in the scene. While this could be accomplished with an advisee-worn head-mounted display, this would both encumber the advisee more and require more infrastructure. Instead one might have a stand-alone display, and/or even project imagery (or laser pointer light) directly onto the patient. The best approach for a particular scenario will involve tradeoffs between infrastructure (bandwidth, tracking, etc.), advisee encumbrances, and corresponding added value.

New Advisor/Advisee Interfaces. Our current work in scene reconstruction typically involves using a collection of cameras and lights (see Figure 2). We are also working on integrating digital projectors and *imperceptible structured light* techniques as described above. Our long-term goal however is not a disjoint collection of devices, but small atomic units with tightly integrated cameras, projectors, and digital lighting. These Image-Based Input-Output (IBIO) devices might look simply like portable or permanent surgical lamps. However in addition to providing general task lighting they would support projection of arbitrary graphics (pointers or advisor's hands) onto the "illuminated" surfaces, while simultaneously *capturing* and reconstructing the areas in 3D. We envision a day when IBIO devices are standard-issue for EMTs (police officers, etc.), and deployed throughout medical facilities, for use in both consultation and training paradigms.

Currently we can display dynamic 3D reconstructions for a single advisor with stereo and head-tracked imagery, using either a head-mounted display or digital projectors. While the techniques for doing so are well known, what is less obvious is how to provide the same for a two or three (or more) person team of advisors such as a trauma team assessing a remote patient (stationary or in transit). While auto-stereoscopic displays offer many advantages, they are typically small and designed for inward "fish tank" viewing. We are eager to explore panoramic projector-based systems with passive head tracking designed to reduce the number of viewing rays that need to be computed and/or projected at any moment.

New Usage Paradigms. We believe that there are other paradigms where the technology described here could prove useful for training or analysis. In particular we are interested in real-time multicasting and remote observation of medical procedures for training purposes, and TiVo-like recording of advisor and advisee 3D scene reconstructions, audio, and vital signs for off-line analysis and training.

Beyond medicine we are interested applications related to scientific collaboration—e.g., telepresence to support and record (in 3D)distributed scientific experiments, and applications related to assisting the elderly—e.g., communication with family members, and continuous unobtrusive monitoring of mobility.

4. Acknowledgments

At UNC-Chapel Hill we want to acknowledge Jim Mahaney and John Thomas for their ongoing technical support. At the University College of Borås in Sweden we want to acknowledge graduate student Hanna Maurin for her tireless work on a series of formal evaluations of 3D telepresence applied to medical consultation.

This effort is primarily supported by National Library of Medicine contract N01-LM-3-3514: "3D Telepresence for Medical Consultation: Extending Medical Expertise Throughout, Between and Beyond Hospitals," and in part by NSF grant EIA-0303590: "Ubiquitous Pixels: Transforming Collaboration & Teaching with Pervasive Wall Displays."

References

[1] Point Grey Research, September 2005.

[2] 3rdTech. 3rdtech, July 15 2000.

[3] Bonnie Danette Allen and Greg Welch. A general method for comparing the expected performance of tracking and motion capture systems. In *Proceedings of 12th ACM Symposium on Virtual Reality Software and Technology (VRST)*, Monterey, California, November 2005. ACM SIGCHI & ACM SIGGRAPH.

[4] T. J. Coats and A. Goode. Towards improving prehospital trauma care. *Lancet*, 357(9274):2070., 2001.

[5] W. Fitzmaurice, George. Situated information spaces ad spatially aware palmtop computers. *Communications of the ACM*, 36(7), July 1993.

[6] W. Fitzmaurice, George and William Buxton. The chameleon: Spatially aware palmtop computers. In *ACM CHI94*, Boston, MA USA, 1994.

[7] Ken Hinckley, Randy Pausch, John C. Goble, and Neal F. Kassell. Passive real-world interface props for neurosurgical visualization. In *CHI '94: Proc. of the SIGCHI conference on Human factors in computing systems*, pages 452–458, New York, NY, USA, 1994. ACM Press.

[8] University of Washington HIT Lab. Artoolkit, September 2005.

[9] A. A. Meyer. Death and disability from injury: a global challenge. *J Trauma*, 44(1):1–12., 1998.

[10] Mathias Mohring, Christian Lessig, and Oliver Bimber. Video see-through AR on consumer cell-phones. In *Proc. of the Third IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'04)*, pages 252–253, Washington, DC, USA, 2004. IEEE Computer Society.

[11] W. Pasman, A. van der Schaaf, L. Lagendijk, R., and W. Jansen, F. Accurate overlaying for mobile augmented reality. *Computers and Graphics*, 23:875–881, 1999.

[12] F. B. Rogers, M. Ricci, M. Caputo, S. Shackford, K. Sartorelli, P. Callas, J. Dewell, and S. Daye. The use of telemedicine for real-time video consultation between trauma center and community hospital in a rural setting improves early trauma care: preliminary results. *J Trauma*, 51(6):1037–41., 2001.

[13] R. S. Smith. Telemedicine and trauma care. *South Med J*, 94(8):825–9., 2001.

[14] Andrei State, Greg Welch, and Adrian Ilie. An interactive camera placement and visibility simulator for image-based vr applications. In *Proceedings of The Engineering Reality of Virtual Reality 2006 (3D Imaging, Interaction, and Measurement; IS&T/SPIE* 18th *Annual Symposium on Electronic Imaging Science and Technology)*, San Jose, CA, January 2006.

[15] Jonathan Steuer. Defining virtual reality: Dimensions determining telepresence. *Journal of Communication*, 42(4):73–93, Autumn 1993.

[16] S. Tachakra. Depth perception in telemedical consultations. *Telemed J E Health*, 7(2):77–85., 2001.

[17] S. Tachakra, P. Jaye, J. Bak, J. Hayes, and A. Sivakumar. Supervising trauma life support by telemedicine. *J Telemed Telecare*, 6(Suppl 1):S7–11., 2000.

[18] S. Tachakra, A. Sivakumar, R. Everard, S. Mullett, and R. Freij. Remote trauma management–setting up a system. *J Telemed Telecare*, 2(Suppl 1):65–8., 1996.

[19] P. Taylor. A survey of research in telemedicine. 1: Telemedicine systems. *J Telemed Telecare*, 4(1):1–17, 1998. Using Smart Source Parsing.

[20] P. Taylor. A survey of research in telemedicine. 2: Telemedicine services. *J Telemed Telecare*, 4(2):63–71, 1998. Using Smart Source Parsing.

[21] Greg Welch, Diane Sonnenwald, Ketan Mayer-Patel, Ruigang Yang, Andrei State, Herman Towles, M.D. Bruce Cairns, and Henry Fuchs. Remote 3D medical consultation. *To appear in the Journal of Mobile Multimedia (JMM)*, 2005.

[22] Ruigang Yang. *View-Dependent Pixel Coloring—A Physically-Based Approach for 2D View Synthesis*. Ph.d. thesis, University of North Carolina at Chapel Hill, 2003.

[23] Ruigang Yang, Marc Pollefeys, Hua Yang, and Greg Welch. A unified approach to real-time, multi-resolution, multi-baseline 2d view synthesis and 3d depth estimation using commodity graphics hardware. *International Journal of Image and Graphics (IJIG)*, 4(4):1–25, 2004.