3D Telepresence for Medical Consultation: Extending Medical Expertise Throughout, Between and Beyond Hospitals

Agency No. N01-LM-3-3514

Final Report

March 17, 2008

Professor Henry Fuchs, P.I.*
Professor Bruce Cairns, M.D., Co-P.I.§
Professor Ketan Mayer-Patel, Co-P.I.*
Professor Diane Sonnenwald, Co-P.I.†
Professor Greg Welch, Co-P.I.*

** The Department of Computer Science*
The University of North Carolina at Chapel Hill
Chapel Hill, North Carolina 27599-3175, USA

§ The Department of Surgery
The University of North Carolina at Chapel Hill
Chapel Hill, North Carolina 27599-7228, USA

† The Swedish School of Library and Information Science
Göteborg University and the University College of Borås
501 90 Borås, Sweden

1 Introduction

Trauma, or serious physical injury, is a major health problem accountable for more productive years lost than heart disease, cancer and stroke combined [1, 2]. A trauma victim's recovery often depends on how soon the victim receives appropriate medical care. In general, paramedics are the primary medical personnel to provide health care to trauma victims at the scene of an accident. They diagnose each victim and, based on their diagnosis, perform medical procedures within their legal authorization and transport the victim to nearest appropriate hospital or medical facility. In complex trauma cases when the paramedic is having difficulty diagnosing the patient, needs advice regarding a procedure that should be performed, and/or needs legal permission to perform a specific procedure, paramedics will consult with a physician at a hospital or medical facility. Today paramedic physician consultation occurs via cell phone or radio. During this consultation, a paramedic must quickly and accurately do the following: verbally describe the victim, the accident scene and the victims symptoms; answer the consulting physicians questions; discuss treatment options with the physician; monitor the victims progress; and simultaneously perform complex medical procedures to save the victims life. All of these activities must often be performed within minutes. An incorrect description and subsequent decision and/or action could result in death, or further complications and a longer recovery time for the victim. Thus, visual technologies could be beneficial in emergency healthcare by providing the consulting physician with a directly transmitted view of the patient and accident scene when patients are severely injured or when there are long transport times to nearest hospital and the patient is in need of immediate care beyond the level paramedics are authorized to provide.

We report here on the progress of a multi-year effort to develop, apply, and evaluate technology for view-dependent *3D telepresence* technology for remote medical consultation. We refer to the approach as *three-dimensional medical consultation* (3DMC). Our primary aim is to enhance and expand medical diagnoses and treatment in a variety of life-critical trauma scenarios. Our long-term goal is to provide both a *advising* health care provider and a distant medical *advisee* with a high-fidelity visual and aural sense of 3D presence with each other. Primarily, but not exclusively, we envision scenarios involving extemporaneous consultation related to unforeseen events where time is critical, anxiety is high, and a physician or technician would welcome an expert consultation for concurrence or guidance in diagnosis and management, but physical presence is not possible.

Our hypothesis is that the shared sense of presence offered by view-dependent 3D telepresence technology will be superior to current 2D video technology, improving communication and trust between geographically-separated *advisor* and *advisee* medical personnel, enabling new extensions of medical expertise throughout, between, and beyond medical facilities. To validate this hypothesis, we designed our program to answer two fundamental questions: can we develop the technology for 3D telepresence in medicine, and will high-fidelity 3D telepresence be useful to the medical community? We have structured our effort to include corresponding technology *research* and *evaluation* components, along with a *systems integration* component designed to focus the research and facilitate the evaluation.

Our application-driven research efforts focus on the key technological barriers to 3D telepresence including real-time acquisition and novel view generation, network congestion and variability, and tracking and displays for producing accurate 3D depth cues and motion parallax. Our systems integration effort focuses our research and development on the scenarios (a)-(d) shown in Figure 1, which correspond to uses of permanent, portable, and hand-held advisor and advisee technologies. In an effort to balance usefulness and practicality, our evaluation effort is aimed specifically at assessing the fundamental effectiveness of *emergent airway management* in scenario (c)—where the advisor and advisee are using permanent but geographically-separated facilities. Our evaluation

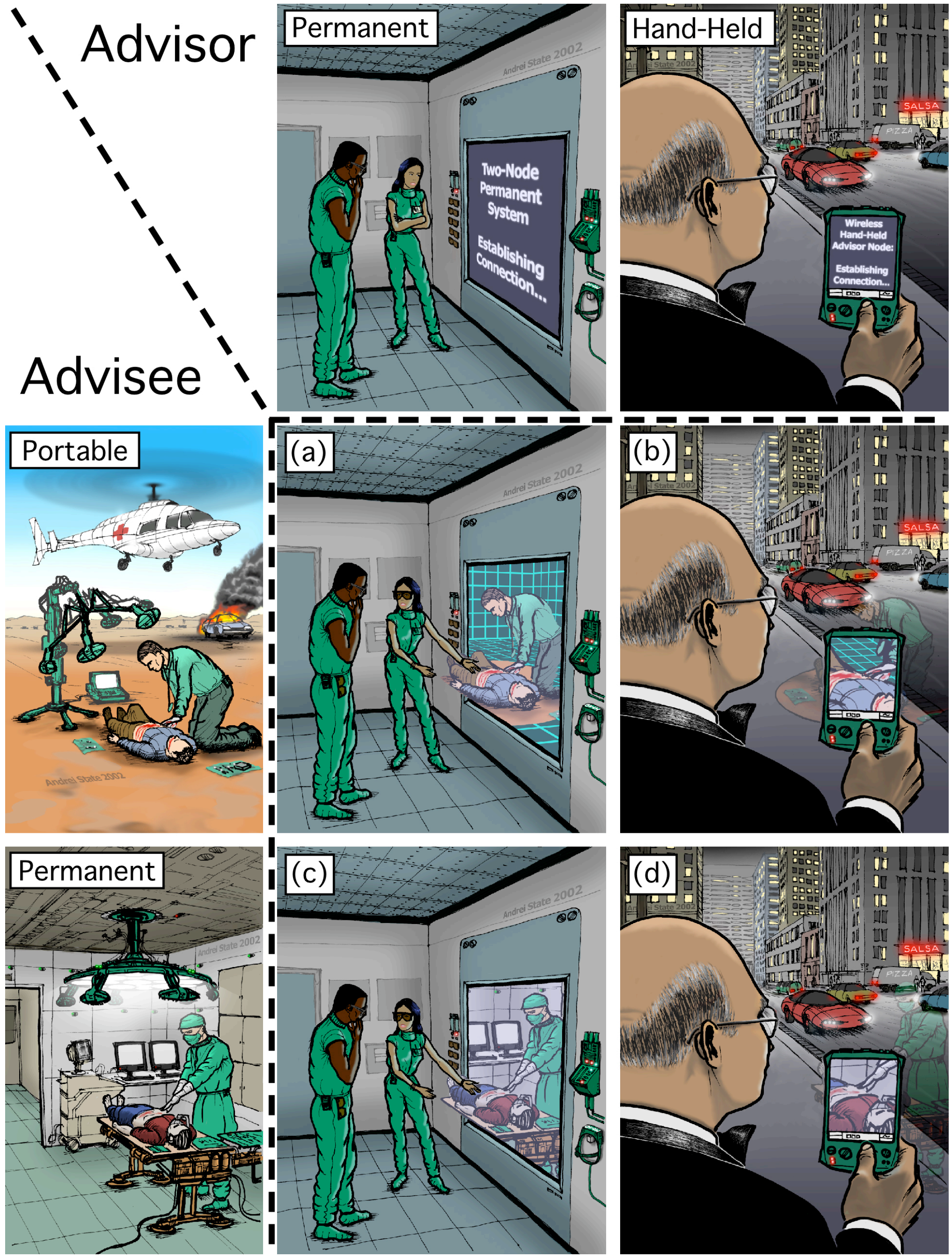

Figure 1: Future vision of 3D Telepresence for Medical Consultation. The left column illustrates examples of person-portable and permanent 3D telepresence technologies used by an *advisee*. The top row illustrates examples of permanent and hand-held technologies used by an *advisor*. Images (a)-(d) illustrate the shared sense of presence for corresponding advisor-advisee scenarios.

effort is aimed at investigating the potential quality of the diagnoses and treatment using 3D telepresence, and the *acceptability* of the technology by participating paramedics—an essential element for a successful 3DMC implementation.

Once the effectiveness of the system in controlled experiments has been established, we envision future efforts to adapt this technology for use in a variety of clinical scenarios, including remote hospital to tertiary center emergency consultations, portable in-transit diagnosis and stabilization systems (e.g., ambulance or helicopter transport), intraoperative consultations, and tumor boards.

In the remainder of this section we motivate the work with a discussion of traditional 2D medical consultation (tele-medicine), we then introduce the notion of 3D medical consultation, and explain why the concept is difficult to achieve in practice. In Section 2 we describe a proof-of-concept prototype system; in Section 3 we describe our research efforts and some visual results; and in Section 4 we discuss an ongoing formal evaluation.

1.1 2D Medical Consultation

Two-dimensional (2D) video-based medical consultation has been explored widely in the past 15–20 years, and there is renewed interest because of increasing concerns about natural and man-made disasters such as pandemics, hurricanes, and terrorist attacks. 2D televideo has been demonstrated to be acceptable for face-to-face consultation—supplementing the telephone, and useful for visual examinations of wounds, abrasions, etc. However with the latter use in particular, the two issues that seem to arise in the literature related to most relevant case studies are (a) the difficulty associated with obtaining the desired 2D camera views, and (b) depth perception.

For example, camera view difficulties were mentioned in multiple places in the final report for the U.S. National Library of Medicine's National Laboratory for the Study of Rural Telemedicine [3]. One example is in the discussion of the use of the a 2D televideo system to observe children with swallowing disorders. The report states

> "Limitations of telemedicine services for management of feeding and growth issues include the need to rely on the interpretations of others during physical exam. At times the camera angles were not ideal to allow for clear pictures of the mouth during feeding."

Similarly, included in the concerns identified with the university's "Clinical Studio" are the need for periodic movement of cameras and improper camera locations.

> "Full-motion video and audio of the neurological examination is a reliable means of visualizing the patient between remote locations. This technology is not difficult and can be done by ER staff. However the images are in two dimension hence certain aspects of the exam could be enhanced by more than one camera angle."

The problem was also identified in [4] where they describe work using a computer-based telemedicine system for semi- and non-urgent complaints at a short-term correctional facility.

> "The lack of remote control on the patient care camera at the remote site by the examining emergency medical physicians requires the nurse to spend considerable time operating the camera and responding to technical instructions. This problem has been resolved in a recent system upgrade, but it was another important reason for nonuse."

Beyond obtaining the desired 2D view of a remote patient, in [5] Tachakra states that "Impaired depth perception is a significant problem in telemedicine." and notes that "The most important cue of depth is due to binocular disparity." Tachakra describes several "coping strategies" that can be used to overcome the inherent limitation of the 2D imagery. Chief among the coping strategies is the practice of "Rotating the camera in the transverse plane about 30° at a time...." This is not surprising given that object occlusion and motion parallax are two of the most powerful depth cues. Tachakra surmises that this controlled camera rotation "...enables the consultant to build a three-dimensional mental image of the object by briefly storing a range of two-dimensional views."

There are two primary problems illustrated by the above examples. First, for *any* chosen configuration of remote 2D cameras, it is unlikely that the available views will always match the consulting physician's desired views. One could try and address the visibility problem using multiple cameras. But switching between numerous disjoint views as a security guard might with a surveillance system is not very natural. With a very large number of cameras and user head tracking, one could imagine automatic switching based on view position and orientation. But the quantity and configuration of cameras necessary to achieve smooth and appropriate switching over an operating room, as well as the 2D video storage and bandwidth needs, would be impractical. While pan-tilt-zoom cameras can help address this problem, they require additional technical skills, impose an additional cognitive load, and require additional time to adjust (difficult in a trauma situation).

Second, in cases where depth perception would aid in the consultation, users must resort to secondary visual cues or verbal clarification from a remote collaborator, which both impose additional cognitive loads compared to the very natural views afforded if the consulting physician were able to "be there" with the patient or the collaborating medical personnel.

With respect to face-to-face collaboration between medical personnel, previous research in computer supported cooperative work (e.g., [6, 7]) and theory of language [8] suggests that working remotely using 2D videoconferencing lacks the richness of collocation and face-to-face interaction, e.g., multiple and redundant communication channels, implicit cues, spatial co-references, that are difficult to support via computer-mediated communications. This lack of richness is thought to impair performance because it is more difficult to establish the common ground that enables individuals to understand the meaning of each others utterances. Other research (e.g., [9–11]) suggests that working remotely may not be compatible with existing work practices, and thus not adopted by individuals.

With respect to telemedicine, this limitation was cited in Georgetown University Medical Center's final report for their NLM-sponsored work under the NLM HPCC program. The report notes that in contrast to a face-to-face visit, the 2D technology limits the physicians view of the patient, and as a result some patients felt that the physician could not always "see" how the patient was "really doing."

1.2 3D Medical Consultation

To address the 2D problems outlined in Section 1.1 above, we are working on systems for *three-dimensional* medical consultation (3DMC). The basic idea is to use a relatively small number of cameras to "extract" (estimate) a time-varying 3D computer model of the remote environment and events. When coupled with head (or handheld viewer) position and orientation tracking, this should offer a remote consultant a continuum of dynamic views of the remote scene, with both direct and indirect depth cues through binocular stereo and head-motion parallax. See Figure 1 for example scenarios. We believe that some day in the future such 3D visual interfaces will be a standard part of mobile emergency patient care systems [12].

We believe such a system will provide the consulting physician with an increased *sense of presence* with the remote patient and medical personnel, thus improving communication and trust. There is evidence to support the increased sense of presence. In [13], the authors report on three studies where they vary display parameters and attempt to assess a user's sense of presence. The authors report that the first and second study results indicated that the reported level of presence was significantly higher when head tracking and stereoscopic cues were provided. They report that the third study indicated that the level of presence increased with the visual field of view.

There is also evidence to suggest that an immersive 3D display will increase a user's *performance* on certain tasks. For example [14] reported a moderately positive relationship between perceived presence and task performance. In [15] the authors present the results of a study where they found that users performing a generic pattern search task decrease task performance time by roughly half when they change from a stationary 2D display to a head-mounted (and tracked) 3D display with identical properties. In [16] the authors present the results of a study where distant collaborators attempted to solve a Rubik's cube type puzzle together. The authors compared face-to-face (real) task performance with networked performance using both a immersive 3D display and a conventional 2D desktop display. They found the task performance using the networked immersive 3D display and in the real scenario were very similar, whereas desktop performance was "much poorer." Most recently, in [17] the authors describe a careful 46-person user study aimed at determining whether or not immersive 3D virtual reality technology demonstrated a measurable advantage over more conventional 2D display methods when visualizing and interpreting complex 3D geometry. The authors found that the head-tracked 3D system showed a statistically significant advantage over a joystick-controlled 2D display.

1.3 Why is Obtaining 3D So Difficult?

The most common approach to 3D scene reconstruction is to use cameras and effectively "triangulate" points in the scene. This involves automatically picking some feature in one camera's 2D image, finding the same feature in a second camera, and then mathematically extending lines from the cameras into the scene. The place where the lines intersect corresponds to the 3D location of the feature in the room. If one can do this reliably for a sufficient number of points in the scene, many times per second, then with some assumptions about the scene, and a lot of compute power, one can turn the dynamic collection of disjoint 3D points into a coherent dynamic 3D computer model that one can use like a fight simulator.

However there are at least three areas of fundamental difficulty associated with trying to reconstruct dynamic 3D models of real scenes: feature visibility, feature quality, and reconstruction algorithms. Features might not exist or might be confusing/ambiguous, they are hard to detect, isolate, resolve, and correlate, and automating the overall reconstruction process in light of these difficulties is a very hard problem. The state of the art is limited to static environments for large spaces, or dynamic events in relatively small controlled spaces.

1.4 Project Status and This Article

The work we report on here is associated with a multi-year project sponsored by the U.S. National Institutes of Health's National Library of Medicine (Craig Locatis and Michael Ackerman). Currently the project includes research in real-time computer vision/graphics and network adaptation strategies, and a formal evaluation of the likely effectiveness and acceptance of 3D medical consultation paradigm. We have constructed a proof-of-concept prototype system and are in the process of completing the formal evaluation. Development of a production-quality system is beyond the scope of the project. (Part of our formal evaluation has been designed to indicate whether such future work is warranted.)

2 Prototype System

Here we briefly describe our proof-of-concept 3DMC prototype, to give the reader a concrete notion of the components used in the research described in Section 3. As shown in Figure 2, our prototype consists of multiple components that would be associated with the patient site and the remote consultant: a portable camera unit (a), a portable compute/rendering cluster (b), and two consultant display device paradigms (c) & (d).

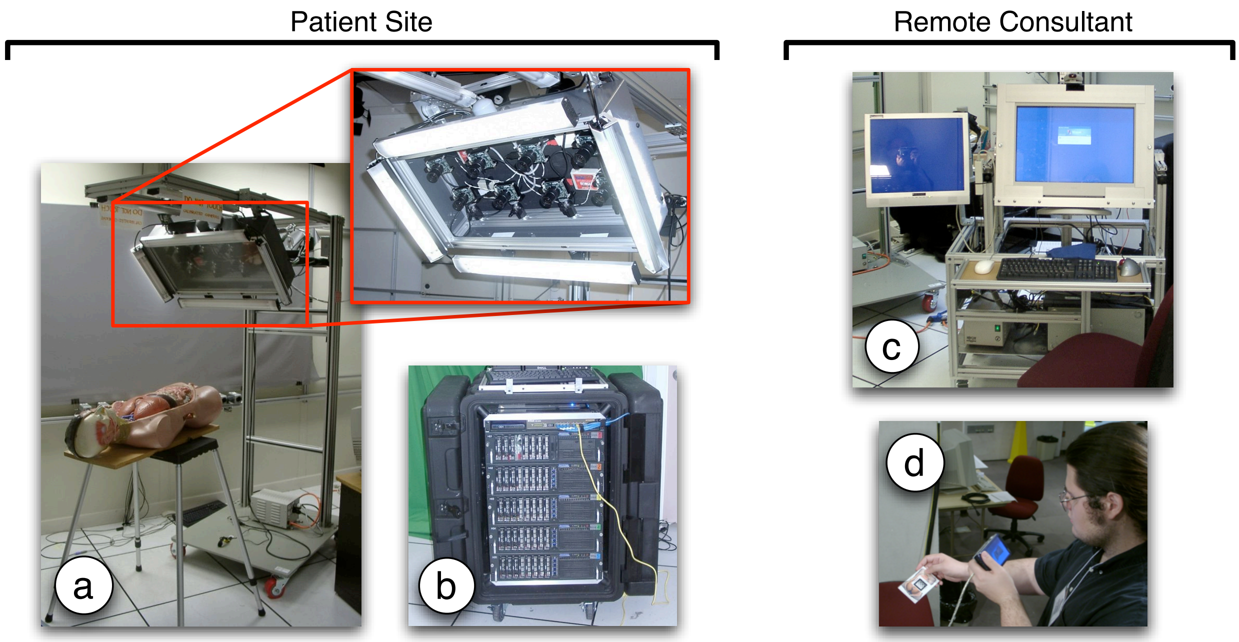

Figure 2: Our proof-of-concept 3DMC prototype, with patient site components on the left and remote consultant components on the right: (a) portable camera unit with eight Firewire cameras and high-frequency area lights; (b) compute cluster; (c) a transportable consultant viewing station with 2D and 3D (head tracked or autostereo) displays; (d) a tracked PDA mobile display.

The portable camera unit (PCU) shown in (a) of Figure 2 is a rolling unit holding a camera-lighting array with eight 640×480 resolution digital (IEEE 1394a) color cameras from Point Grey Research [18]. 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 region of interest on the patient. Mounted around the cameras are multiple Stocker-Yale high-frequency fluorescent fixtures for flicker-free illumination. The entire array is mounted on a rolling cart with adjustable length and angle, and significant weight (underneath the base) to prevent tipping. The PCU includes an AC isolation transformer (mounted on the base) to meet the current leakage requirements of UNC Hospital's medical engineering staff.

The compute/rendering cluster (b) in Figure 2 consists of five dual-processor servers in a transportable rack case. Four of the servers are connected to the PCU camera array via Firewire cables. These servers function as *Camera Servers*, compressing the PCU camera images and forwarding them via a dedicated gigabit Ethernet to the 5^{th} server. Each camera server can optionally record the video streams to disk. The 5^{th} server then decompresses the video streams, loading the color images into texture memory of the graphics card for view-dependent 3D reconstruction as described in Section 3.1. The unit also includes an AC isolation transformer. Note that because the PCU and the compute/rendering cluster in our prototype are connected via Firewire cables, they must generally be moved together. In the hospital the PCU is designed to be inside a patient room, and the cluster just outside the door. In a real production system (in the future) the PCU (a) and compute/rendering servers (b) could be combined into a single unit.

The consultant viewing station (c) in Figure 2 consists of a rolling cart with a dedicated server (lower shelf) that is connected to the compute/rendering cluster (b) by a single gigabit Ethernet cable. This Ethernet cable is the realization of the networking boundary shown in the middle of Figure 3. It is the only link between the compute/rendering cluster (b) and the consultant viewing station (c). The connection could be across the hospital or across the world. The station has a high-speed and high-resolution 2D monitor (top right of cart), an Origin Instruments opto-electronic tracker (top of the 2D monitor) to track the viewer's head position and orientation, and an *autostereoscopic* display mounted on an articulated arm (left)*. The unit also includes an AC isolation transformer.

Our current implementation of the tracked PDA mobile display (d) in Figure 2 uses a DragonFly camera [18] mounted on a Toshiba e800 PDA. The camera is currently attached to the rendering PC (above) via a Firewire cable, which uses ArToolKit [8] to compute the relative position and orientation of the PDA. (This is discussed further in Section 3.3.) The current prototype is not truly portable because of the wired (Firewire) link to a computer, so we plan on implementing the tracking on a PDA with a built in camera in the future. Wagner and Schmalstieg have ported and optimized ArToolKit for PDAs [19, 20], and although their results indicated that the primary bottleneck is image capture rate, new PDAs are coming out with cameras better suited to video rate capture. This would allow a wireless interface.

3 Research

Here we describe the three fundamental areas of research related to this project, including the computer vision methods for reconstruction of a 3D model of a dynamic scene (Section 3.1), two remote consultation display paradigms (Section 3.3), and some application-specific networking work (Section 3.4). These areas and their relationships are reflected in Figure 3 as follows: (a) and (b) on the left are associated with the reconstruction of a dynamic 3D model of the patient/procedure, (c) & (d) on the right are associated with the displays the remote consultant would use, and the networking research is associated with the geographical distance separation indicated in the middle.

3.1 3D Reconstruction

The 3D reconstruction process involves two major steps: the reconstruction of 3D *points* from 2D images and the reconstruction of 3D *surfaces* from the 3D points. To reconstruct 3D points from 2D images we use a novel approach called *View-dependent Pixel Coloring* [21]. 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.

As described in [22] we use the graphics hardware on the 5^{th} PC to perform the 3D reconstruction very quickly as the images arrive from the Camera Servers (Section 2). The basic idea is to use the

*Autostereoscopic displays provide one or more viewers with a fixed number of stereo views (for example eight) of a 3D scene, without the use of special user-worn glasses. See `http://www.newsight.com`.

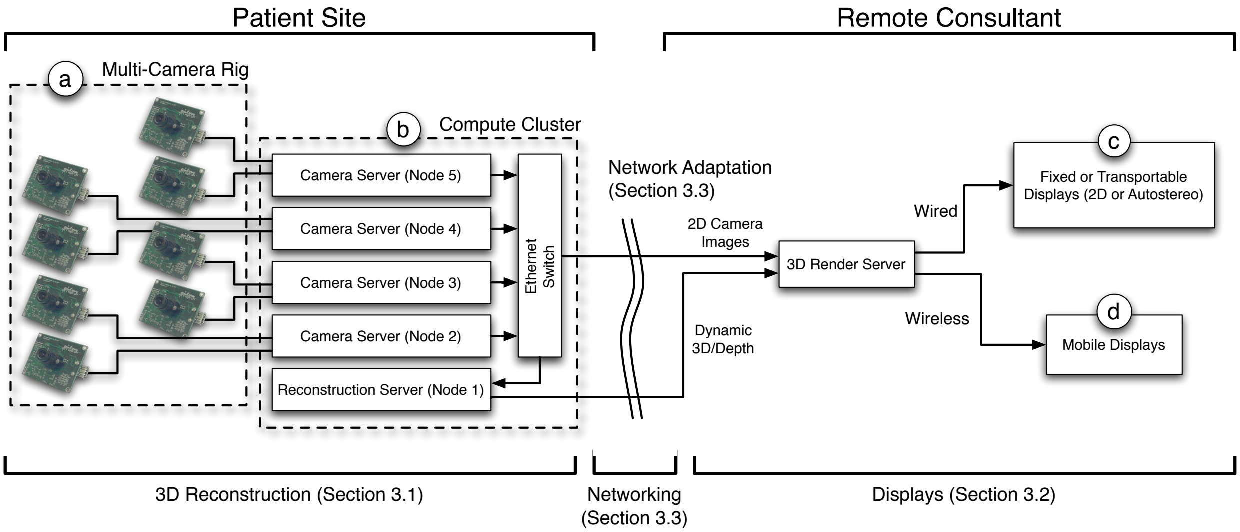

Figure 3: System diagram showing the patient site components on the left and the remote consultant components on the right: (a) rigid multi-camera rig with eight Firewire cameras; (b) compute cluster with four camera nodes and one 3D reconstruction node; (c) a fixed or transportable viewing station with 2D or 3D (head tracked or autostereo) displays; (d) a mobile display such as a tracked PDA.

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.

Figure 4 shows some results from our current prototype (Figure 2). Those views were reconstructed on line, in real time. Note that the views were reconstructed and rendered from completely novel view points—none the same as any of the cameras, at different times during the live sequence.

Figure 4: A 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.

3.2 Camera Planning

Historically computer vision researchers and practitioners have largely relied on rudimentary geometric analysis, experience, and intuition when making camera configuration choices such as the number and physical arrangement of cameras, sensor sizes, lens choices, frame rates, and shutter times. And yet these choices can have a tremendous impact on most applications. No estimation algorithm can overcome poor choices of devices, parameters, or geometric arrangement. If the necessary image samples are not available at a sufficient rate and quality throughout the desired working volume, the results will be inherently limited in those areas. In effect the device choices set an upper bound on how well the system as a whole will perform.

In some circumstances, for example if the camera configuration is relatively homogeneous and the design choices are correspondingly limited, one can attempt to *automatically* optimize the hardware design as in [23–25] and [26,27]. As the design becomes more complex, such optimizations become more or completely intractable. An alternative approach in these cases is to provide a human with an appropriate indication of the "cost" of a design, and then allow them to vary the parameters in an attempt to minimize the cost. This *intelligence amplification* approach [28, 29] has been used to evaluate specific designs [30] and is the basis for design tools like our Pandora [31], an interactive software simulator developed under this contract for human-in-the-loop optimization of multi-camera setups.

3.2.1 System-Level Optimization

In [32] we introduced a more general method for evaluating, comparing, and interactively optimizing the expected performance of tracking and motion capture systems. Given a candidate hardware design and some model for the expected user motion, we compute and interactively visualize estimates for the asymptotic (steady-state) uncertainty throughout the working volume. As with systems for the simulation and visualization of fluid or air flow, the idea is to modify the hardware design, looking for "hot" and "cold" spots in uncertainty, unusual shapes or variations in uncertainty, or uncertainty *sensitivities* to the hardware configuration. This interactive design optimization process is intended to augment traditional design and simulation practices, offering another tool to achive the best possible hardware design. See Figure 5 for some examples.

3.2.2 Measurement-Level Optimization

In [33] we introduced the notion of a *measurement*-level optimization. The idea is that in addition to an optimal device sample *rate* there is an optimal sample *duration*, for systems that sample devices over some non-zero time interval. For example, because cameras integrate light over a non-zero shutter time, estimating camera motion or dynamic scene structure using feature or color matching always involves a tradeoff between maximizing the signal (dynamic range or contrast) and minimizing the motion-induced noise (blur). If the shutter time is too short, the uncertainty in the measurements will be greater than necessary. If the shutter time is too long, the measurements will be corrupted by scene or camera motion. While some others have looked at this problem [34–36], as well as camera motion analysis [37, 38], our approach is general in that it applies to any sensors in general (not just cameras), and it fits nicely within the same stochastic framework as our system-level optimization.

3.3 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 (Figure 1, center). Our current prototype addresses this scenario with

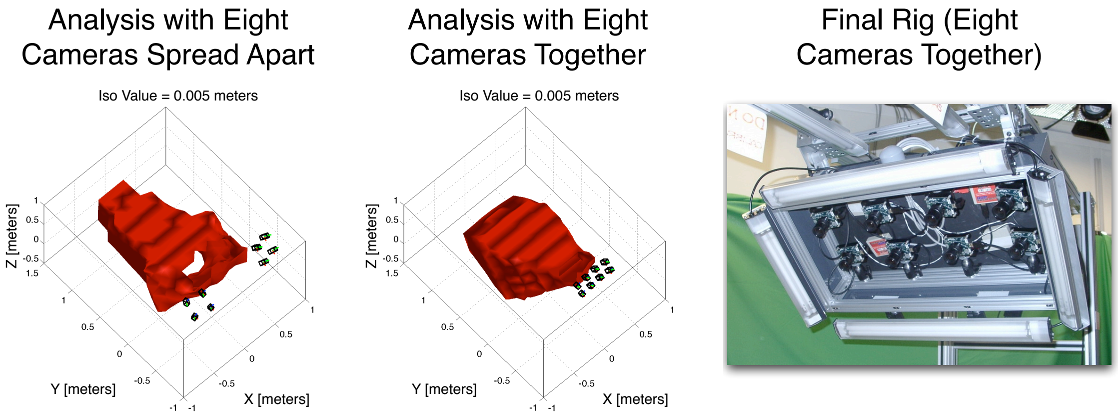

Figure 5: Examples of the 5mm steady-state uncertainty isosurfaces corresponding to a multi-camera rig with eight PtGrey DragonFly cameras spread apart into two groups of four (left) or arrayed together (middle). The corresponding final camera rig, with four high-frequency fluorescent fixtures, is shown on the right.

a high-resolution monitor and a system for tracking the viewer's head position and orientation. 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. Thus the user can observe the reconstructed view with natural/intuitive monoscopic head-motion parallax. We are also working on time-division multiplexing (shuttered) stereoscopic displays, and new autostereo displays that support multiple simultaneous viewers, with no glasses, and *independent* views.

We also want to provide the best possible 3D experience when the medical advisor is away from the hospital (Figure 1, right). For a remote display we are looking at *personal digital assistants* (PDAs). Most medical personnel are already accustomed to carrying a pager and mobile telephone, and some a personal digital assistant (PDA). 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" [39–42] providing a view of the remote patient, with natural interactive viewpoint control to help address occlusions and to provide some sense of depth.

We have investigated a two-handed patient "prop" paradigm as shown in Figure 6. Hinckley, et al. introduced the idea, using a dolls head or rubber ball and various tools as 'props' for neurosurgeons visualizing patient data [43]. Hinckley found that users could easily position their hands relative to one another quickly—a task we all do frequently. For 3D medical consultation 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 (PDA cover) in one hand and the PDA in the other, moving them around with respect to each other as needed to obtain the desired view. This paradigm provides the advisor with an instant visual target to aim their "magic lens" at, and 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, opening up a number of potential tracking solutions that were otherwise not feasible. This work is described in more detail in [44].

Figure 6: Left: Our first tracked PDA prototype used a HiBall-3000 ^{TM} tracking system [45], with sensors mounted on the PDA (Toshiba e800, left hand) and the surrogate (right hand). Right: Our current prototype uses a PointGrey DragonFly camera [18] 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 [46] to track the surrogate with respect to the PDA (camera).

Our current prototype has three main software components: a Tracking Server; a PDA Server (that also acts as a client to the Tracking Server); and a PDA Client. The Tracking server gets images from the PDA camera, and uses ARToolkit [46] to track the surrogate (PDA cover) with respect to the PDA. The PDA Server, which currently runs on the viewing station server (Section 2), continually gets a complete representation of the reconstructed data from the compute/rendering cluster (b) via a dedicated Ethernet connection as described in Section 2. It also obtains the estimated position and orientation of the PDA from the Tracking Server using the Virtual-Reality Peripheral Network (VRPN) protocol [47]. It then renders a view of the most recent reconstruction from the estimated PDA position and orientation, and compresses the view to send to the PDA. The PDA Client (running on the PDA) receives compressed images and displays them, as well as relaying user input back to the PDA Server, such as thumbwheel-controlled field-of-view settings. Each of these components may be run on the same or separate machines.

3.4 Networking

One of the central challenges we faced in developing 3DMC is the problem of adapting multiple video streams, audio, and other real-time continuous information streams to limited network and processor resources. Streaming media adaptation has been a widely studied area of research in other more mainstream application domains such as video-on-demand, video conferencing, voice-over-IP, and packet network television. In general, the problem involves accurately tracking current network conditions in order to estimate loss and throughput and consequently adjusting video quality and frame rate to match. The context of the application (i.e., live vs. pre-recorded, interactive vs. passive, open loop vs. closed loop, etc.) often determines which mechanisms and techniques can be effectively brought to bear on the problem. For the most part, state of the art solutions have cast the problem as an optimization problem within a rate-distortion framework.

Tele-immersion systems such as 3DMC raise a number of new challenges that must be addressed. While techniques and algorithms developed for more mainstream applications can still be brought to bear, the multi-stream nature of 3DMC complicates adaptation considerably. At issue is the need to negotiate inter-stream rate allocation decisions in a way that is meaningful to the overall application and aware of the semantic relationships between video streams. For example, an 8-camera 3DMC apparatus will need to dynamically aportion available bandwidth to different cameras depending on where medical personnel are looking, the task trying to be accomplished, the ease by which the reconstruction algorithm is able to extract structure from the visual information, the amount of overlap between the fields of view of each camera, and myriad other factors.

For each video stream in a multi-stream teleimmersion application, quality adaptation trade-offs exist not only between competing dimensions within a stream (e.g., frame rate vs. frame quality), but also between streams themselves (e.g., camera A vs. camera B). A central challenge for realizing an appropriate video adaptation in such complex, multi-stream environments, is the ability to compactly express and efficiently evaluate adaptation policy that must simultaneously negotiate all possible competing trade-offs. A rule-based approach simply does not scale and is brittle to any change in the system (e.g., number of cameras, additional adaptation dimensions, etc.). This insight has led us to embrace the following central design principle for our work on adaptation within 3DMC:

> **Principle of Multidimensional Adaptation**
> Multidimensional adaptation is best served by approximating the desired adaptation policy within a purely mechanical adaptation process.

Guided by this central principle, we have developed a utility-driven framework for addressing video adaptation in multi-stream environments. The central mechanism is a *utility graph* within which each application data unit (i.e., video frame) from each data source (i.e., camera) is modeled as a node in a directed graph. The utility graph exists within a multi-dimensional utility space. Each dimension is an axis of adaptation (e.g., time, resolution, area covered by viewpoint, etc.). The edges of the graph model data dependencies that arise due to differential encoding. A distance function relates the data needs of the user to the utility graph, and the adaptation framework simply chooses to transmit the data unit that maximizes utility per unit of cost (e.g., bandwidth). The graph evolves over time as the data sources produce new data units and is constantly re-evaluated as network conditions change.

Applications approximate the intended adaptation policy within our utility-driven framework via the design of the utility space and the formulation of the distance function. This is where expert knowledge of the application domain must be brought to bear in order to realize the appropriate adaptive behavior as well as possible. This can be as much a matter of art and experience as one of science and will necessarily be specific to the application. But, once an appropriate utility space and distance function have been specified, the evaluation of the numerous trade-offs possible becomes entirely mechanical. The utility graph and associated distance function are simply evaluated at any time an adaptation decision must be made. Doing so identifies the application data unit with the highest utility for the lowest cost. The key point is that the complexities of multidimensional adaptation in multi-stream environments such as 3DMC can be reduced to a mechanical process.

The main contributions we have been able to make with our work on adaptation in 3DMC include:

- Development of a utility-driven framework for expressing complex multistream adaptation relationships among data.

- Development of algorithms and metrics suitable for dynamically optimizing utility-cost trade-offs in real-time in a manner that adheres to our design principle of mechanical adaptation.
- Extensions to our framework for integrating the effects of packet loss and retransmission.
- Experimental validation and evaluation of our approach published in the literature.

In the remainder of this section, we describe in more detail our adaptation framework and its extension for handling packet loss. Significantly more detail can be found in [48, 49].

3.4.1 The Utility Space

The first step in using the adaptation framework is defining a *utility space*. Each dimension of the utility space is a possible axis of adaptation. Every *application data unit (ADU)* is associated with a specific point in this utility space. An ADU is an application-specific atomic unit of data that may potentially be chosen to be transmitted by the adaptation framework. Furthermore, the user's data interest is also represented as a point in this space. We call this point the user *point of interest (POI)*. The POI represents the ideal coordinates within the utility space of the most useful piece of information to the user. Thus, the utility of each ADU can be measured as the inverse of the distance (as measured by a specified distance function) between the coordinates of the ADU to those of the POI.

For example, in the eight camera 3DMC prototype we have four dimensions to the utility space: time (t), resolution (r), and two spatial dimensions (x and y). Each camera is associated with a point in the $x - y$ plane that represents the current center of its viewpoint. Each camera produces a layered, multiresolutional image every 1/30 of a second. Three resolutions are available: low, medium, and high. These resolutions are mapped to values along the resolution axis at r_{low}, r_{med}, and r_{high} respectively. Each resolution of each video frame from each camera defines a possible ADU in our system. This means that if Camera 1 (C1) currently has a viewpoint center located at (x_{C1}, y_{C1}), then at time t_0, it will produce three ADUs, one for each resolution of the video frame. The coordinates of the low resolution ADU within the utility space would be the tuple: $(x_{C1}, y_{C1}, t_{t0}, r_{low})$. The medium and high resolution versions of the video frame would have similar coordinates differing only in the value of the resolution component.

3.4.2 The Utility Graph

The collection of ADUs produced by all data sources is modeled as a *utility graph* embedded within the utility space. Each ADU is represented by a node in the graph located at the coordinates associated with the ADU in the utility space. Directed edges between nodes indicate encoding dependencies. An edge from node i to node j indicates that a dependency exists such that the ADU represented by node i must be available before the ADU represented by node j can be properly received and decoded. Furthermore, each edge is associated with a cost that represents the amount of the limiting resource required for the node to be decoded. The cost of decoding a node that does not depend on any other node is represented by a "self edge" that begins and ends at the same node.

Each node can be in one of three states: *resolved*, *available*, or *idle*. Resolved nodes represent ADUs that have already been selected and transmitted by the adaptation framework. Available nodes represent ADUs that could be usefully selected by the adaptation framework. This means that either on edge exists between a resolved node and the available node or that a self-edge for the available node exists. Idle nodes represent ADUs that are not useful because there exists some encoding dependency on a different node that has yet to be resolved.

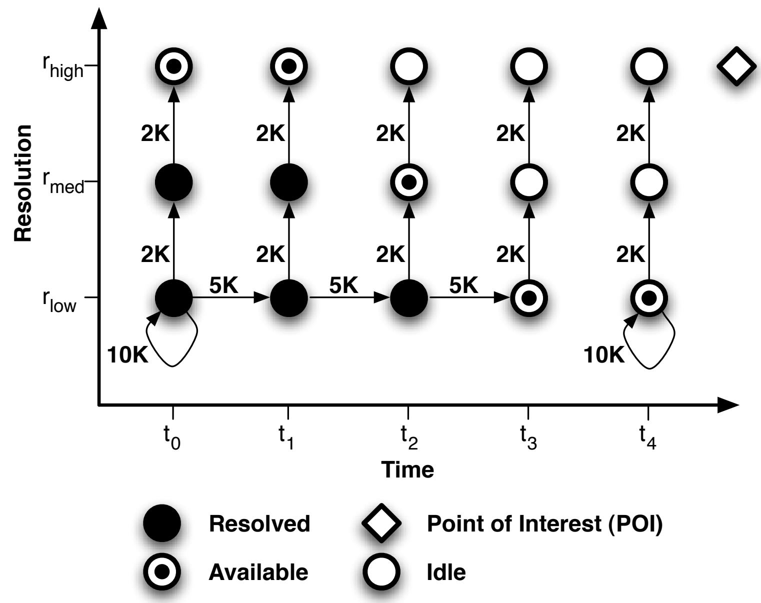

Figure 7: Example Utility Graph

In 3DMC, the high and medium resolution images are encoded differentially from the next lower resolution image of the same camera and time instance. Thus edges exist from nodes representing low resolution images to nodes representing the corresponding medium resolution images. Likewise, edges exist from medium resolution nodes to high resolution nodes. Furthermore, the low resolution images are differentially encoded temporally as well. Thus, there is an edge from each low resolution ADU node to the next low resolution ADU node in time from the same camera. The constrained resource is bandwidth and thus the cost associated with each edge is the number of bytes associated with the differential encoding of the frame. At fixed intervals, temporal differential encoding is **not** used and a low resolution image is produced that is entirely self contained. Those familiar with nomenclature used in the MPEG community will recognize these a *I-frames*. The size of these frames is represented by a self-edge at these nodes. I-frames are useful for limiting error propagation due to packet loss and providing periodic entries into the dependency chain induced by differential encoding.

3.4.3 Utility-Cost Ratio

The notions of utility and cost are fundamental concepts essential to adaptation. Individual elements are either more or less useful to an application than others. We define this notion as the utility of information. At the same time, access to a unit of data comes at some cost, often measured in time or required resources. Given a set of available elements, the process of adaptation attempts to maximize the utility of the received information and minimize the associated cost. When noted as a ratio of utility with respect to cost, adaptation becomes an attempt to maximize the utility-cost ratio, or UCR.

3.4.4 Mechanical Adaptation

Once the utility space and utility graph have been defined, adaptation decisions are simply made by the mechanical process of selecting the available node with greatest utility for least cost. Figure 7

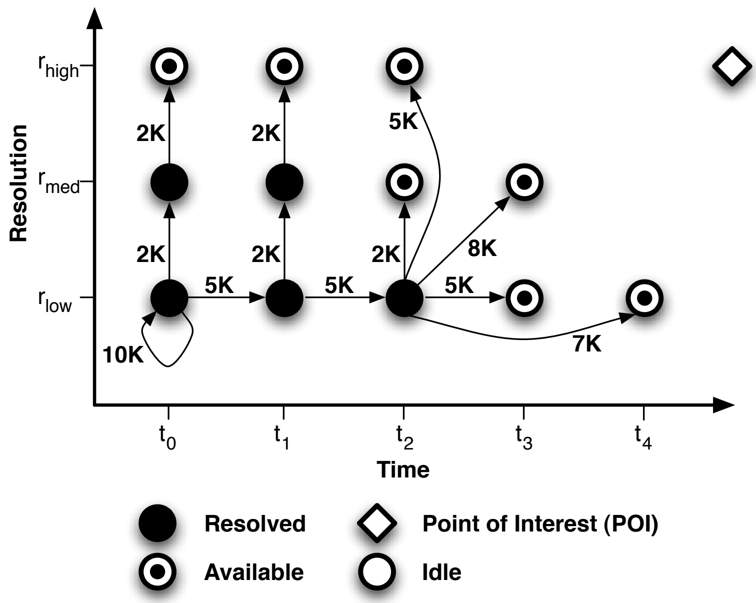

Figure 8: Example Utility Graph With Smart Encoding

presents an illustrative example. In this example, we have simplified the 4D utility space of our 3DMC system by eliminating the two spatial dimensions x and y and showing the utility graph for ADUs produced by a single camera in the two dimensions of time and resolution. Nodes that are resolved are filled solid, available nodes are filled with a bullet, and idle nodes are unfilled. Edges are labeled with their cost (i.e., number of bytes). The POI is represented by a diamond and is located to indicate that the ideal, most useful information to the user would be the most current, highest resolution image. In this example we can see five available nodes including high resolution frames at t_0 and t_1, a medium resolution frame at t_2, and low resolution frames at t_3 and t_4.

The adaptation framework makes the decision as to which node to select by calculating the *utility/cost ratio (UCR)*. The utility of each node is simply the inverse of the distance between the node and the POI. The cost of each node is the cost associated with the edge or self-edge leading to that node. We use a simple weighted Euclidean distance function where the weights allow the system developers to tune the trade-offs negotiated by the system between the different dimensions of the utility space. It is in this way that the distance function essentially embodies the notion of adaptation policy. The node with the highest UCR is selected. Once a node is selected, the system transmits the associated data, the status of the node is changed from available to resolved, and the status of any dependent idle nodes are updated. As the POI moves along the time dimension and new frames are produced, the utility graph is incrementally updated.

3.4.5 The Problem of Packet Loss

The problem of packet loss is that it is only some time in the future after an adaptation decision and possibly subsequent dependent decisions have been made that the system learns of the packet loss. If we do nothing to account for this new information, we may make suboptimal (or even completely useless) adaptation decisions because subsequent ADUs may have encoding dependencies that rely on the lost information. Indeed, a key insight in understanding why loss is so problematic is realizing that while differential encoding is a necessary source coding technique for increasing representational efficiency, it makes the resulting data stream brittle to channel errors and complicates adaptation by making it difficult to understand the true utility of a particular packet.

In fact, the utility of a packet sent as part of a prior ADU increases as subsequent encoding-dependent ADUs are chosen for transmission. This potential future utility of an ADU is not considered in the original decision because of uncertainty as to whether that potential will ever be realized. However, it is vital to consider an ADU's actual realized utility once packet loss has been discovered in order to appropriately account for any subsequent adaptation decisions that now depend on this ADU. Below, we present two possible approaches for dealing with the issue of packet loss within our adaptation framework while being true to our central design principle of purely mechanical utility-driven adaptation.

3.4.6 Smart Encoding

The first approach, called *Smart Encoding*, assumes full reliability semantics of the underlying transport-level protocol. The basic idea is to make differential encoding decisions that are fully aware of prior adaptation decisions. The result is that no ADU is ever dependent on a prior ADU that has not already been selected for transmission. Furthermore, no encoding relationship is ever jeopardized by packet loss because the reliability semantics of the transport-layer ensure that no permanent packet loss ever exists.

One benefit of this approach is that encoding dependency chains can be significantly lengthened since there is never cause for concern that a prior frame needed for reference is damaged. A drawback of the approach is that encoding efficiency is generally reduced whenever the temporal distance between a frame and its encoding basis is larger than one frame period. Another drawback is that while the transport-level protocol is delegated with the responsibility of reliable delivery, latency for such delivery is potentially unbounded.

Figure 8 illustrates our previous example utility graph shown in Figure 7 with Smart Encoding in effect. In this example, the current state of the utility graph (i.e., which prior ADUs have already been selected for transmission) is made available to the media encoders. The encoders in turn make new ADUs available for selection. These new ADUs, however, are always encoded relative to ADUs known to have been transmitted. In the resulting utility graph, we can see that many of the available nodes are now encoded relative to ADUs that are "farther" away. The effect of this encoding distance will naturally be seen in the cost of the encoding. For example, the low resolution node at t_4 is encoded relative to a frame that is 2 frame periods away and thus results in an inflated cost since temporal coherence is diminished. Similarly, the high resolution frame at t_2 directly uses the low resolution frame as an encoding basis instead of the medium resolution frame since the medium resolution frame has not yet been selected for transmission. Another feature of Smart Encoding is that encoders are able to delay generation of particular ADUs until after an adaptation decision is known. This is seen in the example by the lack of nodes representing high resolution frames at t_3 and t_4 and the medium resolution frame at t_4.

In summary, Smart Encoding allows the utility framework to deal with packet loss by employing a reliable transport-level protocol, thereby sidestepping the issue of packet loss altogether. Since ADUs selected for transmission are now guaranteed to be received, the framework is "smart" about how it encodes ADUs and leverages that knowledge by letting encoders extend dependency chains indefinitely and encode ADUs directly against prior ADUs known to have been transmitted.

3.4.7 Smart Reliability

Our second possible approach is to fold the decision of whether to retransmit the lost packet back into the utility-driven adaptation framework and evaluate the retransmission decision as a possible adaptation choice. To do so, whenever a lost packet is discovered, we add a self edge to the

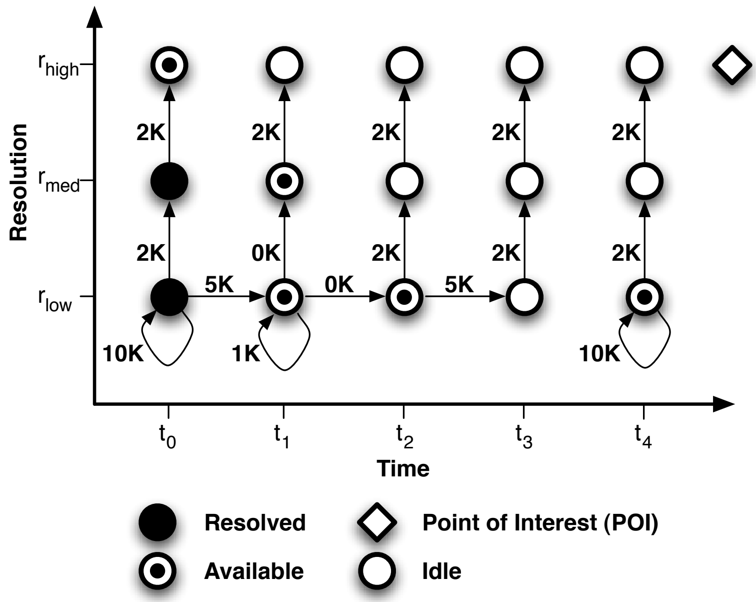

Figure 9: Example Utility Graph With Smart Reliability

afflicted node in the graph that represents its retransmission costs and mark this node as available. Furthermore, any subsequent dependent nodes that had already been selected for transmission are marked as available and all edges leading to them are given zero cost. When the next adaptation decision is made, the utility associated with the afflicted node is evaluated as the maximum utility of any node reachable from it with zero cost.

Figure 9 illustrates the example utility graph in Figure 7 with Smart Reliability. In this figure, we suppose that we have discovered that a packet from the low resolution ADU at time t_1 has been lost. The utility graph has been updated with a self-edge to represent the retransmission cost as well as zero cost edges to the other dependent ADUs that have already been transmitted. Other nodes that were previously available such as the high resolution image at t_1, the medium resolution image at t_2, and the low resolution node at t_3 must be marked as idle because they depend on ADUs that can not be decoded properly without first repairing the lost packet. At this point, the adaptation framework will decide between retransmitting the lost packet or transmitting packets associated with a non-dependent ADU such as the low resolution ADU at time t_4. The framework does so through the same mechanical process it has always used, namely evaluating each of the possible transmission choices and choosing the one that results in the greatest utility for the least cost.

One advantage of Smart Reliability is that because the utility of retransmission is evaluated at the time of retransmission against the utility of transmitting new independent data, the choice to retransmit can be abandoned if either there are few dependent ADUs (i.e., the potential utility was never realized) or if the utility of the dependent ADUs has been diminished due to latency (i.e., the frames are far enough in the past not to matter anymore). Smart Reliability provides a guard against the potential for unbounded latency found in Smart Encoding which relies on a fully reliable transport-level protocol to ensure the efficacy of its previous adaptation decisions. Furthermore, the encoders are once again independent of the adaptation framework and do not need to know about prior adaptation decisions in order to make encoding decisions.

Both of our approaches attempt to address the packet loss problem by incorporating a sort of joint source/channel coding decision. In the case of Smart Encoding, the encoding relationships exploit the reliability semantics of the underlying channel. In the case of Smart Reliability, the channel coding mechanisms (i.e., retransmission) are driven by the utility derived from previously known encoding relationships and subsequent adaptation decisions. In our experimentation we have found that while both approaches are effective, Smart Reliability generally outperforms Smart Encoding in most circumstances.

4 Evaluation

In this section we describe our work on evaluating the potential benefits of the emerging (not yet fully developed) 3D telepresence technology for medical consultation. Although we and many other groups are developing components of 3D telepresence techology, many millions of dollars of further investment will likely be required before the technology is sufficiently advanced for full-scale use. However as indicated in the National Academy's report on telemedicine evaluation [50], it is of critical importance to examine the acceptability and practicality of technology in medicine. Funding agencies, for example, would like to know today whether such a large investment of resources is merited, and avoid spending these resources if the technology has little potential to improve emergency health care. To address this issue traditional human-computer interaction experimental and field evaluation methods must be augmented with other techniques because the 3D telepresence technology does not yet exist. Thus we use a 3D proxy that simulates the 3D technology, and compared the quality of emergency trauma care provided when paramedics consulted with a physician via the 3D proxy, via 2D video, and worked alone.

4.1 Hypothesis and Approach

We propose that the use of 3D telepresence technology can improve the physicians understanding of a trauma victim and activities at a remote location, including diagnosis and treatment procedures performed by a paramedic at the remote scene. The improved understanding should enable the paramedic to provide better medical care. Thus our hypotheses are:

- H1: *Paramedics working in consultation with a physician via 3D telepresence technology will provide better medical care to trauma victims than paramedics working in consultation via 2D video or paramedics working alone.*
- H2: *Paramedics working in consultation with a physician via 3D telepresence technology will report higher levels of self-efficacy than paramedics working in consultation via 2D video or paramedics working alone.*
- H3: *Paramedics will rate the quality of interaction with the consulting physician higher in the 3D condition than in the 2D condition.*

Evaluating technology in emergency medical situations has unique challenges, many of which can be attributed to the complex context in which trauma situations occur. When arriving at the scene of an accident, paramedics must work quickly and accurately in order to save lives, and they must do this wherever the patient is located, which could be in a ditch next to a highway at night during a heavy rain storm. Patient healthcare priorities and privacy, together with the dynamics of emergency field care make it extremely difficult, if not impossible, to collect evaluation data in the field. Yet we should take into account as many aspects of emergency medical situations as possible during an evaluation to increase the validity of our results. Thus in our evaluation we simulated a realistic emergency medical scenario, and compared paramedics performance under three conditions working alone, working in consultation with a physician via 2D video technology, and working in consultation with a physician via a 3D proxy. In our post-test, between-subjects design [51] we also employed rigorously developed task performance measures in order to insure the validity and reliability of the evaluation results.

4.2 Trauma Simulation

One key aspect of the simulation design was choosing an appropriate trauma emergency medical task for the participating paramedics to perform. The task needed to have a high level of complexity, consisting of demanding medical decision-making and technical skills. It needed to be a task paramedics were qualified to perform with measurable outcomes, and a task they would seek advice on. In addition the task needed to be societally relevant, i.e., an important medical problem. The task we selected was the diagnosis and treatment of a trauma related *difficult airway*, including performing a surgical cricothyrotomy. In a surgical cricothyrotomy an incision is made through the skin in the neck and the underlying cricothyroid membrane to allow air to pass to the lungs. Paramedics are trained to manage the difficult airway and perform a surgical cricothyrotomy, yet this is extremely challenging. A surgical cricothyrotomy is usually regarded as the last resort in pre-hospital care situations as it can be very risky for the patient. When we asked paramedics about performing a surgical cricothyrotomy, they reported:

> "Cric[othyrotomy] skills scare the crap out of you."
>
> "It's something I hope I don't have to do."
>
> "It's difficult making that next step, the decision that you're gonna have to do a cric[othyrotomy]."

Even the most experienced physicians in airway management recognize the sense of urgency and anxiety associated with control of the difficult airway, because patients without an adequate airway will die within minutes if they do not receive appropriate treatment [52]. The inability to secure an airway is the most common cause of preventable death in pre-hospital care of injured patients [53].

Figure 10: Simulation setup.

A state-of-the-art computerized mannequin, the METI Human Patient Simulator™, was used to simulate the trauma patient suffering from a difficult airway. The mannequin (created by the manufacturers as a middle-aged Caucasian male) can simulate a wide range of medical conditions and responds to treatment in a life-like way. For example, its pupils dilate in response to light, its chest rises and falls when breathing, and its heart rate, breathing pattern and oxygen levels respond to drugs and medical procedures. The simulated trauma situation was a car accident in which a victim was thrown from a car and found moaning and groaning. Police had responded first on the scene and put the victim on a backboard as illustrated in Figure 10. To increase the ecological validity we added background sounds and lights that included traffic noises, a rotating beacon, peoples voices and emergency vehicle sirens approaching and leaving the scene.

4.3 Conditions

Figure 11: Consulting physician's view in the 2D condition.

As mentioned previously we compared paramedics performance and perceptions of working under three conditions: a paramedic working alone, a paramedic working in consultation with a physician via 2D video technology, and a paramedic working in consultation with a physician via a 3D proxy. In the remainder of this report these conditions are referred to as Alone, 2D and 3D proxy respectively. State-of-the-art, high quality 2D video was used in the 2D condition. Three views of the mannequin were provided to the remote consulting physician using digital cameras directly connected to three 21-inch high resolution monitors (Figure 11). One camera was a remote-controlled pan-tilt-zoom camera that the consulting physician could control. All cameras were placed in optimal positions for our particular trauma situation. That is, expert physicians determined the best locations for the three cameras to enable effective observation of the diagnosis and management of a difficult airway on the mannequin. Furthermore, the consulting physician had a full screen view of the patient monitor showing the mannequins heart rate, blood pressure and oxygen saturation rates in real time. The consulting physician observed the patient monitor and camera views in a custom built work station. In addition, the paramedic also had a 2D video view of the remote physician during the simulation. This view was located at the end of the mannequins stretcher.

Since the 3D telepresence technology is not yet sufficiently developed to allow us to use it even under ideal lab conditions, we designed a 3D proxy, or surrogate. For the 3D proxy, the consulting physician was physically present in the same room as the mannequin and paramedic. The consulting physician was allowed to freely move around in the room. However, the physician could not touch anything in the room and could only point to things using a laser pointer. This simulates the current vision and technical goals for 3D telepresence technology. As described earlier, the 3D technology should ideally enable the user to see a 3D representation of the remote scene and be able to dynamically change their viewpoint of that scene, mirroring the shared visual work space collaborators would have when working face-to-face. In our context it should allow the physician to virtually walk around the stretcher to get different views of the accident victim, or bend down to look more closely at the neck area. The physician should also be able to interact with the remote scene through a laser pointer that was displayed in his/her view and at the remote scene, e.g., to allow the physician to virtually point at the correct place for the paramedic to make the incision in the neck.

Social facilitation theory [54, 55] suggests that an individuals performance is affected by the physical (collocated) and virtual presence of an audience, in the sense that a person being observed by an audience will perform easier tasks better and more difficult tasks worse. This effect occurs when at least one person is present. If two or more people are in the audience, however, the impact is the same as if only one person is in the audience. In our experiment a physical (collocated) audience that consisted of a researcher-observer and a virtual audience that consisted of an expert mannequin operator observing the study participant from an adjacent room were present during all experiment sessions across all conditions. Social facilitation theory indicates that the physical presence of an additional physical or virtual audience member, i.e. the consulting physician in the 3D proxy and 2D conditions, would have no additional impact on paramedics performance. This is because all paramedics in the study had both a physical and virtual audience each consisting of at least one person no matter if they worked alone or with a physician via 2D or 3D proxy.

Two emergency care physicians acted as the consulting physicians in the 2D and 3D proxy conditions, with each physician participating in equal numbers across the two conditions. To help reduce the potential impact of any individual differences between the physicians, a physician paramedic interaction script was used by the physicians. The script consisted of appropriate, constructive phrasing of responses to typical questions and advice to give the paramedic at certain times during the scenario. The main reason for this was to minimize individual differences between the physicians regarding tone of voice and advice given to paramedics. The script was based on actual physician paramedic interaction observed during the pilot study and was developed in collaboration with the physicians.

4.4 Study Participants

To determine the optimal number of sessions to be conducted, we reviewed the literature for similar studies and found that 10 to 20 sessions per condition was common. See for example [56, 57]. Thus we had 20 participants per condition, for a total of 60. The 60 participants, 48 males and 12 females, were all certified paramedics working in southeastern US. In the US there are three certification levels for emergency medical technicians (EMTs): basic, intermediate, and paramedic. We chose the paramedic population because it is representative of individuals who perform the most advanced medical care among EMTs and their training and skill levels are somewhat consistent. For example, they are the only EMTs that are required to have training in performing a cricothyrotomy. Study participants were recruited through announcements on bulletin boards, email lists, newsgroups, personal emails, phone calls, advertisement in local papers, and through flyers posted in emergency medical services (EMS) offices, ambulance stations, and emergency rooms. They received $50 as appreciation for their participation in an experiment session which typically lasted 2 hours. The average years of total EMS work experience of the sixty paramedics who participated in our study was 11 years, with a range of 1 to 26 years. Their average paramedic work experience was 7 years, with a range of 1 to 24 years. The paramedics were randomly assigned across conditions, with equal distribution of gender and years of experience across all three conditions. Of all participants, 14 persons had previously performed a cricothyrotomy on a real patient, 7 in the Alone, 4 in the 2D, and 3 paramedics in the 3D proxy condition. In the 2D and 3D proxy conditions 5 and 6 paramedics respectively had met the physician before, a situation that mirrors daily work of paramedics. Paramedics usually have one or two specific hospitals where they take patients on a regular basis and where they know the some of the physicians in the ER. At other times, patients need to be transported to other hospitals or facilities where the paramedic does not know or has not met the physicians before.

4.5 Experiment Sessions

The medical simulation sessions took an average of 11 minutes, with a range of 6–23 minutes. During each session, the paramedic needed to diagnose and treat the victim (mannequin) as discussed earlier. Each simulation session was video-recorded using four cameras that captured paramedics actions on and surrounding the mannequin and medical monitor output (heart rate, oxygen saturation levels). The video recordings from each session were graded to evaluate the paramedics performance, using a grading protocol developed in collaboration with two physicians as discussed in the following sections. In addition, after each session each paramedic completed a questionnaire and participated in an interview , as described in the next section.

4.6 Evaluation Measures

4.6.1 Task Performance

Although management of a difficult airway, including performing a cricothyrotomy, is taught to paramedics and medical students throughout the world, a standard grading protocol does not exist in the published literature. To develop a grading protocol, we researched medical education literature (e.g., [58–60]) and previous research on performance assessment in simulated medical scenarios, e.g., [61–64]. The most common performance aspects to measure include combinations of: time for problem solving and decision making; technical, cognitive and behavioral skills; and number of appropriate/inappropriate procedures. The measures are often based on a recognized treatment algorithm, or in compliance with principles determined by medical researchers as good practice. Our main challenge was to construct an assessment protocol that could measure task performance comprehensively, based on the information that was possible to accurately estimate from the video recordings of each session. Our initial grading protocol was based on the ASA Practice Guidelines for Management of the Difficult Airway developed by American Society for Anesthesiologists [65]. The protocol was then discussed in-depth with two local emergency medical physicians who have decades of experience in managing difficult airways and performing surgical crics. The result was a grading protocol that captured the *performance time of key events*, *subtask execution*, and *harmful interventions*. Additional details regarding the grading protocol can be found in [66].

4.6.2 Self-Efficacy

Self-efficacy refers to the perceived capability to perform a certain task and is considered a powerful determinant of how well a task will be performed in the future [67]. Previous experiences, especially successful ones, are the strongest source of influence on self-efficacy, but also observing other peoples experiences can increase ones self-efficacy [67]. Thus a paramedics perceptions of self-efficacy after participating in our simulation can help predict the paramedics future performance in managing a difficult airway and performing a cricothyrotomy. If after consulting with a physician in the 3D proxy condition paramedics report higher levels of self-efficacy than the other paramedics, the theory of self-efficacy predicts that the paramedics with the higher levels of self-efficacy will actually perform better in the future, improving patient outcomes in the future. Since we did not find a self-efficacy questionnaire for difficult airway management or for performing a cricothyrotomy, we developed questions based on Banduras recommendations [68] and in consultation with emergency medical physicians.

4.6.3 Physician-Paramedic Interaction

Physician-paramedic interaction was measured by a post-question and post-interviews. The post-questionnaire included four questions derived from previous research that investigated interaction between dyads mediated by technology [69–71]. The questions focus on qualitative perceptions of the interaction because specific details can be difficult for study participants to recall. Physician-paramedic interaction was further measured by investigating the usefulness of information provided by the physician. Building on previous research by Levin and Cross [72], usefulness of information was measured through five items on the post-questionnaire that asked to what extent the information received from the consulting physician helped or hindered key aspects of diagnosing and treating the victim in the simulation and future victims.

To further our understanding of the post-questionnaire responses, each participant also participated in a semi-structured interview after their simulation session. The interviews ranged from 25–75 minutes in length, with an average time of 34 minutes. The interviews were recorded and later transcribed. Each interview consisted of open-ended questions focusing on the interaction with the physician during the simulation, the respondents perceptions of the simulated scenario, current practice, and future technology for remote consultation. As pointed out by Robson [73] the use of open-ended interview questions has many advantages: it allows greater flexibility; encourage respondents co-operation; and has the potential to produce unexpected answers.

The interviews were analyzed using both open coding and axial coding [74]. During open coding a subset of the interviews were read thoroughly and carefully by two researchers, and the researchers identified coding categories, or coding frames. After the initial set of categories was discussed among the researchers, another subset of interviews was analyzed. Additional coding categories emerged from this analysis. These codes were used to analyze a third set of interviews, and no new codes emerged, and the researchers were in agreement on the application of the codes. In the final step, i.e., during axial coding, all interviews were re-read and analyzed using the coding categories. For the purposes of this report, we report on the following codes: references to *Session Interaction* and *Future Technology*.

4.7 Results

Details regarding the statistical analyses described below can be found in [66, 75].

4.7.1 Ecological Validity

To investigate the participants perspective regarding the ecological validity of our simulation, i.e. how closely the simulation mirrors real world conditions, we included several questions in the post-questionnaire about ecological validity of the simulation and participants engagement during the simulation. All participants reported that they thought the simulation was realistic, they were intensely absorbed in the activity, and that they fully concentrated on the scenario (Table 1). An analysis of variance (ANOVA) yielded no significant differences between responses due to condition, suggesting that face validity was equally high across conditions.

	Result	
Question*	Mean	SD
The simulation was realistic	5.80	1.246
I was absorbed intensely in the activity	6.05	0.899
I concentrated fully on the activity	6.20	0.971
* Response Scale: 1 (strongly disagree) to 7 (strongly agree).		

Table 1: Ecological validity results.

4.7.2 Task Performance

Task Execution Times We compared average task performance times across conditions using an analysis of variance (ANOVA) test. No statistically significant differences were found between task performance times due to condition. Still, it is interesting to observe that for all tasks (except for airway incision—tube insertion) the 3D proxy condition tend to show less variation in task performance times than the other conditions. Figure 4 illustrates these differences between the conditions for total task and cricothyrotomy performance. Although there were no significant differences with respect to performance times between conditions, other factors such as years of professional experience, i.e., years of paramedic experience and years of total EMS experience (defined as years of experience as paramedic EMT, and basic or intermediate EMT) influence task performance times for the Alone and 2D conditions but not for the 3D proxy condition. This suggests that impact of professional experience on task performance may decrease with the use of 3D telepresence technology.

Subtask Performance An ANOVA test showed no statistically significant differences across the conditions for the execution of the manual mask ventilation or intubation subtasks. However, there were statistically significant results with respect to the cricothyrotomy subtask. Three paramedics in the Alone condition did not perform a cricothyrotomy. A Pearson Chi-Square test shows this to be statistically significant ($p = 0.043$) . In our simulation, as in many real life situations, the trauma victim dies when a cricothyrotomy is not performed when it is needed.

Harmful Interventions As shown in Table 2 the highest number of harmful interventions occurred when paramedics worked alone. Seven harmful interventions occurred when a paramedic worked in consultation with a physician via 2D video technology. In comparison only one harmful intervention occurred in the 3D proxy condition.

	Condition		
Harmful Intervention	Alone	2D	3D Proxy
Nasal intubation	1	1	0
Chest decompression	4^*	0^*	0^*
Not locating the cricothyroid membrane	3	1	0
Improper incision	3	1	0
Airway tube slippage	$0^\dagger$	$4^\dagger$	$1^\dagger$
Totals	*11*	*7*	*1*

* Result is statistically significant using a Pearson Chi-Square test ($p = 0.007$).

† Result is significant using a Pearson Chi-Square test ($p = 0.076$).

Table 2: Total number of harmful interventions performed.

4.7.3 Self-Efficacy

There are two categories of self-efficacy items: basic airway management tasks, and cricothyrotomy tasks. Basic airway management tasks are those tasks performed frequently by paramedics to insure their patients are getting oxygen into their lungs. Cricothyrotomy tasks are used less frequently, i.e. when basic airway management is not sufficient. ANOVA tests were performed to determine if there are statistically significant differences in responses due to conditions. The results are somewhat surprising. For all basic airway management tasks paramedics in the 2D condition reported *lower* levels of self-efficacy than paramedics working alone or in the 3D proxy condition. For one task, manually ventilating patients, the differences between paramedics working alone and in the 2D condition are statistically significant ($p \leq 0.05$) . For two other tasks, observing intubation problems and deciding on an alternative strategy when intubation fails, the differences are statistically significant at $p = 0.08$.

For all cricothyrotomy tasks, paramedics in the 2D condition again reported the lowest levels of self-efficacy. Paramedics in the 3D proxy condition reported the highest levels of self-efficacy. The differences in self-efficacy between the 2D and 3D proxy conditions are statistically significant at the $p = 0.05$ level.

Negative correlations between self-efficacy and years of professional experience were found in the Alone and 2D conditions (all statistically significant at $p = 0.05$. That is, the less work experience the paramedics in these two conditions had, the lower they rated their ability to treat the patient in the simulated scenario. There were no correlations between work experience and self-efficacy in the 3D proxy condition.

4.7.4 Physician-Paramedic Interaction

The data analysis shows that interaction between the physician and paramedic was statistically significant (at the $p \leq 0.05$ level) in one dimension, i.e. free-constrained. In fact, all physician-paramedic interaction in the 3D condition was perceived more positively than in the 2D condition but the differences for the dimensions, good-bad, accurate-distorted and easy-difficult, are not statistically significant. However, the difference with respect to the dimension of free-constrained is so profound, and the other interaction items so consistent, that when combined the differences between interaction in the 2D and 3D conditions is statistically significant. Overall, interaction in the 3D condition is viewed more positively.

Usefulness of information responses were analyzed using an analysis of variance (ANOVA) test comparing means between conditions. This analysis shows that participants believed that the information received from the consulting physician under the 3D condition, the 3D proxy, was statistically significantly (at the $p \leq 0.05$ level) more useful than the information received under the 2D condition. In particular, participants responses indicate that information provided regarding all key aspects of diagnosing and managing a difficult airway, except the intubation task, was more useful in the 3D proxy condition. We suspect this is because intubation is a task frequently performed by paramedics and thus they require little or no advice regarding this task.

Benefits from Paramedic-Physician Interaction Participants in both the 2D and 3D proxy conditions valued the interaction with the physician. Compared to their current way of working, collaborating with a physician over the radio, using video and a 3D proxy had several advantages: better patient healthcare was provided and paramedics improved their healthcare skills. As participants reported:

> "With him there we got a better airwayand increased the patients chance of survival." (3D proxy participant)
>
> "If I had been on the scene without a physicianit probably would have taken 1.5 or 2 minutes longer before [the patient] got a surgical airway." (2D participant)
>
> "Everything he was telling me was precise. I knew exactly what to look for. He explained the procedureI learned something." (3D proxy participant)
>
> "It helped me feel more confident already after [this one session]. (2D participant)

Challenges in Paramedic-Physician Interaction Participants in the 2D condition reported more challenges interacting with the physician than participants in the 3D Proxy condition. In the 2D condition, 15 participants (75%) described problems with the paramedic-physician communication. In comparison, 7 participants (35%) in the 3D Proxy condition reported problems. Of those 7, six focused on an initial awkwardness in the communication. Participants explained:

> "I'm humming along treating a patient then all of a sudden [there was] this voice over there and then the light, the laser goes beep and it just threw me for a second."

2D participants also reported an initial awkwardness:

> "When you start [interacting] there's an initial ice breaking thing... [an] initial awkwardness of it."

However, 2D participants reported four additional challenges. Some reported that the physician should be more assertive, more vocal and ask more questions during the diagnosis and treatment process. Others wanted the physician to give more feedback. Still others reported that the physician could not always see what needed to be seen and this made communication difficult, and one participant noted that he needed to change his behavior in the 2D condition. For example, 2D participants said:

> "I would prefer is that rather than prompt me to look at something, [the physician] were to say 'I think he's quit breathing' rather than 'how's his breathing doing?' "
>
> "[the physician] could have actually been even more vocal for example, reading what he's seen on the monitor"
>
> "I think it would be better if [the physicians] asked you questions. You know, initiated more."
>
> "I like feedback...I didn't get the feedback...more feedback would have been better."
>
> "[The physician] even asked me to move my hand at one time so he could actually see that I was in the right place could have used [some guidance] on [whether] I should go deeper with the scalpel because I wasn't sure."
>
> "I had to remember to give [the physician] more information...once you kinda' get into it and step back [you] realize, ok, I need to interact more and do this a little different."

With one exception (one 3D proxy reported also reported he wished the physician had been more assertive), all the challenges were reported only by 2D participants. We found no correlation between reports of these challenges and paramedics performance, paramedics years of professional experience or participating physician. Therefore these challenges appear to be implicitly linked to 2D video, and even under the best conditions (as in our experiment) the use of 2D video for emergency medical consultation will introduce communication problems- problems that appear not to emerge with 3D telepresence. Increased depth perception and the ability to dynamically change views appear to be important features of 3D telepresence technology. In addition to challenges arising form the lack of these features in 2D video, we often saw physicians changing their viewpoint during the experiment sessions—bending down to get a side-angle view, standing up on tiptoe and bending over the victim and paramedic hands. The physicians did not need to ask paramedics to move so the physician could see the patient better. The paramedic was free to focus on the medical task at hand, and did not need to worry about the physicians view.

Future Implementations of 3D Telepresence Technology to Support Physician-Paramedic Interaction Study participants mentioned social and technical features of 3D telepresence technology that are important to facilitate physician-paramedic interaction in emergency care situations. Participants reported that paramedics may feel threatened and intimidated because the 3D technology makes the paramedics performance visible to the physician and others in new ways. For example, participants told us:

> "It was nice that [the physician] was there and he had your back and he was going to walk you through it. But then again its kind of intimidating because you feel like you

> get trained to do this right you're scared you might mess up, and they say, we want you trained better than this." (3D proxy participant)
>
> "It kind of makes somebody nervous being monitored by a physician, someone of such higher training. And youre afraid to make a mistake because this person could be the person that ends up saying [whether] you get to do more, and where you work or not." (2D participant)

We recommend that best practice procedures should be implemented to eliminate or reduce the threat of intimidation introduced by the use of 3D telepresence technology. Possible best practices mentioned by participants included opportunities for paramedics and physicians to get to know one another personally and professionally, open and non-judgmental communication practices, and increased understanding regarding joint responsibilities and priorities in the

One question regarding the application of 2D video and 3D telepresence technology focuses on whether it is important for the paramedic to see the physician during the emergency consultation. Recall that in both conditions paramedics saw the physician. In the 2D condition the physician (face view) was shown on a screen at the end to the patients gurney. In the 3D proxy condition, the paramedic could see the physician face-to-face. We asked participants whether seeing the physician is useful. The results were mixed, with about half across both conditions reporting that seeing the physician is not useful and half reporting that it can be beneficial. Those participants not wishing to see the physician explained:

> "If it's a really bad patient, I would never look up. I would just start talking to him. As long as I could hear him, thats all I would care about." (2D participant)
>
> "I just want to be able to talk [to the physician]. I don't necessarily need a camera [to see him]. If he has the ability to see [the patient] I don't need to see him." (3D proxy participant)

Yet other participants explained that seeing the physician increased their confidence, reduced their stress, and provided additional learning opportunities.

> "To see his face, his facial expressions, to match his face expressions with the spoken word actually helped me with my confidence in that scenario.... Just having the face there in real time kinda helps me focus on what I needed to focus on.... And just help me to calm me down." (2D participant)
>
> "To see [the physician] not freaking out, it tends to make the medic not freak out so bad. If you see somebody calm then you're more bound to be calm." (3D proxy participant)
>
> "When you see peoples faces you could see better if they understand. They can see better if you are confused or if you visit something about something and why—they are not left to make assumptions." (2D participant)
>
> "If you can actually see the doctor and he can actually show you [what to do is easier] than...talking to a radio and trying to understand and make sense of what he's saying." (2D participant)
>
> "If [the physician] starts using some terms and stuff that we aint understanding, then he could show us." (3D proxy participant)

We found no correlation between wanting to see the physician or not and paramedics task performance or years of professional experience or participating physician. Similar to research on video-conference and phone meetings today, we do not know why some people prefer seeing the person they are interacting with, and others not.

We saw several features of our 3D proxy frequently utilized during the 3D proxy sessions. For example, physicians used the laser pointer frequently to identify the location and size of the required incision and to point to specific pieces of medical equipment that the paramedic needed to use. The paramedics paid attention to the physicians pointing. As Clarke [12] and other research have indicated the ability to point to physical objects facilitates mutual understanding and task completion. 3D proxy participants reported:

> "He was able to physically point to various things so that was an added bonus."
>
> "I liked that he was able to point and tell me what goes where and all."

4.8 Discussion

Results indicate there is partial support for hypothesis H1. Overall task performance is worst for paramedics working alone. Three paramedics in the alone condition did not perform a cricothyrotomy and a total of 19 harmful interventions were committed. Not performing a cricothyrotomy in the simulation (and in real life) leads to death, and performing harmful tasks may lead to further complications, permanent damage or death. A total of 7 harmful interventions were performed when paramedics consulted with a physician via 2D video, whereas only 1 harmful intervention was performed when the consultation occurred via the 3D proxy. Thus a paramedic working alone is the least desired condition and the 3D proxy condition is more desired from a patient healthcare perspective.

Although no statistically significant differences with respect to task performance times across groups emerged from the data analysis, the results suggest 3D telepresence technology may reduce variation in task performance time across all levels of experience among paramedics. Furthermore, only one (out of five) task performance time in the 3D conditions was influenced by years of professional experience. In comparison, three and four task performance times were influenced by years of professional experience when paramedics worked alone or in consultation via 2D video, respectively. Recall that paramedics were assigned to conditions based on their years of professional experience, such that there was an equal distribution of years of professional experience across all three conditions. Thus 3D technology might have the potential to reduce differences in diagnosis and treatment caused by differences in years of professional experience. For patients this could mean that they would receive the same high level of care regardless a paramedics years of professional experience.

The results show support for hypothesis H2. As discussed earlier perceptions regarding of self-efficacy predict future task performance [67]. Paramedics consulting via the 3D proxy reported the highest levels of self-efficacy. This suggests that the 3D telepresence technology might have a positive impact on future task performance.

In contrast, the paramedics consulting via 2D video reported the lowest levels of self-efficacy—even lower than paramedics working alone. These results, although not statistically significant, suggest that an important area for future research. We should investigate whether the use of 2D video-conferencing technology for emergency medical care actually harms paramedics future task performance. Could the physician-paramedic interaction during a 2D video consultation session—where the physician and paramedic must exchange basic information regarding the patients condition and actions undertaken by the paramedic—erode a paramedics confidence and his or her future performance? For example, during the 2D sessions we saw a physician asking a paramedic while trying to ascertain what was not working well: "Did you make the hole [surgical incision] big enough?" The physicians just could not see this important detail with the 2D cameras. However from a paramedics perspective the question is frustrating because no paramedic would deliberately make an incision too small. Perhaps physician-paramedic interaction that is perceived as counter-productive from the paramedics standpoint has a long term negative impact on task performance.

Hypothesis H3 that predicted paramedics would judge the quality of interaction with the physician as higher in the 3D condition was partially supported. The paramedics reported that their interaction was freer or less constrained in the 3D condition than in the 2D condition. In addition, all information provided by the physician, except that regarding intubation, was judged to be more useful in the 3D condition. Usefulness of information is an important aspect of emergency medical care because receiving useful information has an impact not only on current task performance but also future task performance. However, the interaction was not necessarily better, more accurate or easier. This result based on questionnaire responses is explicated in the post-interview data. Paramedics in both the 2D and 3D proxy conditions reported benefits from collaborating with the physician, and an initial awkwardness in the interaction. However, 2D participants reported additional challenges, including needing the physician to be more assertive, making direct observations, asking more questions and providing more feedback, and having to change their behaviour so the physician could see and understand their situation better.

In summary, 3D telepresence technology shows potential to improve patient healthcare in complex emergency medical situations. Our results further illustrate that social practices, such as developing a culture of open communication among physicians and paramedics, and technology features, such as remote laser pointing and dynamic views, should be included in any future implementations of 3D telepresence technology for emergency healthcare. Future research includes conducting a field, or interview, study that examines issues of acceptability and practicality of 3D telepresence technology with a variety of stakeholders in emergency healthcare, including emergency room nurses, physicians and directors, as well as hospital administrators.

5 Investigating the Potential Impact of 3D Telepresence in the Context of Emergency Health Care Practices and Policies in the U.S.

The experimental evaluation described in Section 4 was augmented by a second study which investigates the acceptability and practicality of the technology within the context of the emergency medicine culture and healthcare system. As indicated in the National Academys report on telemedicine evaluation [50], it is of critical importance to examine the acceptability and practicality of telemedicine technology. Many factors may impact acceptability and practicality, including professional culture and image and health care system structure [50]. Often these factors, which may negatively impact and even halt adoption, are not discovered until the technology is deployed in operational settings at large costs. There is a need to investigate these factors earlier to identify improvements to the technology which could enhance its acceptability and practicality to guide policy makers.

To investigate such factors we conducted semi-structured interviews and focus group interviews with actors in the emergency healthcare system who could influence and be influenced by remote 3D remote medical consultation. We interviewed 24 professionals, including:

- Emergency room physicians in a major hospital and remote medical center (the head of emergency medicine department, county medical director, assistant county medical director,

and family medicine physicians working in the ER at remote medical center)
- Trauma physician
- Emergency room residents
- Emergency room nurses (ER nurse, nurse manager, nurse educator) in a major hospital and remote medical center
- Emergency department administrator (responsible for strategic and financial planning, budget, marketing etc) in a major hospital
- IT managers in a major hospital
- Emergency room IT manager and tech support
- Medicare administrators
- Clinical directors of emergency medical services in a major hospital and remote medical center
- EMS operations manager, in charge of county EMS
- State medical director responsible for paramedic training, education, scope of practice and performance improvement on a state level.

During the interviews, a short video that included artist renderings demonstrating our vision of telepresence technology in emergency medical situations was used to introduce the technology to study participants. We encouraged each participant to discuss the technology in the context of their work using open-ended questions. The interview questions was guided by theories such as innovation adoption [76], socio-technical change (e.g., [77]), and technology acceptance and use [78]. Specifically we asked participants to discuss their perspectives regarding:

- the benefits and disadvantages of the technology for patients, emergency healthcare professionals, their department, their organization, and the healthcare industry;
- features the technology would need to provide in order to be accepted and used;
- features of the technology that might stop it from being used;
- how the technology might change their current way of working;
- how the technology might change the way emergency healthcare is provided features of the technology that might stop it from being used;
- whether the technology would be a status symbol for their organization; and
- how decisions are made regarding the purchase and implementation of new technology in their organization.

The interviews ranged from 24 to 110 minutes in length, with an average length of 50 minutes. Gender distribution among participants was relatively equal (11 women and 13 men). The interviews were digitally recorded and transcribed. The interview data was analyzed using both open coding and axial coding [79]. During open coding interviews were read thoroughly and carefully to identify coding categories, or themes. After an initial set of categories was developed and discussed among the research team, two researchers analyzed interview data to determine if additional coding categories were needed and if there was consensus regarding data coding. An agreement of 0.82 using Cohens Kappa intercoder reliability measure was reached. A Kappa value above 0.75 is considered excellent [73].

Data analysis indicates that across all job categories, interview participants saw many potential benefits of 3D remote consultation. Benefits mentioned include: improvement of patient care for specific, complex emergency procedures; skill development for emergency services staff; improvement of patient care in remote areas, disaster areas, and developing countries; documentation of field procedures to protect medical staff from law suits and to help subsequent doctors treat the patient; a marketing tool and status symbol to attract patients and their families; a way to move resources back to communities and possibly allow patients to stay in their communities instead of being transported to remote clinics or hospitals; and, a tool to facilitate research on emergency health care in the field.

However there are many challenges facing the adoption of 3D remote medical consultation. In addition to the technical challenges discussed elsewhere in this report, there are challenges with respect to multiple infrastructures, including: telecommunications infrastructure (security, reliability, priority, quality); EMS infrastructure (medical protocols, technology readiness, technology cost financing, legal authority); IT hospital operations infrastructure (support of remote technology, technology operations costs); public and private insurance healthcare infrastructure (billing and legal responsibilities among hospitals, EMS providers and regional clinics); and ER infrastructure (physician availability, privacy, billing, legal responsibility and authority.) Participants mentioned some solutions to these challenges, including research trials that would provide strong clinical evidence of benefits. Further detailed results from the data analysis are being documented in referred publications.

Acknowledgements

We want to acknowledge James Manning, M.D. and Eugene Freid, M.D. for their valuable medical collaboration on the project; Andrei State and Herman Towles for their technical collaboration; Graduate Research Assistants B. Danette Allan, Adrian Ilie, Hanna Maurin Söderholm†, Vincent Noel, Michael Noland, Max Smolens and Hua Yang for their contributions to the effort; and Jim Mahaney and John Thomas for their technical support.

References

[1] T. Coates and A. Goode, "Towards improving prehospital trauma care," *Lancet*, vol. 357, no. 9274, p. 2070, 2001.

[2] A. Meyer, "Death and disability from injury: A global challenge," *J Trauma*, vol. 44, no. 1, pp. 1–12, 1998.

[3] M. Michael G. Kienzle, "Rural-academic integration: Iowas national laboratory for the study of rural telemedicine," tech. rep., National Laboratory for the Study of Rural Telemedicine, 2000.

[4] M. David G. Ellis, P. James Mayrose, M. Dietrich V. Jehle, M. Ronald M. Moscati, and M. Guillermo J. Pierluisi, "A telemedicine model for emergency care in a short-term correctional facility," *Telemedicine Journal and e-Health*, vol. 7, no. 2, pp. 87–92, 2001.

[5] S. Tachakra, "Depth perception in telemedical consultations," *Telemed J E Health*, vol. 7, no. 2, pp. 77–85., 2001.

[6] P. Dourish, A. Adler, V. Bellotti, and A. Henderson, "Your place or mine? learning from long-term use of audio-video communication," *Computer Supported Cooperative Work*, vol. 5, no. 1, pp. 33–62, 1996.

[7] G. Olson and J. Olson, "Distance matters," *Human-Computer Interaction*, vol. 15, no. 2-3, pp. 139–178, 2000.

[8] H. Clarke, *Using Language*. Cambridge, UK: Cambridge University Press, 1996.

[9] S. L. Star and K. Ruhleder, "Steps toward an ecology of infrastructure: Problems of design and access in large information systems," *Information Systems Research*, vol. 7, pp. 111–134, 1996.

[10] W. Orlikowski, "Learning from notes: Organizational issues in groupware implementation," *The Information Society*, vol. 9, no. 3, pp. 237–252, 1993.

[11] J. S. Olson and S. Teasley, "Groupware in the wild: lessons learned from a year of virtual collocation," in *Proc. of the 1996 ACM conference on Computer supported cooperative work*, pp. 419–427, ACM Press, 1996.

[12] Y. Chu, X. Huang, and A. Ganz, "Wista: A wireless transmission system for disaster patient care," in *Proceedings of BROADNETS: 2nd IEEE/CreateNet International Conference on Broadband Networks*, (Boston, MA, USA), pp. 118–122, Omnipress, October 2005.

[13] C. Hendrix and W. Barfield, "Presence within virtual environments as a function of visual display parameters," *Presence: Teloperators and virtual environments*, vol. 5, no. 3, pp. 274–289, 1996.

[14] M. P. Snow, *Charting Presence in Virtual Environments and Its Effects on Performance*. PhD thesis, Virginia Polytechnic Institute and State University, December 1996.

[15] R. Pausch, M. A. Shackelford, and D. Proffitt, "A user study comparing head-mounted and stationary displays," in *Proc. of IEEE Symposium on Research Frontiers in Virtual Reality*, pp. 41–45, IEEE Press, 1993.

[16] R. Schroeder, A. Steed, A.-S. Axelsson, I. Heldal, A. Abelin, J. Wideström, A. Nilsson, and M. Slater, "Collaborating in networked immersive spaces: As good as being there together?," *Computers & Graphics, Special Issue on Mixed Realities - Beyond Conventions*, vol. 25, pp. 781–788, October 2001.

[17] D. W. Mizell, S. P. Jones, M. Slater, B. Spanlang, and D. Swapp, "Immersive virtual reality vs. "flat-screen" visualization: A measurable advantage," *(Submitted for publication.)*, 2003.

[18] "Point Grey Research." `http://www.ptgrey.com/`, September 2005.

[19] D. Wagner and D. Schmalstieg, "First steps towards handheld augmented reality," in *ISWC '03: Proceedings of the 7th IEEE International Symposium on Wearable Computers International Symposium on Wearable Computers*, (Washington, DC, USA), IEEE Computer Society, October 2003.

[20] D. Wagner and D. Schmalstieg, "Handheld augmented reality displays," in *Proceedings of 2nd Emerging Display Technologies Workshop (EDT 2006)* (G. Welch, ed.), pp. 35–36, 2006.

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

[22] R. Yang, M. Pollefeys, H. Yang, and G. 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)*, vol. 4, no. 4, pp. 1–25, 2004.

[23] X. Chen, *Design of Many-Camera Tracking Systems For Scalability and Efficient Resource Allocation*. PhD thesis, Stanford University, 2002.

[24] G. Olague and R. Mohr, "Optimal camera placement for accurate reconstruction," *Pattern Recognition*, vol. 35, no. 4, pp. 927–944, 2002.

[25] U. Erdem and S. Sclaroff, "Optimal placement of cameras in floorplans to satisfy task requirements and cost constraints," in *OMNIVIS*, p. workshop, 2004.

[26] L. Davis, E. Clarkson, and J. P. Rowland, "Predicting accuracy in pose estimation for marker-based tracking," in *Proceedings of Second IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'04)*, pp. 28–35, Institute of Electrical and Electronics Engineers, IEEE Computer Society Press, October 2003.

[27] L. Davis, F. Hamza-Lup, and J. P. Rowland, "A method for designing marker-based tracking probes," in *Proceedings of Second IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'04)*, Institute of Electrical and Electronics Engineers, IEEE Computer Society Press, November 2003.

[28] R. Ashby, *Design for an Intelligence-Amplifier*, vol. 34 of *Annals of Mathematics Studies, Automata Studies*. Princeton University Press, 1956.

[29] R. Ashby, *Design For a Brain*. New York City, N. Y.: John Wiley & Sons, 2nd ed. ed., 1960.

[30] E. Foxlin, M. Harrington, and G. Pfeifer, "Constellation: A wide-range wireless motion-tracking system for augmented reality and virtual set applications," in *Computer Graphics* (M. F. Cohen, ed.), Annual Conference on Computer Graphics & Interactive Techniques, pp. 371–378, Orlando, FL USA: ACM Press, Addison-Wesley, SIGGRAPH 98 conference proceedings ed., 1998.

[31] A. State, G. Welch, and A. Ilie, "An interactive camera placement and visibility simulator for image-based vr applications," in *Proceedings of the Engineering Reality of Virtual Reality 2006*, (San Jose, CA), IS&T/SPIE 18th Annual Symposium on Electronic Imaging Science and Technology, January 2006.

[32] B. D. Allen and G. Welch, "A general method for comparing the expected performance of tracking and motion capture systems," in *VRST '05: Proceedings of the ACM symposium on Virtual reality software and technology*, (New York, NY, USA), pp. 201–210, ACM Press, 2005.

[33] G. Welch, B. D. Allen, A. Ilie, and G. Bishop, "Measurement sample time optimization for human motion tracking/capture systems," in *Proc. IEEE VR 2007 Workshop on 'Trends and Issues in Tracking for Virtual Environments'* (G. Zachmann, ed.), (Charlotte, NC, USA), IEEE, Shaker Verlag, Aachen, Germany, March 11 2007.

[34] M. Tsuzuki, T. Naito, and S. Yamamoto, "Image acquisition method with wide dynamic range of moving objects for its," in *Proceedings of the IEEE International Vehicle Electronics Conference, 2001 (IVEC 2001)*, pp. 187–192, September 25–28 2001.

[35] S. Yuan, A. Devor, D. A. Boas, and A. K. Dunn, "Determination of optimal exposure time for imaging of blood flow changes with laser speckle contrast imaging," *Applied Optics*, vol. 44, no. 10, pp. 1823–1830, 2005.

[36] D. W. Tyler, A. H. Suzuki, M. A. von Bokern, D. D. Keating, and M. C. Roggemann, "Optimal SNR exposure time for speckle imaging: experimental results with frequency-dependent detector noise," in *Proc. SPIE Vol. 2198, p. 1389-1397, Instrumentation in Astronomy VIII, David L. Crawford; Eric R. Craine; Eds.* (D. L. Crawford and E. R. Craine, eds.), vol. 2198 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, pp. 1389–1397, June 1994.

[37] W. Sohn and N. D. Kehtarnavaz, "Analysis of camera movement errors in vision-based vehicle tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 1, pp. 57–61, 1995.

[38] S.-W. Shih, Y.-P. Hung, and W.-S. Lin, "Accuracy analysis on the estimation of camera parameters for active vision systems," in *ICPR '96: Proceedings of the 1996 International Conference on Pattern Recognition (ICPR '96) Volume I*, (Washington, DC, USA), p. 930, IEEE Computer Society, 1996.

[39] G. W. Fitzmaurice, "Situated information spaces ad spatially aware palmtop computers," *Communications of the ACM*, vol. 36, July 1993.

[40] G. W. Fitzmaurice and W. Buxton, "The chameleon: Spatially aware palmtop computers," in *ACM CHI94*, (Boston, MA USA), 1994.

[41] M. Mohring, C. Lessig, and O. 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)*, (Washington, DC, USA), pp. 252–253, IEEE Computer Society, 2004.

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

[43] K. Hinckley, R. Pausch, J. C. Goble, and N. F. Kassell, "Passive real-world interface props for neurosurgical visualization," in *CHI '94: Proceedings of the SIGCHI conference on Human factors in computing systems*, (New York, NY, USA), pp. 452–458, ACM Press, 1994.

[44] G. Welch, M. Noland, and G. Bishop, "Complementary tracking and two-handed interaction for remote 3d medical consultation with a pda," in *Proceedings of Trends and Issues in Tracking for Virtual Environments, Workshop at the IEEE Virtual Reality 2007 Conference* (G. Zachmann, ed.), (Charlotte, NC USA), Shaker, March 11 2007.

[45] "3rdTech, Inc.." `http://www.3rdtech.com/`, September 2005.

[46] "Artoolkit." `http://www.hitl.washington.edu/artoolkit/`, September 2005.

[47] "Virtual reality peripheral network." `http://www.cs.unc.edu/Research/vrpn/`, November 2006.

[48] D. Gotz and K. Mayer-Patel, "A general framework for multidimensional adaptation," in *Proceedings of ACM Multimedia 2004*, October 2004.

[49] S. Krishnan and K. Mayer-Patel, "A utility-driven frmaework for loss and encoding aware video adaptation," in *Proceedings of ACM Multimedia 2007*, September 2007.

[50] M. Field, ed., *Telemedicine: A guide to assessing telecommunications for health care.* Washington, DC: National Academy Press, 1996.

[51] W. Shadish, T. Cook, and D. Campbell, *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton Mifflin, 2002.

[52] A. Gawande, *Complications.* New York: Henry Holt, 2001.

[53] A. Bair, E. Panacek, D. Wisner, R. Bales, and J. Sakes, "Cricothyrotomy," *J Emerg Med*, vol. 24, no. 2, pp. 151–156, 2003.

[54] E. Bradner and G. Mark, "Social presence with video and application sharing," in *GROUP '01 Conf*, pp. 154–161, New York: ACM Press, 2001.

[55] N. Cottrell, D. Wack, G. Sekerak, and R. Rittle, "Social facilitation of dominant responses by the presence of an audience and the mere presence of others," *J Pers. Soc. Psycho*, vol. 9, no. 3, pp. 245–250, 1968.

[56] C. Gale, "The effects of gaze awareness on dialog in a video-based collaborative manipulation task," in *Proc. CHI '98 Conf*, pp. 345–346, New York: ACM Press, 1998.

[57] C. Gutwin and M. A. Roseman, "Usability study of workspace awareness widgets," in *Proc. CHI '96 Conf*, New York: ACM Press, 1996.

[58] J. Ali, R. Cohen, R. Adam, T. Gana, I. Pierre, H. Bedaysie, E. Ali, U. West, and J. Winn, "Teaching effectiveness of the advanced trauma life support program as demonstrated by an objective structured clinical examination for practicing physicians," *World J Surg*, vol. 20, no. 8, pp. 1121–1126, 1996.

[59] D. Chapman, K. Rhee, J. Marx, B. Honigman, E. Panacek, D. Martinez, B. Brofeldt, and S. Cavanaugh, "Open thoracotomy procedural competency: Validity study of teaching and assessment modalities.," *Ann. Emerg. Med*, vol. 28, no. 6, pp. 641–647, 1996.

[60] D. Johnson, D. Macias, A. Dunlap, M. Hauswald, and D. Doezema, "A new approach to teaching prehospital trauma care to paramedic students," *Ann Emerg Med*, vol. 33, no. 1, pp. 1477–1490, 1999.

[61] A. Byrne and J. Greaves, "Assessment instruments used during anaesthetic simulation: review of published studies," *Br. J. Anaesth*, vol. 86, pp. 445–450, 2001.

[62] J.-H. Devitt, M.-M. Kurrek, M.-M. Cohen, K. Fish, P. Fish, A.-G. Noel, and J.-P. Szalai, "Testing internal consistency and construct validity during evaluation of performance in a patient simulator," *Anesth-Analg*, vol. 86, pp. 1160–1164, 1998.

[63] D. Gaba, S. Howard, K. Fish, B. Smith, and Y. Sowb, "Simulation-based training in anesthesia crisis resource management (acrm): A decade of experience," *Simulation & Gaming*, vol. 32, no. 2, pp. 175–193, 2001.

[64] K. Stringer, S. Bajenov, and M. Yentis, "Training in airway management," *Anaesthesia*, vol. 57, no. 10, pp. 967–983, 2002.

[65] A. S. of Anesthesiologists Task Force on Difficult Airway Management, "Practice guidelines for management of the difficult airway," *Anesthesiology*, vol. 98, pp. 1269–1277, 2003.

[66] H. Söderholm, D. Sonnenwald, B. Cairns, J. Manning, G. Welch, and H. Fuchs, "The potential impact of 3d telepresence technology on task performance in emergency trauma care," in *Proceedings of the ACM Group 2007 Conference*, New York: ACM Press, 2007.

[67] A. Bandura, *Self-Efficacy: The Exercise of Control*. New York: H. Freeman, 1997.

[68] A. Bandura, "Guide for constructing self-efficacy scales," tech. rep., Available from Frank Pajares, Emory University, 2001.

[69] J. Short, E. Williams, and B. Christie, *The Social Psychology of Telecommunications*. New York: Wiley, 1976.

[70] L. Chidambaram and B. Jones, "Impact of communication medium and computer support on group perceptions and performance: A comparison of face-to-face and dispersed meetings," *MIS Quarterly*, vol. 17, no. 4, pp. 465–491, 1993.

[71] D. H. Sonnenwald, M. C. Whitton, and K. L. Maglaughlin, "Evaluating a scientific collaboratory: Results of a controlled experiment," *ACM Trans. Comput.-Hum. Interact.*, vol. 10, no. 2, pp. 150–176, 2003.

[72] D. Levin and R. Cross, "The strength of weak ties you can trust: The mediating role of trust in effective knowledge transfer," *Management Science*, vol. 50, no. 11, pp. 1477–1490, 2004.

[73] C. Robson, *Real World Research*. Massachusetts: Blackwell, 2002.

[74] B. L. Berg, *Qualitative Research Methods for the Social Sciences*. Boston: Allyn and Bacon, 1989.

[75] D. H. Sonnenwald, H. Maurin, B. Cairns, E. Freid, J. Manning, G. Welch, and H. Fuchs, "Experimental comparison of the use of 2d and 3d telepresence technologies in distributed emergency medical situations," in *Proceedings of the American Society of Information Science and Technology (ASIS&T 2006)* (A. Grove, ed.), (Austin, Texas), American Society of Information Science and Technology, November 3–9 2006.

[76] E. Rogers, *Diffusion of Innovations*. New York, NY, USA: The Free Press, 1995.

[77] W. Bijker, *Of bicycles, bakelit and bulbs: Towards a theory of socio-technical change*. Boston, MA, USA: MIT Press, 1995.

[78] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "User acceptance of information technology: Toward a unified view," *MIS Quarterly*, vol. 27, no. 3, pp. 425–478, 2003.

[79] B. L. Berg, *Qualitative research methods for the social sciences*. Boston, MA, USA: Allyn and Bacon, 1989.