Chapter 8
Remote 3D Medical Consultation

Greg Welch, Diane H. Sonnenwald, Henry Fuchs, Bruce Cairns, M.D., Ketan Mayer-Patel, Ruigang Yang, Andrei State, Herman Towles, Adrian Ilie, Srinivas Krishnan, and Hanna M. Söderholm

Abstract Two-dimensional (2D) video-based telemedical consultation has been explored widely in the past 15–20 years. Two issues that seem to arise in most relevant case studies are the difficulty associated with obtaining the desired 2D camera views, and poor depth perception. To address these problems we are exploring the use of a small array of cameras to synthesize a spatially continuous range of dynamic three-dimensional (3D) views of a remote environment and events. The 3D views can be sent across wired or wireless networks to remote viewers with fixed displays or mobile devices such as a personal digital assistant (PDA). The viewpoints could be specified manually or automatically via user head or PDA tracking, giving the remote viewer virtual head- or hand-slaved (PDA-based) remote cameras for mono or stereo viewing. We call this idea remote 3D medical consultation (3DMC). In this article we motivate and explain the vision for 3D medical

G. Welch (✉), H. Fuchs, K.M.-Patel, A. State, H. Towles, A. Ilie, and S. Krishnan
The Department of Computer Science, The University of North Carolina at Chapel Hill, Campus Box 3175, Chapel Hill, North Carolina 27599-3175, USA
e-mail: welch@cs.unc.edu, fuchs@cs.unc.edu, kmp@cs.unc.edu, andrei@cs.unc.edu, herman@cs.unc.edu, adyilie@cs.unc.edu, krishnan@cs.unc.edu

D.H. Sonnenwald
School of Information and Library Studies, University College Dublin, Belfeild, Dublin 4, Ireland
e-mail: Diane.Sonnenwald@ucd.ie

B. Cairns, M.D.
The Department of Surgery, The University of North Carolina at Chapel Hill, Campus Box 7228, Chapel Hill, North Carolina 27599-7228, USA
e-mail: bruce_cairns@med.unc.edu

R. Yang
The Department of Computer Science, The University of Kentucky, 232 Hardymon Building, Lexington, Kentucky 40506-0195, USA
e-mail: ryang@cs.uky.edu

H.M. Söderholm
The Swedish School of Library and Information Science, Göteborg University and the University College of Borås, 50190 Borås, Sweden
e-mail: Hanna.Maurin@hb.se

S. Coquillart et al. (eds.), *Virtual Realities*, DOI 10.1007/978-3-211-99178-7_8,

consultation; we describe the relevant computer vision/graphics, display, and networking research; we present a proof-of-concept prototype system; and we present some early experimental results supporting the general hypothesis that 3D remote medical consultation could offer benefits over conventional 2D televideo.

8.1 Introduction

We report here on a multi-year project to develop and evaluate technology for view-dependent 3D telepresence technology to support medical consultation across geographic distances. We refer to this technology as *three-dimensional medical consultation* (3DMC). Our long-term vision is to enhance and expand medical diagnoses and treatment in life-critical trauma situations. Our goal is to connect an advising health care provider, such as an emergency room physician, with a distant medical advisee and patient, such as a paramedic treating a trauma victim, using view-dependent 3D telepresence technology to provide a high-fidelity visual and aural sense of presence such that they can more effectively communicate and share information when diagnosing and treating the patient (see Fig. 8.1). Primarily, but not exclusively, we envision the technology enabling better patient healthcare through extemporaneous medical consultation across geographic distances in dynamic situations where patient diagnosis and treatment is time-critical and complex, but physical co-presence of medical experts and patients is not possible.

The basic technical idea for 3DMC 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 consultant a continuum of dynamic views of the remote scene, with both direct and indirect depth cues through binocular stereo and head-motion parallax. Example scenarios are illustrated in Fig. 8.1. We believe that some day in the future such 3D technology could be a standard part of mobile emergency patient care systems (e.g., [6]) that today use 2D video technology.

We hypothesize 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 medical personnel, enabling new opportunities to share medical expertise throughout, between, and beyond medical facilities. To investigate this hypothesis our research addressed two fundamental questions: can we develop the technology for 3D telepresence in medicine, and will the technology be useful to the medical community? Thus our research consisted of three inter-related components: 3D technology research, prototype development, and evaluation.

Our 3D technology research explored the key technological barriers to 3D telepresence today, including real-time acquisition and novel view generation, tracking and displays for producing accurate 3D depth cues and motion parallax, and network congestion and variability. Our prototype development synthesized the results from our technology research efforts to create a system that aims to provide permanent,

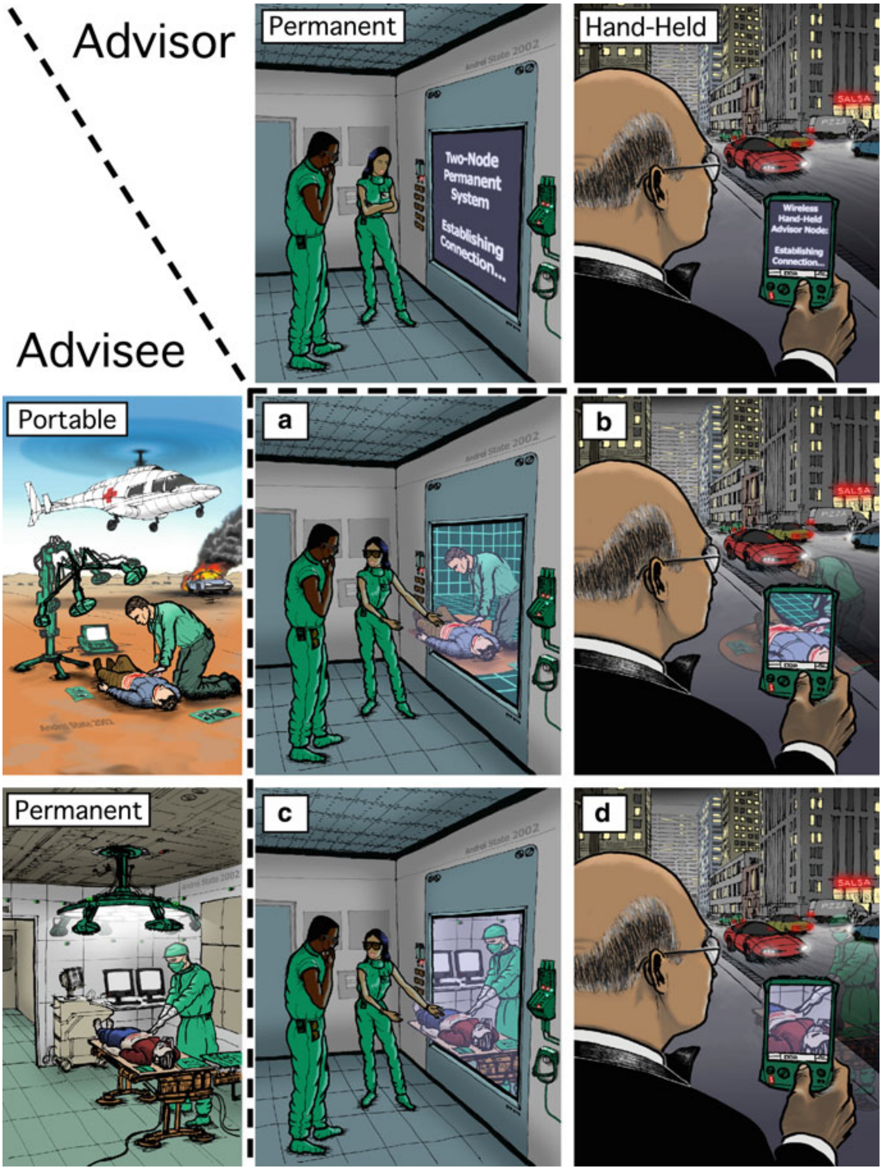

Fig. 8.1 Future vision of 3D telepresence for medical consultation. The *left column* illustrates examples of person-portable and permanent 3D telepresence technologies used by an *advisee*. The *top row* illustrates examples of permanent and hand-held technologies used by an *advisor*. Images (**a**)–(**d**) illustrate the shared sense of presence for corresponding advisor-advisee scenarios

portable, and hand-held access to 3DMC. Our evaluation effort assessed the potential effectiveness of 3DMC. Although some early results from each component have been reported elsewhere in conference papers, this is the first time all project components are presented in a holistic manner.

8.1.1 Medical Consultation via Video Technology

Medical consultation using two-dimensional (2D) video-conferencing and televideo technology has been explored in a variety of medical settings, such home-based health care [9, 20], prison-based healthcare [8, 10] and rural health care [23, 43]. Two limitations with respect to the technology arise repeatedly in the literature: the difficulty associated with obtaining the desired camera views and depth perception.

For example, camera view difficulties were mentioned in multiple places in the final report for the US National Library of Medicine's National Laboratory for the Study of Rural Telemedicine [23]. One example is in the discussion of the use of the 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 exams. At times the camera angles were not ideal to allow for clear pictures of the mouth during feeding" [23, p. 110].

The problem was also identified in [8] 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" [8, p. 92].

Patients have also found this same limitation in 2D video technology. Georgetown University Medical Center [26] reports that in contrast to a face-to-face visit, the use of 2D video technology limits the physician's view of the patient, and as a result patients felt that the physician could not always "see" how the patient was "really doing."

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 or feasible in time-critical health care situations. 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 (which is difficult in a trauma situation).

In addition to the challenges in obtaining the desired 2D view of a remote patient, 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" [36, p. 77]. Similarly, a university "Clinical Studio" which used video conferencing to perform neurological examinations reported: "[Videoconferencing] technology is not difficult and can be [handled] by [Emergency Room] staff. However the images are in two-dimensions hence certain aspects of the exam could be enhanced by more than one camera angle" [36, p. 187].

In situations 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 and with the collaborating medical personnel. Tachakra describes several "coping strategies" that can be used to overcome the inherent limitation of 2D imagery. Chief among the coping strategies is the practice of "rotating the camera in the transverse plane about 30° at a time." 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" [36, p. 83]. This is not surprising given that object occlusion and motion parallax are two of the most powerful depth cues.

However it is often not realistic to require camera rotation as prescribed by Tachakra in emergency, time-critical health care situations in the field. For example, the time needed to rotate a camera and view the rotating imagery reduces the amount of time available to perform life-saving procedures. It reduces the number of on-site personnel who can provide assistance to a trauma victim as it requires the full-time effort of a trained on-site person. And, in some situations it may be physically very difficult to rotate a camera, e.g., when a victim of a car accident is lying on a hillside along the side of a road. To address these limitations, we developed 3D telepresence technology that provides depth perception and dynamic views.

8.1.2 Sense of Presence and Task Performance via 3D Technology

Previous research shows that 3D technology generally enables an increased sense of presence. For example, Hendrix and Barfield [16] report on three studies, in which they vary display parameters and attempt to assess a user's sense of presence. The results from the first and second study indicate that the reported level of presence is significantly higher when head tracking and stereoscopic cues are provided. The third study indicates that the level of presence increases with the visual field of view.

There is also evidence to suggest that view-dependent or immersive 3D displays increase users' task performance. For example, in a study of how various system parameters affect the illusion of presence in a virtual environment, Snow [32] reports a moderately positive relationship between perceived presence and task performance. Pausch and colleagues [29] 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. Schroeder and colleagues [31] present the results of a study where distant collaborators attempted to solve a Rubik's cube type puzzle together. The authors compare face-to-face task performance with networked performance using both an immersive 3D display and a conventional 2D desktop display. They report that task performance using the networked immersive 3D display and in the face-to-face scenario were very similar, whereas desktop performance was "much poorer." Most recently, Mizell and colleagues [24] describe a careful 46-person user study aimed at determining whether or not immersive 3D virtual reality technology demonstrates a measurable advantage over more conventional 2D display methods when visualizing and interpreting complex 3D geometry. The authors report that the head-tracked 3D system shows a statistically significant advantage over a joystick-controlled 2D display.

Visual exploration of a 3D environment by means of a hand-held device, as illustrated in the right column of Fig. 8.1, has been previously proposed [14] and continues to be actively investigated [18, 19]. Applications include both viewing synthetic data and viewing computer-enhanced imagery of a user's real world surroundings. The latter is a very active research topic due to the growing interest in mobile (indoor or outdoor) Augmented Reality [30]. Today's mobile devices do not offer stereoscopic viewing (at least not a comfortable kind: the color anaglyph technique can of course be implemented on any platform), and so these types of displays can provide only (tracked) 2D perspective views into the remote environment. Yet even the earliest controlled study we are aware of [14] demonstrated good depth perception performance, comparable to typical desktop perspective displays such as the ones used in contemporary video games; a recent study [18] indicates that one can achieve a high degree of immersion and presence with "Hand-held Virtual Reality."

Thus previous research suggests that a 3DMC system may potentially improve information sharing and task performance in emergency medical situations, leading to improved patient health care. Yet there are significant technical challenges that must be overcome in order to create a 3DMC system.

8.1.3 3DMC Technical Challenges

To create a 3DMC system, we must reconstruct a dynamic remote 3D scene in real time. 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 flight 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.

8.2 Research

Three fundamental areas of technology research are required to create a 3DMC system: (1) computer vision methods for reconstruction of a 3D model/view of a dynamic scene; (2) remote consultation display paradigms; and (3) network resource management to support transmission of the 3D view to the remote consultant.

These areas and their relationships are reflected in Fig. 8.2 as follows: (a) and (b) on the left are associated with the reconstruction of a dynamic 3D model of the patient/procedure, (c) and (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.

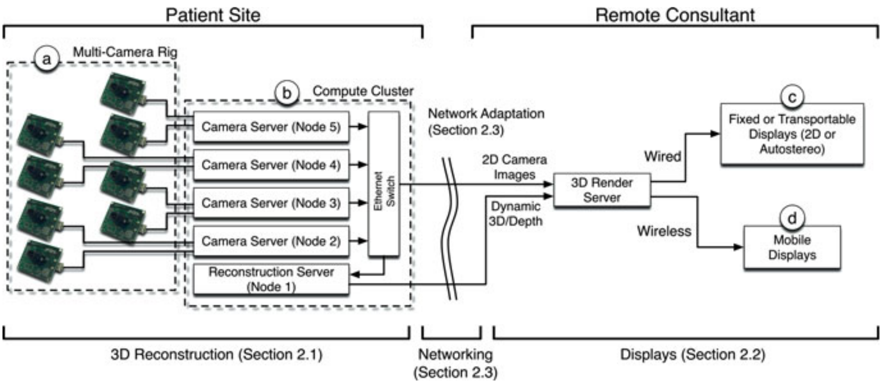

Fig. 8.2 System diagram showing the patient site components on the *left* and the remote consultant components on the *right*: (**a**) rigid multi-camera rig with eight Firewire cameras; (**b**) compute cluster with four camera nodes and one 3D reconstruction node; (**c**) a fixed or transportable viewing station with 2D or 3D (head tracked or autostereo) displays; (**d**) a mobile display such as a tracked PDA

8.2.1 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 developed a novel approach called *View-dependent Pixel Coloring* [41]. 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 [42] we use the graphics hardware to perform the 3D reconstruction very quickly as the images arrive from the Camera Servers described in Sect. 8.3. The basic idea is to use the graphics hardware to rapidly render the camera images onto a series of virtual (computer graphics) planes swept through the scene, searching in parallel for the best color matches (least variance) at a dense set of points on the planes. As shown in Fig. 8.3, for a desired new view C_n (the red dot in Fig. 8.3), we discretize the 3D space into planes parallel to the image plane of C_n. Then we step through the planes. For each plane D_i, we project the input images on these planes, and render the textured plane on the image plane of C_n to get a image (I_i) of D_i. While it is easy to conceptually think of these as two separate operations, we can combine them into a single homography (planar-to-planar) transformation for which the graphics hardware is optimized to perform. For each pixel location (u, v) in I_i, we compute the mean and variance of the projected colors. The final color of (u, v) is the color with minimum variance in $\{I_i\}$, or the color most consistent among all camera views. The index i implicitly encodes the depth plane location, so the result contains both the most consistent color and the corresponding depth value at each pixel location.

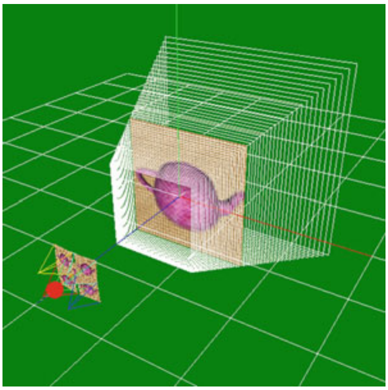

Fig. 8.3 A configuration where there are five input cameras, the small sphere in the lower left represents the new view point. Spaces are discretized into a number of parallel planes

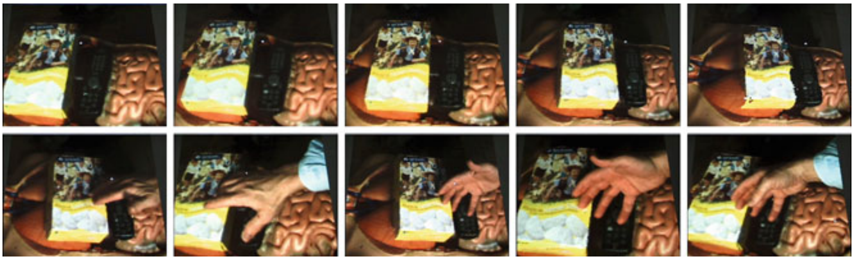

Fig. 8.4 A sequence of novel view images reconstructed using the system shown in Fig. 8.6. We set a box of Girl Scout cookies on top of the training torso to provide more obvious scene geometry

To achieve real-time performance we must make a few simplifying assumptions. First we assume that there are no occlusion problems, which allows us to compute photo-consistency values in a single pass. Second we assume that the scene is Lambertian, so that we can use color variance as the photo-consistency test. Third we use a smoothness constraint that aggregates weighted photo-consistency values from neighboring voxels, a practice commonly used in stereo. Under these assumptions, VDPC can be implemented entirely on commodity graphics hardware, taking advantage of the hardware's inherent parallelism.

When rendering stereoscopic views, we typically use VDPC to render one reference frame and then warp this frame to the second view based on the calculated depth map. In the places of occlusions we use neighboring pixels with large depth values (e.g., pixels further away from the center of projection) to fill in the gaps. Given the small baseline between the stereoscopic views, the artifacts caused by this in painting process are usually negligible. The final image quality depends on the displacement between the desired view and the input camera views. The larger baseline, the more artifacts are likely to become visible.

Figure 8.4 shows some results from our current prototype (Fig. 8.6). 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.

8.2.2 Remote Consultation 3D 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. See, for example, Fig. 8.1a,c. Our current prototype addresses this scenario with 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 (Fig. 8.1, right). For a remote display we looked 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 was to investigate the development or adaptation of tracking technology and user interface paradigms that would allow a remote medical advisor to use a PDA as a "magic lens" [12, 13, 25, 28], providing a view of the remote patient, with natural interactive viewpoint control to help address occlusions and to provide some sense of depth.

We investigated a two-handed patient "prop" paradigm as shown in Fig. 8.5. Hinckley et al. introduced the idea, using a doll's head or rubber ball and various tools as 'props' for neurosurgeons visualizing patient data [17]. 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

Fig. 8.5 *Left*: Our first tracked PDA prototype used a HiBall-3000^{TM} tracking system [1], with sensors mounted on the PDA (Toshiba e800, *left hand*) and the surrogate (*right hand*). *Right*: Our current prototype uses a PointGrey DragonFly camera [3] 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 [2] to track the surrogate with respect to the PDA (camera)

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 [40].

We have developed 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 [2] to track the surrogate (PDA cover) with respect to the PDA. The PDA Server, which currently runs on the viewing station server (Sect. 8.3), continually gets a complete representation of the reconstructed data from the compute/rendering cluster (b) via a dedicated Ethernet connection as described in Sect. 8.3. It also obtains the estimated position and orientation of the PDA from the Tracking Server using the Virtual-Reality Peripheral Network (VRPN) protocol [4]. 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.

8.2.3 Networking

In our target 3DMC scenarios the network path represents a significant bottleneck. We must carefully manage this resource in order to ensure that at all times we transmit the data that are most useful to the overall application and the goals of the user. In particular, 3DMC has the potential to generate many media streams with complex interstream semantic relationships, and the utility of the information from one data source may depend on the quality and utility of information from some other data source. For example, given two video cameras that share a significant overlap of field of view, it may be preferable to allocate available bandwidth to capture and transmit a high-quality image for only one of the two streams while allowing the quality of the other stream to degrade. Alternatively, it may be better to allocate bandwidth equally in order to achieve similar quality for both streams – useful for VDPC or stereo correlation and high-quality 3D reconstruction. Thus the challenge we face is twofold. First, how can we compactly and intuitively specify an adaptation policy to support specific user-level goals? Second, how can we efficiently evaluate that policy?

We need a framework for addressing the problems of adaptation that is more flexible than previous approaches, which often rely on statically defined priorities (e.g., prioritize audio over video) or simple rule-based decisions (e.g., when available bandwidth is X, do Y). In the framework we are developing, all possible tradeoffs available to the application are mapped as nodes in an N-dimensional “utility space.” Each dimension represents a particular axis for adaptation. Edges between nodes represent both encoding dependencies as well as encoding costs. The nodes and edges form a graph embedded within the utility space. The current information needs of the system are modeled as a “point of interest” within this space. The location of this point of interest changes to reflect the how the user is interacting with the system and the dynamics of the application. The utility of any given tradeoff is inversely proportional to the distance between the node that represents the tradeoff and the point of interest. Adaptation is now simply the process by which we select the most useful tradeoff available as defined by the ratio of utility to cost. Real-time evaluation is feasible since the adaptation is now a simple mechanical process of maintaining the set of possible tradeoffs in the graph and their distance to the point of interest. Additional technical details can be found in [15].

While the use of a utility space provides us with a mechanical means of driving adaptation and allows parsimonious specification of adaptation policy, the construction of a utility space for a specific application is more art than science. An application developer must incorporate appropriate domain knowledge in order to make choices about which adaptation dimensions are going to be modeled, how these dimensions are scaled relative to each other, the specific distance function that will be used to establish utility, and how the actions of the user are reflected by the point of interest. For 3DMC, we have identified five dimensions for adaptation: one each for time, video resolution, and relative change of visual content; and two that capture the notion of region of interest (i.e., field of view). Preliminary experiments show the system is able to make complex, non-trivial adaptation decisions in an emulated eight-camera setup such as in Fig. 8.6 [21]. Much of the remaining challenge is to develop and evaluate specific utility functions that correspond to the actual perceived quality of real users.

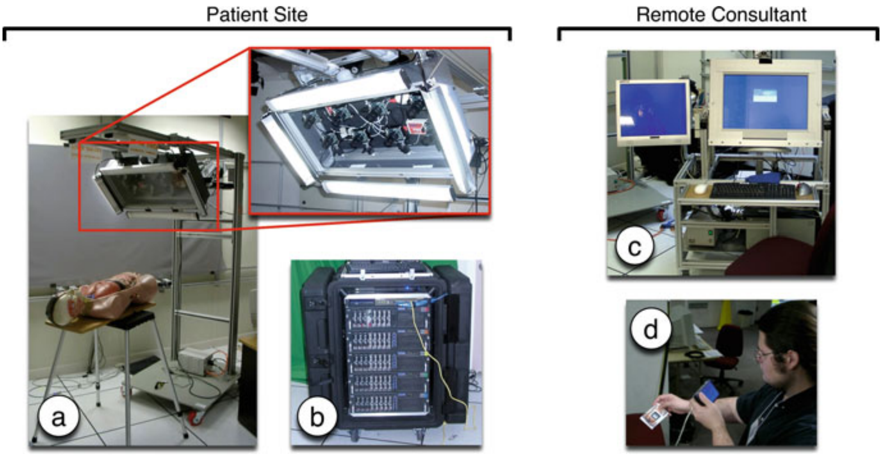

Fig. 8.6 Our proof-of-concept 3DMC prototype, with patient site components on the *left* and remote consultant components on the *right*: (**a**) portable camera unit with eight Firewire cameras and high-frequency area lights; (**b**) compute cluster; (**c**) a transportable consultant viewing station with 2D and 3D (head tracked or autostereo) displays; (**d**) a tracked PDA mobile display

8.3 Our 3DMC Prototype System

As shown in Fig. 8.6, 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) and (d).

The portable camera unit (PCU) shown in (a) of Fig. 8.6 is a rolling unit holding a camera-lighting array with eight 640×480 resolution digital (IEEE 1394a) color cameras from Point Grey Research [3]. 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 Fig. 8.6 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 5th server. Each camera server can optionally record the video streams to disk. The 5th 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 Sect. 8.2.1. 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 we have placed the PCU 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 Fig. 8.6 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 Fig. 8.2. 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).[1] The unit also includes an AC isolation transformer.

[1] Autostereoscopic displays provide one 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.

The prototype of the tracked PDA mobile display (d) in Fig. 8.6 uses a DragonFly camera [3] mounted on a Toshiba e800 PDA. The camera is attached to the rendering PC (above) via a Firewire cable, which uses ArToolKit [2] to compute the relative position and orientation of the PDA. (This is discussed further in Sect. 8.2.2.) 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 [38, 39], 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.

Our 3DMC proof-of-concept system currently exhibits numerous shortcomings and limitations. Its acquisition system of eight cameras provides insufficient density for high quality reconstruction and insufficient range of views to enable reconstruction over a large viewing volume. The reconstruction algorithm (Sect. 8.2.1) makes simplifying assumptions in favor of speed, so the resulting 3D reconstructions often exhibit small gaps or other artifacts. The image quality of the reconstructions is often noticeably poorer than the quality of any of the eight input images.

The displays also exhibit several limitations. The commercial autostereo display, as virtually all such displays, presents fusible stereo imagery only within a range of distances from the display surface. Often a viewer instinctively moves closer to the display surface in an effort to more closely observe the medical scene, but is disappointed in losing the stereo imagery rather than gaining more detail. The PDA display exhibits even more severe limitations; it currently supports neither stereo nor head-tracking, although parallax is afforded through two-handed motions.

8.4 Evaluation

Because an extremely large amount of resources are needed to develop and deploy 3DMC we investigated its potential effectiveness already at this stage in its research and development cycle. If 3DMC has no possibility of having a positive impact on emergency medical care then there is no need to continue developing it. To evaluate its potential effectiveness we conducted a controlled experiment that compared the outcomes of a paramedic diagnosing and treating a trauma victim under one of the following three conditions: working alone, working in consultation with a physician using today's state-of-the-art video-conferencing, and working in consultation with a physician via a 3D proxy.

For the 3D proxy condition, the consulting physician was physically present in the same room as the mannequin and paramedic. The 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 3DMC. Details regarding this approach and its validity can be found in [35].

Fig. 8.7 *Human patient simulator* (HPS) staged as an accident victim for the experimental evaluation. Note the 2D video cameras and display mounted on the overhanging structure. Inset shows the HPS airway details

The experiment was a posttest, between-subjects design in which an emergency medical situation was realistically simulated. The medical situation involved the diagnosis and treatment of a trauma victim suffering from a blocked airway, when a victim's airway collapses. Each paramedic eventually had to perform a procedure beyond his/her level of everyday expertise – an emergency cricothyrotomy, i.e., make a surgical incision to open up the victim's windpipe, in order to save the victim's life. The victim was actually a METI human patient simulator, a sophisticated computerized mannequin that responds realistically to medicine and other treatments (Fig. 8.7). In all, 60 paramedics participated in the experiment; 20 per each condition.

Our evaluation measures included task performance and paramedics' perceptions of self-efficacy. A task performance protocol was developed in collaboration with emergency care physicians. It took into account task and subtask order, quality of task execution, harmful interventions performed, and task completion time. Each experiment session was recorded using four different video cameras (Fig. 8.7) that captured the monitor showing the victim's vital signs and multiple views of the paramedic, mannequin and physician (when present). These video-recordings were subsequently analyzed to grade task performance.

The theory of self-efficacy refers to a person's judgment of their capability to perform a certain task [5]. It fundamentally proposes that people are products as well as producers of their own environment and social systems. It suggests that people have self-beliefs that let them exercise control over their thoughts, beliefs and feelings which, in turn, effects how they behave. Self-efficacy affects a person's behavior such that people with high self-efficacy try harder when facing a difficult task and are less sensitive to failures and performance obstacles than people with low self-efficacy. Low levels of self-efficacy lead to more stress, slower recovery after failures, and higher likeliness to give up when facing problems. The primary, most influential source of self-efficacy is previous task performance experiences.

Self-efficacy is considered a major determinant of future task performance, i.e., task performance creates beliefs in ability which influence future task performance. There have been many studies investigating the validity of self-efficacy as a predictive measure. For example, Downey [11] compares and validates computer self-efficacy instruments with respect to predictive power on future performance, attitude and usage.

Because experiences, especially successful ones, are the strongest source of influence on self-efficacy [5] a paramedic's perceptions of self-efficacy after diagnosing and managing a difficult airway in a simulated medical scenario can help predict the paramedic's future performance in diagnosing and managing a difficult airway. That is, the theory of self-efficacy suggests that paramedics with higher levels of self-efficacy after diagnosing and managing a difficult airway will actually perform the same tasks better in the future.

After participating in a session each paramedic completed a self-efficacy post questionnaire. They also participated in a post interview, discussing their perceptions of the session.

Details of the data analysis can be found in [34]. The results illustrate that paramedics working in consultation with a physician via the 3D proxy tended to provide better medical care to trauma victims than paramedics working in consultation via today's 2D videoconferencing or paramedics working alone. Fewer subtask errors and harmful interventions were performed in the 3D proxy condition. Three paramedics working alone did not perform a cricothyrotomy, although a cricothyrotomy was required to save the victim. A total of eleven harmful interventions were performed when paramedics worked alone, and six were performed when paramedics consulted with a physician via 2D videoconferencing. In comparison only two harmful interventions were performed when the consultation occurred via the 3D proxy. However no statistically significant differences with respect to task performance times across conditions emerged from the data analysis.

Paramedics consulting with a physician via the 3D proxy reported the highest levels of self-efficacy. In comparison paramedics collaborating via 2D videoconferencing reported the lowest levels of self-efficacy. Furthermore the less work experience paramedics in the alone and 2D conditions had, the lower they rated their ability to treat similar patients in the future, whereas work experience had no impact at all on feelings of self-efficacy for paramedics in the 3D proxy condition. All of these results were statistically significant at the 0.05 level. These results suggest that the 3DMC may have a positive impact on future task performance, irrespective of a paramedic's years of professional experience.

However paramedics also mentioned several that 3DMC has the potential to make paramedics' work visible and subsequently evaluated in new ways that may have a negative impact on their careers. A paramedic explained:

> *It kind of makes somebody nervous being monitored by a physician, someone of such higher training. And you're 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.*

Ways to avoid such negative consequences that were mentioned by paramedics 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 between paramedics in the field and physicians and nurses in the hospital.

Additional discussion regarding potential benefits and challenges of 3DMC from the perspective of physicians, nurses, medical center IT professionals, and large and small medical center administrators can be found in [33].

8.5 Cost Analysis

We generally adhere to the philosophy that to make something *cost*-effective one must first make it *effective*, and we are not yet prepared to make a strong argument that the systems we envision will be effective much less cost effective. Real-time on-line 3D telepresence is aimed at long-term needs for more efficient (timely and effective) medical consultation, improved health care, and (eventually) reduced costs. However the work is very early (beyond the current state of the art) which makes it difficult to assess the likely costs.

Several studies have attempted to quantify the cost effectiveness of conventional 2D televideo systems used for remote medical consultations, and we believe the end costs for 3D systems would be comparable. For example, [9, 10, 20, 37] and [43] concluded that the 2D systems they evaluated were (or would be) cost effective. In particular, Agha et al. concluded in [43] that "Telemedicine is a cost-effective alternative for the delivery of outpatient pulmonary care for rural populations with limited access to subspecialty services." They claim that the cost effectiveness is related to three major factors: (a) cost sharing; (b) effectiveness of telemedicine in terms of patient utility and successful clinical consultations; and (c) indirect cost savings accrued by decreasing cost of patients' lost productivity. We would expect the systems we propose to also benefit from these factors.

And while it would be imprudent at this point to make specific cost projections for 3D systems, it is possible that a "transportable" 3D telepresence node such as we describe could, in the future, be developed for roughly 2–5 times the fixed (one time) costs of a conventional 2D telemedical consultation system. This increased cost would be due almost exclusively to an increase in the number of cameras and computers, which (incidentally) are continually decreasing in price. One might add cost for networking infrastructure also, however it is possible (perhaps reasonable to assume) that future networking infrastructure between healthcare facilities will improve independently from this proposed work. For example, [27] points out how the Telecommunications Act of 1996 has and continues to improve access and reduce costs for urban and (explicitly in the act) rural health care providers. In any case, the networking techniques we developed (Sect. 8.2.3) are aimed at operation over a variety of capabilities, and as such could (if proven effective) perhaps support some functionality with the existing infrastructure.

8.6 Discussion

In the 2001 PITAC report to the President of the United States, Transforming Health Care Through Information Technology, recommendation five states that the US Department of Health and Human Services should establish an "aggressive research program in computer science" that addresses 'long-term needs, rather than the application of existing information technology to biomedical problems' [27, p. 13]. 3DMC is an aggressive research project in computer science addressing long-term needs for more effective medical consultation, improved health care, and (eventually) reduced medical care costs as described above. The work is in an early stage but we have made progress developing computer vision methods for reconstruction of a 3D model/view of a dynamic scene, remote consultation displays, and network resource management algorithms to support transmission of 3D views. The results were realized in a prototype system.

The Office for the Advancement of Telehealth [27] points out how the Telecommunications Act of 1996 has and continues to improve access and reduce costs for urban and rural health care providers. Our project builds on this tradition, discovering new ways technology can be used to provide emergency healthcare outside hospital settings to trauma victims. Trauma is a significant health problem, frequently referred to as the "hidden epidemic of modern society" because it is responsible for more productive years lost than heart disease, cancer and stroke combined [7,22]. 3DMC can potentially bring needed healthcare expertise to trauma victims before they are transported to hospitals. The sooner victims receive appropriate expert medical care, the shorter their recovery time and lower their medical care costs. Thus 3DMC could have a significant impact on patient healthcare in the future.

Acknowledgements At UNC-Chapel Hill we thank Jim Mahaney and John Thomas for their technical support; and former Graduate Research Assistant Max Smolens for his early contributions to the effort. We thank the paramedics who participated in the evaluation. This effort is primarily supported by National Library of Medicine contract N01-LM-3-3514: "3D Telepresence for Medical Consultation: Extending Medical Expertise Throughout, Between and Beyond Hospitals," and in part by NSF grant EIA-0303590: "Ubiquitous Pixels: Transforming Collaboration & Teaching with Pervasive Wall Displays."

References

1. 3rdTech, Inc. (2005). http://www.3rdtech.com/
2. ARToolkit (2005). http://www.hitl.washington.edu/artoolkit/
3. Point Grey Research (2005). http://www.ptgrey.com/
4. Virtual Reality Peripheral Network (2006). http://www.cs.unc.edu/Research/vrpn/
5. Bandura, A.: Self-efficacy: the exercise of control. W.H. Freeman, New York (1997)
6. Chu, Y., Huang, X., Ganz, A.: Wista: A wireless transmission system for disaster patient care. In: Proceedings of BROADNETS: 2nd IEEE/CreateNet International Conference on Broadband Networks, pp. 118–122. Omnipress, Boston (2005)
7. Coates, T., Goode, A.: Towards improving prehospital trauma care. Lancet **357**(9274), 2070 (2001)
8. David G. Ellis, M., James Mayrose, P., Dietrich V. Jehle, M., Ronald M. Moscati, M., Guillermo J. Pierluisi, M.: A telemedicine model for emergency care in a short-term correctional facility. Telemedicine Journal and e-Health **7**(2), 87–92 (2001)
9. Dimmick, S.L., Mustaleski, C., Burgiss, S.G., Welsh, T.S.: A case study of benefits and potential savings in rural home telemedicine. Home Healthcare Nurse **18**(2), 125–135 (2000)
10. Doty, E., L. H. Zincone, J., Balch, D.C.: Medicine Meets Virtual Reality: Healthcare in the Information Age, chap. Telemedicine in the North Carolina Prison System, pp. 239–241. IOS Press, Eds. J. Weghorst, Hans B. Sieburg, and Karen S. Morgan (1996)
11. Downey, J.: Measuring general computer self-efficacy: The surprising comparison of three instruments in predicting performance, attitudes, and usage. Hawaii International Conference on System Sciences **8**, 210a (2006). DOI http://doi.ieeecomputersociety.org/10.1109/HICSS.2006.268
12. Fitzmaurice, G.W.: Situated information spaces ad spatially aware palmtop computers. Communications of the ACM **36**(7) (1993)
13. Fitzmaurice, G.W., Buxton, W.: The chameleon: Spatially aware palmtop computers. In: ACM CHI94. Boston, MA USA (1994)
14. Fitzmaurice, G.W., Zhai, S., Chignell, M.H.: Virtual reality for palmtop computers. ACM Transactions on Information Systems **11**, 197–218 (1993)
15. Gotz, D., Mayer-Patel, K.: A general framework for multidimensional adaptation. In: MULTIMEDIA '04: Proc. of the 12th annual ACM international conference on Multimedia, pp. 612–619. ACM, New York (2004). DOI http://doi.acm.org/10.1145/1027527.1027671
16. Hendrix, C., Barfield, W.: Presence within virtual environments as a function of visual display parameters. Presence: Teloperators and virtual environments **5**(3), 274–289 (1996)
17. Hinckley, K., Pausch, R., Goble, J.C., Kassell, N.F.: Passive real-world interface props for neurosurgical visualization. In: CHI '94: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 452–458. ACM, New York (1994). DOI http://doi.acm.org/10.1145/191666.191821
18. Hwang, J., Jung, J., Kim, G.J.: Hand-held virtual reality: A feasibility study. In: Proceedings of the ACM symposium on Virtual reality software and technology, pp. 356–363. ACM (2006)
19. Hwang, J., Jung, J., Yim, S., Cheon, J., Lee, S., Choi, S., Kim, G.J.: Requirements, implementation and applications of hand-held virtual reality. IJVR **5**(2), 59–66 (2006)
20. Kaiser Permanente: Outcomes of the kaiser permanente tele-home health research project. Archives of Family Medicine **9** (2000)
21. Krishnan, S., Mayer-Patel, K.: A utility-driven framework for loss and encoding aware video adaptation. In: MULTIMEDIA '07: Proceedings of the 15th international conference on Multimedia, pp. 1026–1035. ACM, New York, NY, USA (2007). DOI http://doi.acm.org/10.1145/1291233.1291459
22. Meyer, A.: Death and disability from injury: A global challenge. J Trauma **44**(1), 1–12 (1998)
23. Michael G. Kienzle, M.: Rural-academic integration: Iowas national laboratory for the study of rural telemedicine. Tech. rep., National Laboratory for the Study of Rural Telemedicine (2000)
24. Mizell, D.W., Jones, S.P., Slater, M., Spanlang, B., Swapp, D.: Immersive virtual reality vs. "flat-screen" visualization: A measurable advantage. (Submitted for publication). (2003)

25. Mohring, M., Lessig, C., Bimber, O.: 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), pp. 252–253. IEEE Computer Society, Washington, DC, USA (2004). DOI http://dx.doi.org/10.1109/ISMAR.2004.63
26. Mun, S.K.: Project phoenix: Scrutinizing a telemedicine testbed. final project report, national library of medicine, contract #n01-lm-6-3544. Tech. rep., Georgetown University Medical Center (2000)
27. Office for the Advancement of Telehealth: Telemedicine report to congress. Tech. rep., U.S. Department of Health and Human Services, Health Resources and Services Administration (2001)
28. Pasman, W., van der Schaaf, A., Lagendijk R.L., Jansen F.W.: Accurate overlaying for mobile augmented reality. Computers and Graphics **23**, 875–881 (1999)
29. Pausch, R., Shackelford, M.A., Proffitt, D.: A user study comparing head-mounted and stationary displays. In: Proc. of IEEE Symposium on Research Frontiers in Virtual Reality, pp. 41–45. IEEE (1993)
30. Reitmayr, G., Drummond, T.W.: Going out: Robust modelbased tracking for outdoor augmented reality. In: Proceedings of 5th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR, pp. 109–118 (2006)
31. Schroeder, R., Steed, A., Axelsson, A.S., Heldal, I., Abelin, A., Wideström, J., Nilsson, A., Slater, M.: Collaborating in networked immersive spaces: As good as being there together? Computers & Graphics, Special Issue on Mixed Realities – Beyond Conventions **25**(5), 781–788 (2001). http://www.elsevier.com/gej-ng/10/13/20/57/34/32/article.pdf
32. Snow, M.P.: Charting presence in virtual environments and its effects on performance. Ph.D. thesis, Virginia Polytechnic Institute and State University (1996)
33. Söderholm, H.M., Sonnenwald, D.H.: Visioning future emergency healthcare collaboration: Perspectives from large and small medical centers. Journal of the American Society for Information Science and Technology **61**(9), 1808–1823 (2010).
34. Söderholm, H.M., Sonnenwald, D.H., Manning, J.E., Cairns, B., Welch, G., Fuchs, H.: Exploring the potential of video technologies for collaboration in emergency medical care. part ii: Task performance. Journal of the American Society for Information Science and Technology **59**(14), 2335–2349 (2008). DOI 10.1002/asi.20939. http://www3.interscience.wiley.com/journal/121378034/abstract
35. Sonnenwald, D.H., Söderholm, H.M., Manning, J.E., Cairns, B., Welch, G., Fuchs, H.: Exploring the potential of video technologies for collaboration in emergency medical care. part i: Information sharing. Journal of the American Society for Information Science and Technology **59**(14), 2320–2334 (2008). DOI 10.1002/asi.20934. http://www3.interscience.wiley.com/journal/121378032/abstract?CRETRY=1&SRETRY=0
36. Tachakra, S.: Depth perception in telemedical consultations. Telemed J E Health **7**(2), 77–85. (2001)
37. Telemedicine Directorate: Advanced concept technology demonstrations letter of finding. Tech. rep., Walter Reed Army Medical Center (2000). http://www.salud.gob.mx/unidades/sic/cits/doctosFuente/2001REPO.pdf
38. Wagner, D., Schmalstieg, D.: First steps towards handheld augmented reality. In: ISWC '03: Proceedings of the 7th IEEE International Symposium on Wearable Computers International Symposium on Wearable Computers. IEEE Computer Society, Washington, DC, USA (2003)
39. Wagner, D., Schmalstieg, D.: Handheld augmented reality displays. In: G. Welch (ed.) Proceedings of 2nd Emerging Display Technologies Workshop (EDT 2006), pp. 35–36 (2006)
40. Welch, G., Noland, M., Bishop, G.: Complementary tracking and two-handed interaction for remote 3d medical consultation with a pda. In: G. Zachmann (ed.) Proceedings of Trends and Issues in Tracking for Virtual Environments, Workshop at the IEEE Virtual Reality 2007 Conference. Shaker, Charlotte, NC USA (2007)
41. Yang, R.: View-dependent pixel coloring—a physically-based approach for 2d view synthesis. Ph.d. thesis, University of North Carolina at Chapel Hill (2003)

42. Yang, R., Pollefeys, M., Yang, H., Welch, G.: A unified approach to real-time, multi-resolution, multi-baseline 2d view synthesis and 3d depth estimation using commodity graphics hardware. International Journal of Image and Graphics (IJIG) **4**(4), 1–25 (2004)
43. Zia Agha MD, M., Ralph M. Schapira, M., Azmaira H. Maker, P.: Cost effectiveness of telemedicine for the delivery of outpatient pulmonary care to a rural population. Telemedicine Journal and e-Health **8**(3), 281–291 (2002)