3D Telepresence *for* Off-Line Surgical Training *and* On-Line Remote Consultation

Greg Welch[*]
Ruigang Yang[†] Bruce Cairns, M.D.[‡] Herman Towles[*]
Andrei State[*] Adrian Ilie[*] Sascha Becker[§]
Dan Russo[§] Jesse Funaro[§] Diane Sonnenwald[¶]
Ketan Mayer-Patel[*] B.Danette Allen[*] Hua Yang[*]
Eugene Freid, M.D.[‖] Andy van Dam[§] and
Henry Fuchs[*]

May 11, 2005

Abstract

We present an overview of multi-year efforts aimed at developing technology to support natural telepresence over *time* for off-line surgical training, and over *space* for on-line or "live" remote consultation. Our goal is to capture and display high-fidelity, 3D graphical reconstructions of the actual time-varying events, allowing observers to change their (stereo) viewpoints naturally, by walking around, leaning forward/backward, rotating their head, etc. We describe the processes we use in our early prototypes, and present some experimental results.

---

[*] University of North Carolina at Chapel Hill, Dept. of Computer Science
[†] University of Kentucky, Dept. of Computer Science
[‡] University of North Carolina at Chapel Hill, Dept. of Surgery
[§] Brown University, Dept. of Computer Science
[¶] Göteborg University and the University College of Borås
[‖] University of North Carolina at Chapel Hill, Dept. of Anesthesiology

1 Background

In 1994 Fuchs et al. outlined a vision and methods aimed at using a "sea of cameras" to achieve 3D telepresence for remote medical procedures [10]. We have since been pursuing related work formally and informally with multiple collaborators over the years. The institutions include Advanced Network and Services (Jaron Lanier and Amela Sadagic), Brown University, Göteborg University and the University College of Borås (Diane Sonnenwald), the University of Kentucky, the University of North Carolina at Chapel Hill, and the University of Pennsylvania (Kostas Daniilidis and Ruzena Bajcsy—now at the University of California at Berkeley).

Our current telepresence research is aimed primarily at two paradigms: off-line surgical training and on-line remote medical consultation. In 2002 we presented an intermediate summary of results from a three-year effort funded by a U.S. National Science Foundation (NSF) grant titled "Electronic Books for the Tele-Immersion Age: A Paradigm for Teaching Surgical Procedures" [38]. A more recent and detailed systems paper will appear in IEEE Multimedia [40]. In Section 3 we present a somewhat condensed version of the work described in [40], introducing some tools and methods for creating and viewing an *immersive electronic book* or IEBOOK, and presenting results that include mock medical procedures with anatomical models and knot-tying aimed at teaching surgical suturing.

Since 2002 we have also been working on a three-year effort funded by a U.S. National Library of Medicine (NLM) contract titled "3D Telepresence for Medical Consultation: Extending Medical Expertise Throughout, Between and Beyond Hospitals." Our primary aim is to enhance and expand medical diagnosis and treatment in a variety of life-critical trauma scenarios, by providing an *advising* health care provider and a distant medical *advisee*, both benefitting from a high-fidelity visual and aural sense of 3D presence with each other. Primarily we envision scenarios involving extemporaneous consultation related to unforeseen events where time is critical, anxiety is high, and a physician or technician would welcome a remote expert consultation for concurrence or guidance in diagnosis and management. In Section 4 we present a current look at this effort, describing the initial research and the future directions.

2 Introduction

Conventional two-dimensional (2D) video recordings have long been available to help surgeons learn new procedures at their leisure, and 2D televideo has proved to be useful for live face-to-face medical consultations and visual examinations of wounds. However, anecdotal accounts and published case studies raise two common visual problems: the absence of natural viewpoint control and depth perception.

For example, camera view difficulties were mentioned in multiple places in the final report for the NLM's National Laboratory for the Study of Rural Telemedicine [19]. 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 dimensions hence certain aspects of the exam could be enhanced by more than one camera angle."

The problem was also identified in [6] 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, [31] points out that "Impaired *depth perception* is a significant problem in telemedicine." [emphasis added] and notes that "The most important cue of depth is due to binocular disparity." The author describes several "coping strategies" that can be used to overcome the inherent limitation of the 2D imagery. Chief among them 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. The author surmises that 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."

We believe there are two common training and consultation 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. 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 indirectly or directly aid in the consultation, users must resort to secondary visual cues or verbal clarification from the remote collaborator, both of which impose additional cognitive loads compared to the natural viewing afforded if the consulting physician was actually next to the patient.

Previous research in both computer supported cooperative work [8,22] and language theory [3] suggests that 2D video-conferencing lacks the richness of face-to-face interaction where there are multiple and redundant communication channels, implicit cues, and spatial co-references. 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., [23, 24, 29]) 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 physician's view of the patient, and as a result some patients felt that the physician could not always "see" how the patient was "really doing."

To address these problems we are working on using an array of cameras and computers to capture in 2D, reconstruct in 3D, store or transmit, and display real surgical procedures in a way that allows an observer to witness the procedure in a realistic, immersive, 3D environment. Our goal is to develop technology to support 3D telepresence over *time* for off-line surgical training, and over *space* for on-line or “live” remote consultation. Our approach is to capture and display high-fidelity, 3D graphical reconstructions of the actual time-varying events, allowing observers to change their (stereo) viewpoints naturally, by walking around, stepping or leaning forward/backward, rotating their head, etc.

In the future, such technology will allow surgical trainees to use an *immersive electronic book* or IEBOOK to witness and explore a past surgical procedure as if they were there, with the added benefit of control over time, instruction from the original surgeon or another instructor, as well as integrated 3D illustrations, annotations and relevant medical metadata. In addition, we believe that on-line real-time versions of similar technology will improve live communication and trust between geographically-separated medical personnel, enabling new extensions of medical expertise throughout, between, and beyond medical facilities.

3 Off-Line Surgical Training

The traditional mode of teaching surgery follows the adage “See one, do one, teach one,” but the complexity and rarity of some procedures limits a surgeon-in-training’s ability to witness these procedures (to “see one”). One example is trauma-related procedures, where surgeons must move quickly to stabilize their patient, and teaching must be secondary to patient care. Such procedures are unplanned, and certain critical traumas are rare.

This and other issues regarding surgical training are not trivial. Most surgical specialties require at least five years of post-medical school training before one is independent and eligible to become board certified, and frequently it takes many years after board certification before most surgeons are entirely comfortable with their craft. In addition, like many other technology driven fields, it is difficult for practicing surgeons to keep up to date with a seemingly dizzying array of new devices and procedures designed to improve care. Combined with an overall increased focus on limiting medical error, there is an intense interest in improving and maintaining surgical competency.

There is currently no educational environment that can replace the apprenticeship environment of the operating room. There are teaching modules that can be used to teach and assess various technical skills, but there is nothing similar to the flight simulator for pilot training that can replace being in the operating room in order to learn how to perform a specific operative procedure.

The most obvious substitute is two-dimensional (2D) video. Such video has long been available to help surgeons learn new procedures, but these videos are universally considered marginally effective at best for a number of reasons. Video relegates the viewer to the role of a passive observer, and as mentioned in Section 2 they cannot interactively change their view and have relatively few depth cues. Subtle, complex motions that can be critical in an operation are easy to miss. Finally, conventional 2D video offers only linear control over the timing of the playback. Replaying the same video only provides the same experience a second time, rather than permitting a new or changed perspective.

Three-dimensional (3D) computer graphics systems can offer people an opportunity to see real or virtual models with the depth, occlusion relationships, and motion parallax that they are used to in every-day life. When coupled with interaction devices and techniques, such systems can occasion personal experiences in human minds. The power of this paradigm has been leveraged for training for many years. Arguably the most successful example is flight simulation for pilot training [27]. Today flight simulators are considered so effective (and cost-effective) that it is not unusual for a pilot to use one to train on a new version of a plane, and then actually fly the real plane for the first time on a *regularly scheduled flight with regular passengers*.

3D graphics techniques have also been used for medical training and education [2, 16, 42]. Previous efforts aimed at training have primarily combined realistic models with interaction devices (sometimes haptic) to simulate the experience of actually performing a particular procedure. Rather than aiming to simulate the "do one" of "See one, do one, teach one," with this work we are aiming at improving the opportunities to "see one," and improving the effectiveness of the associated learning experience.

Our hope is that IEBOOK technology will some day allow surgeons to integrate the technical aspects of the procedure with the anatomy, radiography, and physiology of the patient's condition, thus best simulating and in many ways augmenting the real time educational experience of learning how to perform an operative procedure. See Figure 1 for an early artist's sketch and a screen shot from our current prototype.

We are still in the relatively early stages of development, and have not yet realized our grand vision for a turn-key system with operating room-sized 3D reconstructions, extensive authoring tools, and a widely-available means for immersive or web-based exploration. However we have developed new techniques and tools that together allow us to create a preliminary IEBOOK that embodies some of the basic concepts.

3.1 Overview

The overall process we use to create an IEBOOK includes two major phases: *content creation* and *content viewing.* These phases are depicted left-to-right in Figure 2. Content creation involves the three major steps shown in magenta: *acquisition* of 2D video of an actual surgical event from multiple cameras, *reconstruction* of 3D models, and *authoring* of an IEBOOK. Presently we support two paradigms for content viewing. The primary paradigm is an immersive one (shown in Figure 2), where a head-tracked stereo system allows the user to see and interact with the time-

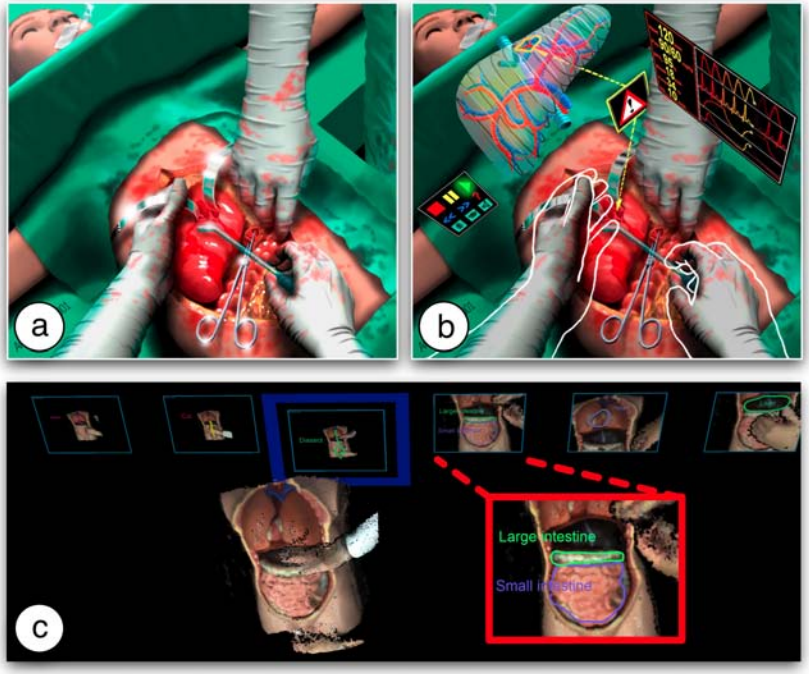

Figure 1: Top: early conceptual sketch of an iebook for training surgeons for treatment of blunt liver trauma: (a) the original surgery, (b) the acquired and replayed/immersively visualized 3D scene with annotations and trainee's hands. (Sketch by Andrei State.) Bottom (c): an iebook in our present immersive display environment. The third of six annotated snapshots is selected and playing. We added a red "zoom box" to magnify the snapshot for the fourth item. See also Figure 6.

varying 3D reconstructions of the event, highlights, and annotations. We also support a web-based paradigm that provides fixed-view anaglyphic (red-blue) stereo movies and images of dynamic 3D reconstructions, and a VCR-like non-stereo interface to dynamic VRML models. Presently our web-based content does not include annotations.

Figure 2: The overall process we use to create an IEBOOK.

3.2 2D Acquisition

The creation of an IEBOOK begins with the dynamic 2D capture (acquisition) of the event of interest. The primary piece of equipment we use for capture is something we call the *camera cube.* The unit, shown in Figure 3, is approximately one meter on each side, and is constructed from rigid modular aluminum framing manufactured by 80/20 Inc.[1] It includes four fluorescent high frequency linear lights and eight 640 × 480 Dragonfly FireWire color cameras from Point Grey Research (PGR).

For *geometric* camera calibration we use Bouguet's Camera Calibration Toolbox for Matlab™ [1]. The procedure involves taking a sequence of *frame sets*[2] of a moving black and white checked calibration pattern, and then using the toolbox to estimate the *intrinsic*[3] and *extrinsic*[4] parameters, and lens distortion coefficients for each camera.

For *photometric* camera calibration, we previously developed a method for matching color models after the acquisition [12], but

---

[1] See `http://www.8020.net`.

[2] By *frame set* we mean a set of eight synchronized images—one from each of the eight cameras on the camera cube.

[3] Focal length, image center, and pixel skew.

[4] Rotation and translation.

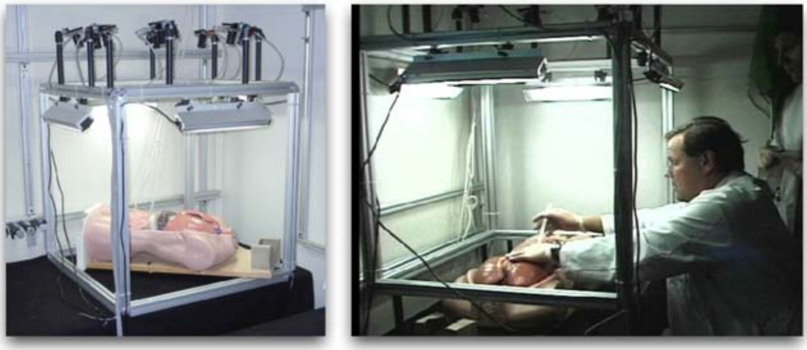

Figure 3: Camera cube used for acquisition. Left: with a medical training model in place. Right: Dr. Bruce Cairns performing a mock procedure to manage severe blunt liver trauma.

more recently have developed a novel closed-loop approach that seeks to maximize dynamic range by adjusting gain and bias registers in the PGR cameras, while estimating higher-order parameters for a set of multi-dimensional polynomials that implement more general linearization and color warping. This combination of hardware and software correction implements what are in some sense hybrid polynomials, where the lower order (linear) terms are effected in hardware and the higher order terms in software. Because the terms cannot be made completely independent, the software alternates between adjusting the hardware registers and computing the polynomial coefficients, while observing the effects in images of a Gretag Macbeth^{TM} color chart.

The actual capture of a dynamic event involves starting image capture server software on each of four servers, and then starting a master synchronization program which then starts the synchronized capture to disk. The capture continues for a preset number of frame sets.

3.3 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 we call View-dependent Pixel Coloring (VDPC). 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*. VDPC simultaneously estimates a view-dependent 3D model (height field) of the first visible surfaces of the underlying 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 surgery (skin, organs, and bodily fluids). We describe the method here at a level that we hope will convey the basic concepts. We refer the reader to [40] and [41] for more theory, implementation details, and a complete discussion of the related work.

The fundamental concepts are illustrated in Figure 4. We begin by defining a 3D *perspective voxel grid* from the *desired viewpoint*. We typically choose the desired viewpoint to be situated above and looking down into the camera cube as shown on the left. Then, as illustrated on the right in Figure 4, for each xy pixel in the desired viewpoint image we effectively traverse the ray away from the desired viewpoint in an attempt to estimate the most likely color for that pixel. To do so we test each voxel along the ray by back-projecting it into each of the eight cameras and looking at the actual camera image color at that point. We choose one "winner" voxel—the one with the most plausible back-projected appearance in all camera samples. We then use the median of the winner's back-projected camera sample colors as the surface (voxel) color estimate, and the position on the ray as the 3D coordinate of the surface point. We mark all winner voxels as opaque

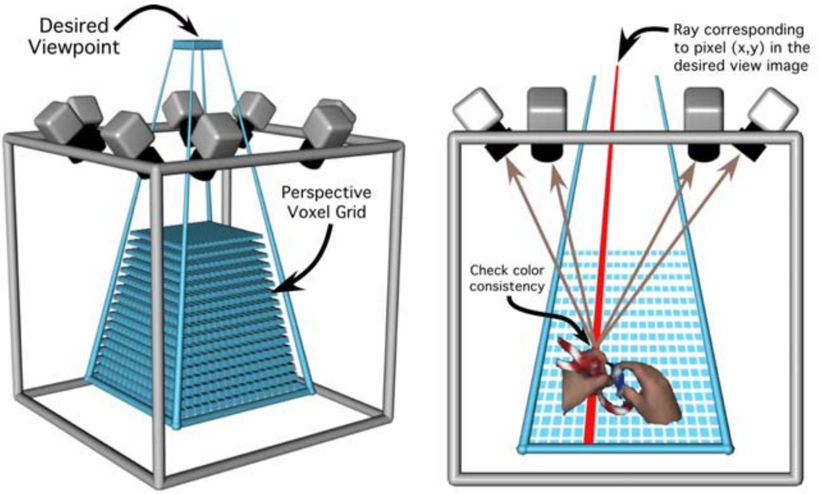

Figure 4: View-Dependent Pixel Coloring (VDPC). Left: an illustration of the *desired viewpoint* and *perspective voxel grid* concepts. Right: an illustration of the color consistency check for a particular voxel along a ray from the desired viewpoint image. The arrows extending from the hand toward the cameras depict the reflectance/imaging of the surface (voxel) into/by the cameras.

surfaces, and all others along the ray between the desired view and the surface as “empty”. As with the seminal *space carving* work by Kutulakos and Seitz [15] this effectively “carves away” voxels that do not appear to be part of a surface. We repeat this volume sweeping process, allowing estimated opaque surface points to occlude other voxels, progressively refining an estimate of a height field corresponding to the first visible surfaces. To improve robustness we employ a *view-dependent smoothness constraint* to (in effect) adapt the reconstruction support in textureless regions, and a *physically-based consistency measure* to factor out specular highlights. Both are described further in [40] and [41].

3.4 Authoring

For immersive content authoring we use a Barco Tan VR-Cube$^{\text{TM}}$ at Brown University. The VR-Cube$^{\text{TM}}$ is approximately three meters on each side, has four projection surfaces (three walls and a floor), and four Marquee 9500LC projectors configured for 1024×768 60Hz field-sequential stereo. The primary authoring input device is a hand-held Tablet PC. The author uses the Tablet PC to navigate the dynamic 3D points in time, to initiate snapshots, to highlight and annotate, and arrange snapshots hierarchically.

Our present approach to authoring includes a novel combination of 2D and 3D interaction techniques. The primary motivation for a hybrid 2D/3D approach is to provide a familiar and tangible means of 2D sketching (the notepad paradigm) while simultaneously offering a natural and immersive means of viewing the dynamic 3D data and the evolving IEBOOK. Figure 5 shows an author annotating an IEBOOK in the Tan VR-Cube$^{\text{TM}}$ and a screenshot of the authoring interface on the Tablet PC.

Using the VCR-like time controls shown on the right in Figure 5, the author can navigate time in the captured sequence, looking for a particular moment in time. She can then move her head to a viewpoint where she can see it well, and "take a snapshot" using a button on the Tablet PC. Using the same Tablet PC interface she can then highlight features, annotate the snapshot, and save the results to a virtual gallery in the IEBOOK as shown in Figure 6. She can then arrange the snapshots hierarchically by dragging them on the Tablet PC application.

Conventional methods for selecting and highlighting in 2D images do not accommodate depth or time. To overcome this limitation we have adapted some conventional methods to enable the selection and highlighting of dynamic 3D point/mesh data [26]. Specifically we have implemented three dynamic 3D highlighting paradigms: *marquee*, *freeform*, and *fill*. These methods are described in detail in [40]. An example of each is shown in Figure 7.

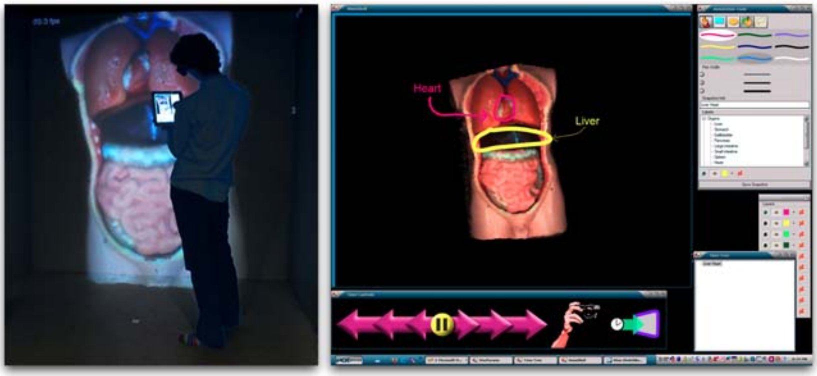

Figure 5: Authoring an IEBOOK. Left: an author in the VR-Cube^{TM}. Right: a screenshot of the interface on the Tablet PC.

The implementations preserve highlights throughout a sequence, and automatically adapt to changes in the 3D topology. The highlight information is stored in the IEBOOK as a series of 2D matrices of Boolean values that describe whether or not each of the vertices in the scene are highlighted in a given frame. Presently there are ten independent "layers" of highlights, each of which can be assigned to a different color. The layers can also each be hidden, allowing the user to see the results with or without their effects.

3.5 Content Viewing

Our primary paradigm for experiencing an IEBOOK is our VR-Cube^{TM} and a subset of the authoring interface described earlier. This is the richest IEBOOK environment we can offer, combining immersive, head-tracked, stereoscopic, 3D imagery, with Tablet PC-based navigation through time and a hierarchy of annotated snapshots. A "student" views the snapshot gallery holding the Tablet PC as in Figure 5. They can navigate through time using

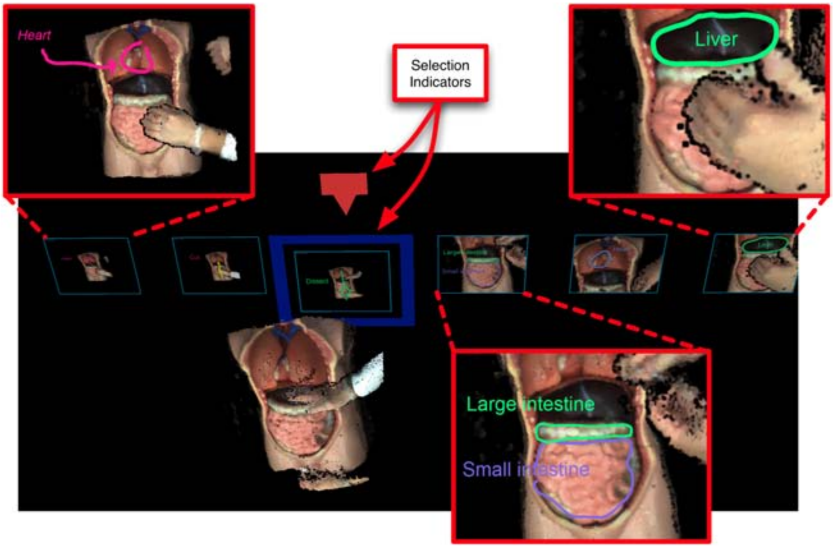

Figure 6: A screenshot from the VR-Cube$^{\text{TM}}$, showing several annotated snapshots arranged in an IEBOOK virtual gallery. We added "zoom boxes" for three of the snapshots, and a label to indicate the current selection markers (blue "halo" and red marker).

the VCR-like controls or by choosing a particular snapshot in the hierarchy (see Figure 6). A halo gives feedback of the active snapshot during the selection process, and once a snapshot is selected the model jumps to the time associated with that snapshot.

When trying to understand a new surgical procedure, one would like a high-quality up-close view of the procedure itself. But it is also critical to learn how to interact and deal with events and the confusion taking place all *around* the viewer. To engender this complete experience we want a display with high-fidelity stereo imagery at arm's length and a wide (surrounding) field of view. Head-mounted displays (HMDs) can provide the former, and projector-based displays can provide the latter, but unfortu-

Figure 7: Examples of our dynamic 3D highlighting paradigms: (a) marquee, (b) freeform, and (c) fill. The tip of a freeform arrow in (b) is partially obscured by the surgeon's hands, illustrating the 3D nature of the highlights. Note that this data set has not been filtered and triangulated.

nately we are not aware of any display system that simultaneously meets both needs. Hence we combined the two paradigms and developed a *hybrid display system* that integrates head-mounted and projector-based displays to simultaneously provide a high-quality stereoscopic view of the procedure as well as a lower quality monoscopic view for peripheral awareness. Details are presented in [13].

While we believe that head-tracked stereo 3D systems provide the greatest sense of immersion, they are typically available only in visualization labs. As another option for experiencing the results of our work, we have been working on methods for making our 3D reconstruction results available on the world-wide web. In consultation with our medical partners, we decided on three forms of media: anaglyphic (red-blue) stereo images, anaglyphic movies, and non-stereo but dynamic (in space and time) VRML models. We developed software to generate anaglyphic stills and movies from the 3D point/mesh data, and separate software that creates the "boxed" VRML data sets shown on the right in Figure 9. The result includes a 3D VCR-like user interface to play and step through a sequence, including a slider bar for "random access" similar to typical computer-based movie players.

3.6 Results

In Figure 8 we show several views of a dynamic reconstruction of Dr. Bruce Cairns, M.D. performing a mock surgical procedure on the physical patient model shown in Figure 3, and of co-author Adrian Ilie manipulating real Rhesus monkey and human skulls.

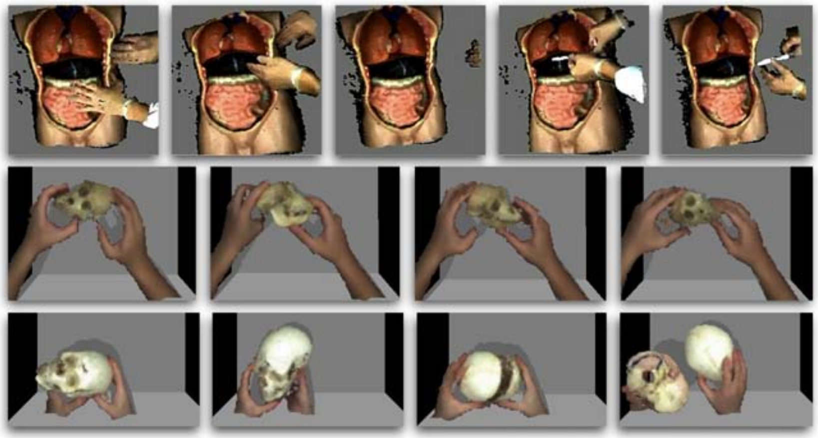

Figure 8: Reconstruction results. Dr. Bruce Cairns performing a mock procedure to manage blunt liver trauma (top), and another co-author manipulating a Rhesus monkey skull (middle) and human skull from a 12-year old child (bottom). The human skull is a teaching artifact that can be separated into multiple parts. Both skull reconstruction sequences (middle and bottom) include simulated dynamic shadows.

One of the skills most needed by new surgeons, and yet difficult to learn, is that of suturing. In particular, skills at knot tying can be critical—the wrong knot can cause a wound to open, or can result in tissue damage. The consensus among surgeons we know is that 2D images and movies of knot tying are woefully inadequate. For this reason we have made some preliminary attempts to acquire and reconstruct some fundamental knot tying events. We are able to reconstruct hands and rope of moderate thickness, and have done so for a few basic surgical knots, as well as some common sailing (or Boy Scout) knots. Figure 9 shows an anaglyphic stereo movie and the VRML interface of a ring knot tying reconstruction. At `http://www.cs.unc.edu/~ootf/Projects/ebooks/reconstructions/` we have made available results of some basic sailing knots tied by Herman Towles, and some common surgical knots tied by Dr. Bruce Cairns. We will continue to add more reconstruction results as we obtain them.

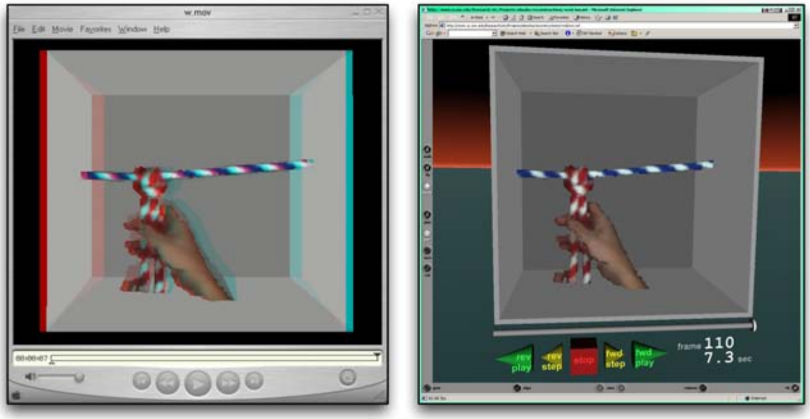

Figure 9: Ring knot tying. Left: anaglyphic stereo movie ("red over right"). Right: VRML interface with VCR-style controls. The movie includes simulated dynamic shadows, the VRML model does not.

At UNC Chapel Hill we have shown several surgeons the lab-based (stereo and head-tracked) and web-based (movies and VRML) reconstructions.[5] The reactions of the surgeons we have asked have been very positive. For example, Dr. Michael Jaskolka, an Oral & Maxillofacial Surgery resident at UNC Chapel Hill, said the

---

[5] Note that we have only shown the surgeons the reconstructions, e.g. as in Figure 8 and Figure 9, not the complete IEBOOK results as in Figure 6.

following.

> “I think that there is fantastic potential and *many* real life applications. I was extremely impressed with the replay and ability to have directed playback through movement of the head mounted display. I think that this technology will be a fantastic learning tool for both new surgeons and for experienced surgeons learning new techniques and methods. This has huge potential impact on patient care in non-tertiary centers.”

The primary limitations cited by most are the size, resolution, and visibility of the reconstructions. This is no surprise to us at this relatively early stage of the work. The choice and arrangment of cameras in our camera cube (Figure 3) was never intended to be optimized for any procedure. The choices were made based on what was practical for a proof-of-concept, supporting reconstructions of modest detail throughout a modest working volume.

4 On-Line Remote Consultation

Trauma is a significant health problem, frequently referred to as the “hidden epidemic of modern society” responsible for more lost productive years than heart disease, cancer, and stroke combined [4, 18]. In addition, since serious trauma can occur at anywhere and anytime, and it is clear that early, appropriate intervention saves lives, issues in trauma management have long been proposed as an ideal application for telemedicine [28]. While there have been some reports of success in limited trials using 2D video teleconferencing [25], issues such as viewpoint control and depth perception continue to limit the application and acceptance of the currently available systems [32, 34–36]. Tachakra reports in [31] that the “impaired depth perception (with 2D conferencing) is a significant problem in telemedicine.”

3D *telepresence* [30] has been described as the "ultimate development in telemedicine" but a concept that "remains in its developmental infancy" [28]. Based on his 2D telemedical application experiences, Tachakra [33] states that "the ideal (telemedicine) videoconferencing environment would produce the minimum feeling of artificiality for its user."

As such our vision of 3D telepresence for medical consultation has a corresponding overriding technology goal:

> *We want to create a physically natural and intuitive visual and aural sense of presence in a remote place, while minimizing a participant's awareness of the technology-related external factors—they should be able to ignore the technology and concentrate on the medical task.*

4.1 Overview

Our effort related to 3D telepresence for on-line medical consultation is relatively new, and hence we have only preliminary results. Our *research* is focused 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. In parallel we are undertaking a *systems integration* effort to focus our research and development on the scenarios (a)-(d) shown in Figure 10, which correspond to uses of permanent, portable, and hand-held advisor and advisee technologies. Finally, in an effort to balance usefulness and practicality, we are undertaking an *evaluation* effort aimed at assessing the fundamental effectiveness of *emergent airway management* in scenario (c) of Figure 10, where the advisor and advisee are using permanent but geographically-separated facilities.

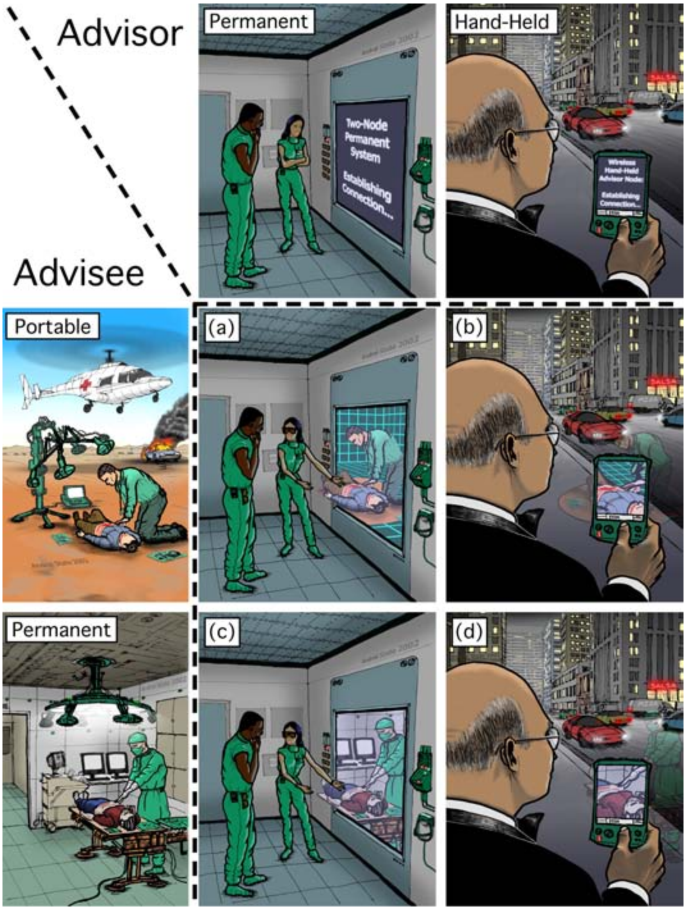

Figure 10: Future vision of 3D Telepresence for Medical Consultation. (Sketch by Andrei State.) The left column illustrates examples of portable and permanent 3D telepresence technologies used by an *advisee*. The top row illustrates examples of permanent and hand-held technologies used by an *advisor*. Images (a)-(d) illustrate the shared sense of presence for the corresponding advisor-advisee scenarios.

4.2 2D Acquisition

Our long-term goal is to equip a real Intensive Care Unit (ICU) room with cameras and projectors similar to the scenario depicted in the lower-left of Figure 10, and to be able to witness procedures remotely in 3D. As a first step in that direction we have built a Portable Camera Unit (PCU) and are using it in an ICU room of the University of North Carolina Hospital's Jaycee Burn Center to capture 2D video of real procedures. The PCU, shown in Figure 11, is a standalone system with two Sony pan-tilt-zoom cameras and a Dell workstation capable of storing to disk JPEG compressed streams at 30fps. We have used the PCU to record several tracheostomy procedures performed by co-author Dr. Bruce Cairns. From these early sessions we have already gained valuable insights into lighting and camera placement, which we expect to find useful when we move toward live 3D reconstruction.

4.3 3D Reconstruction

The full VDPC method presented in Section 3.3 does not run in real time (it typically requires minutes to render each frame set) and thus cannot be used "live" on line. However we have also developed a reduction of VDPC that can be efficiently implemented on modern commodity graphics hardware, which accelerates the view synthesis time by orders of magnitude, thus enabling on-line view synthesis of a live dynamic scene.

To take advantage of current graphics hardware 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.

Figure 11: Portable Camera Unit (PCU) being used to capture 2D video of a real procedure in an ICU room of the University of North Carolina Hospital's Jaycee Burn Center.

Under these assumptions, VDPC can be implemented entirely on commodity graphics hardware, taking advantage of the hardware's inherent parallelism. Rather than a single view ray at a time as illustrated in Figure 4, we use the graphics hardware to compute an entire voxel *plane* a time. The basic idea is first to project the input images on to each voxel plane, then to use programmable graphics hardware to compute the photo-consistency values of voxels and select the most consistent ones. Essential to this implementation are the multi-texture mapping functions and Pixel Shader, both of which have been available on commodity graphics hardware since 2000. We use multi-texture mapping functions to project input images onto voxel planes, and Pixel Shader to compute photo-consistency values and select the best colors. To address the typical aperture (kernel-size) problems related to photo-consistency measures, we use the MIP map hardware to implement a multi-resolution approach that greatly improves the robustness, at little additional computational cost [41].

Comparing this real-time VDPC to the full VDPC approach in Section 3.3 reveals two major differences. First, the visibility constraint is waived—every camera pixel is considered to have an unoccluded view of the desired surface. Second, the smoothness constraint is simplified and optionally applied right after the photo-consistency value is computed. From an algorithmic standpoint, real-time VDPC is similar to the plane-sweeping algorithm in [5], but by using the graphics hardware our approach supports the real-time reconstruction we need for medical consultation.

4.4 Model-Based Tracking and Refinement

Both the off-line and on-line (real-time) VPDC approaches strive to compute dense depth maps as fast as we can acquire images from our cameras. While this dense dynamic depth information is valuable, it is noisy, and as we scale up the number of cameras and the working volume, it will result in massive amounts of data with needs for massive bandwidth.

To address these and other issues, we are also investigating methods for *model-based reconstruction and tracking* of objects in a scene, and *continual refinement* of their models. The idea is that if one has a model of a scene object (e.g., polygonal or analytical) then instead of repeatedly performing dense depth maps corresponding to the object, one could simply track the object's dynamic pose (position and orientation). This offers the possibility of dramatically reducing the real-time bandwidth needs, and of offering higher-quality results from filtering and refinement.

Specifically, we are developing a Kalman filter-based method [14, 39] that uses an initial dense depth from a real-time version of VDPC to construct a 3D feature list corresponding to a set of 2D features detected in each camera view. We are currently using the KLT feature [17, 37] support in the Open Source Computer Vision Library (OpenCV), `http://www.intel.com/research/mrl/research/opencv/`. We are also considering new multi-modal and/or multiple hypothesis methods [11] to detect and eliminate features not associated with an object.

Finally, inspired by [7], we have also started looking at methods for more generalized tracking based on *Jacobian images*[6] of predicted camera images. For example, Figure 12 shows a model of a cube with a marble texture, then three Jacobian images. The white areas of the last (the Complete Jacobian) are the most sensitive to change, and as such places we would expect to have the opportunity to see geometric or texture features in the images. The darker areas exhibit little or no change with respect to motion, which means that they are uninteresting.

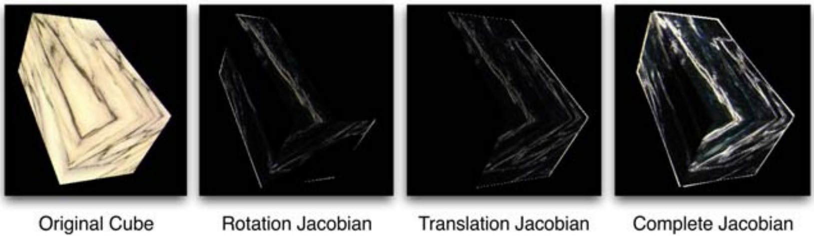
Original Cube
Rotation Jacobian
Translation Jacobian
Complete Jacobian

Figure 12: A model of a cube with a marble texture, and three corresponding Jacobian images: the change in the image with respect to one axis of rotation, one axis of translation, and the sum of all derivatives with respect to all axes of rotation and translation.

---

[6] Jacobian images are the derivatives of the camera images with respect to the position and orientation of the modeled object.

4.5 Network Systems

Comparing the scenarios illustrated in Figure 10, the available networking resources between two permanently installed consultation environments will likely be much richer than the resources available to a portable or mobile system. Within a scenario, network resources will be variable during the course of the consultation. This variation may have a number of causes including competing traffic on the network, bursty packet losses, or dynamic routing. Network variability, present in both high- and low-bandwidth situations, is a reality that can not be escaped. The primary objective of our network systems research is to develop mechanisms to accommodate this variable network quality (i.e., bandwidth, latency, jitter, etc.), both between different scenarios and during a particular scenario. Our approach is to develop an adaptation layer which is charged with estimating network conditions and availability, and with coordinating how network resources are allocated and used by the different flows of information within the application.

Specifically we have been investigating networking mechanisms that provide *aggregate* congestion control information for *groups* of flows between two sites. Traditionally, congestion control and evaluation of current network conditions is conducted independently by each flow of data on an end-to-end basis. Our approach recognizes that in complex multi-stream applications such as 3D telepresence, many flows that are semantically related must coordinate their view of network resources in order to cooperatively use those resources, instead of competing with each other.

We have designed a Coordination Protocol (CP) that provides aggregate congestion control and coordinated usage of network bandwidth for distributed applications. We have developed a kernel-level implementation, and have tested it in a telepresence application architecture. We have also performed an initial investigation of transport-level protocols built on top of CP that may prove useful for addressing our adaptation needs.

Beyond the CP mechanisms, we have been looking at how multidimensional adaptation policies can be effectively *specified* within a complex multi-stream application. The challenge here is to provide an intuitive interface to users for managing tradeoffs between different flows of information. Our approach has been to consider ways in which we may model the utility of different flows, in particular as the user interacts with the information. We have developed an abstract graph-based approach for modeling adaptation requirements and policies, and have begun to explore geometric metaphors for expressing multidimensional adaptation policies and for evaluating information utility within our graph-based abstraction.

4.6 Content Viewing

We discussed some permanent installation viewing paradigms in Section 3.5, but we are also working on the hand-held paradigm illustrated in the right column on images in Figure 10. The basic idea is that a doctor needed as a consultant might be away from a permanent high-fidelity installation, and only have access to a wireless Personal Digital Assistant (PDA) device.

Specifically, we are working toward autonomous or “self” tracking of the PDA, and on remote rendering techniques. Our current prototype uses our lab-based HiBall$^{\text{TM}}$ tracking system to estimate the pose of the PDA *relative to* a second device that serves as a “patient surrogate” (left image in Figure 13). We continually render a 3D view of point cloud data in the PDA, so that it appears at the relative pose of the patient surrogate device. The image to the right in Figure 13 shows Dr. Bruce Cairns and UNC Graduate Research Assistant Max Smolens using the prototype to look at some data from Section 3.6.

Figure 13: Left: a tracked PDA and a separate tracked "patient surrogate" device. Right: Dr. Bruce Cairns uses the PDA in a two-handed mode while student Max Smolens looks on.

4.7 Formal Evaluation

We believe that 3D telepresence could enable earlier diagnosis and treatment outside the realm of the responding on-site Emergency Medical Technician (EMT), emergency response team, or physician. However, currently no data exists to support or refute this premise. In addition, the National Academy's report on telemedicine evaluation [9] states that it is critical to examine the *acceptability* and *practicality* of telemedicine technology. Many factors may impact acceptability, including professional culture and image, the structure of the health care system, and the patients' perception of benefits [9]. Often these factors, which can negatively impact and even halt adoption, are not discovered until the technology is deployed.

As a first step in quantifying the likely effectiveness and acceptance of 3D teleimmersion, we have just begun a multi-year formal evaluation. Our planned evaluation consists of two major efforts: an experimental evaluation and a qualitative field study. The qualitative field evaluation addresses acceptance at the institutional and societal levels [9]. The experimental evaluation addresses evaluation primarily at the clinical level. We will compare medical treatment under four conditions:

- Condition 1: An EMT working alone (typical today).
- Condition 2: An EMT collaborating with an expert physician via state-of-the-art 2D video-conferencing (possible today, but not typical).
- Condition 3: An EMT collaborating with an expert physician who is physically present.
- Condition 4: An EMT collaborating with an expert physician via our best 3D telepresence prototype system.

Condition 1 is intended to verify that the selected task (below) is one that an EMT *cannot* perform alone. Condition 2 will allow a comparison between the best practically-available 2D televideo technology available today, and 3D telepresence technology (conditions 3 and 4). Condition 3 is intended to address expected effectiveness and acceptance, if a 3D telepresence system were *as good as it can be.* The physician will be *physically* present, but will not be able to touch the “remote” patientEMT. Condition 4 will utilize our current best 3D telepresence technology.

The selected experimental task is the management of the *difficult airway*, including the creation of a surgical airway (*cricothyrotomy*). EMTs are expected to be able to manage the difficult airway, yet their training in this task is limited. In addition, managing the difficult airway is challenging even for trained physicians, especially the first time. Even physicians experienced 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 There is a fairly well defined algorithm for performing this task, and there are a number of models that can adequately judge the performance.

We are using the Human Patient Simulator (HPS) shown in Figure 14 as a surrogate patient for the difficult airway task scenario. The HPS is a full sized computerized mannequin that is programmed to represent a variety of physiologic states and a variety of interventions. The HPS is designed to breathe, exhibit heart sounds, and otherwise respond to stress with lifelike physiologic data. It is particularly designed to assess difficult airway management and allows for endotracheal intubation and cricothyrotomy to be performed as shown in the inset on the left in Figure 14.

We recently completed a series of pilot sessions to test the above conditions. For the pilots we modified our portable camera rig (Section 4.2) as shown in Figure 14. We also built a state-of-the-art 2D videoconference station that allows the remote physician to view the EMT working on the HPS. The station (Figure 14) has four monitors that show the HPS patient data and three different views of the HPS and EMT. Two of these views can be controlled by the remote physician via a graphical interface to the pan-tilt-zoom cameras on the Portable Camera Unit.

The pilots were very useful. All EMTs gave us valuable feedback regarding the simulation, scenario, data collection instruments, recruiting strategies and interaction with physicians in general. Iterative improvements to the sessions were made along the way. Early results indicate the simulation scenario is engaging and involves a complex diagnosis and treatment that is challenging for many EMTs working alone. For example, the statement "I was absorbed intensely in the activity" received an average response of 6.6 on a 7-point scale, where 7 indicates "I strongly agree."

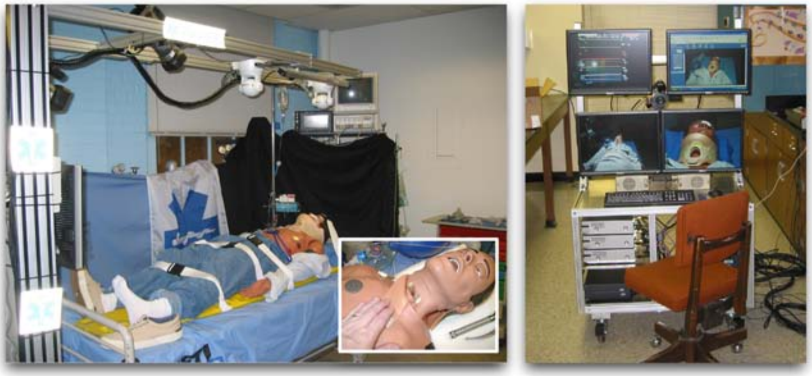

Figure 14: Left: modified Portable Camera Unit (c.f. Figure 11) positioned over the Human Patient Simulator; inset shows a close-up of the access for difficult airway management. Right: 2D videoconference station for the remote physician.

5 Camera Placement

Scaling up the camera cube in Figure 3 is not as simple as it might seem. Clearly we will need more cameras to cover an operating room, yet from a practical standpoint we want to minimize the number. So how many cameras do we need? The problem goes beyond simply looking at the "coverage" of the cameras—we need to consider the quality of the "signal" from the cameras. The quality is affected by things like resolution, accuracy, focal length, sample rate, and various noise sources. In general the reconstruction quality will decrease with increasing distance from cameras, lower sample rates, less reflective surfaces, and lower light levels. One wants to minimize the number of cameras while maintaining sufficient visibility, resolution, and signal quality.

Some researchers have been working on fully automatic methods for determining optimal camera placement, but only for a small number of cameras under constrained geometry [21]. Not only do we need to optimize a less constrained placement of a relatively large number of cameras, we also want to optimize the focal length, frame rate, etc. With such a large optimization space, the problem quickly becomes intractable. Indeed when looking at the classical art gallery "museum guard" problem, Ntafos found that the minimal guard coverage of a three-dimensional grid is NP-complete [20].

To address this problem we are developing two interactive tools (geometric and stochastic) to support human-in-the-loop optimization of the design of the acquisition system. The idea is to provide system designers with intelligence amplification tools, so that they might find (or at least develop intuition about) candidate solutions.

5.1 Geometric Planning Tool

Our first tool is a software simulator designed to help us plan camera positions and orientations in a scene where we want to perform 3D reconstruction. This tool is *geometric* in that it only takes into account a camera's field of visibility, resolution, and occlusions, as opposed to the *stochastic* tool described below in Section 5.2, which also takes into account the expected scene dynamics, the camera frame rate, measurement noise, etc.

Using this geometric planning tool we can evaluate an arrangement of cameras within a given 3D environment with respect to the following issues:

1. Which parts of the scene are visible to each camera?
2. For each camera, what is the imaging resolution of the visible parts of the scene?

The simulator is an interactive 3D graphics program that loads a set of 3D models representing the desired scene (room, furniture, people, props, etc.) and allows the user to specify a number of cameras in terms of position, orientation, and field of view. See Figure 15. The program treats the cameras as shadow-casting light sources. Any part of the scene that is shadowed from a light source is invisible to the corresponding camera. By using a different color for each camera's light source, and by applying additive blending, we can visualize the areas covered by one or more cameras, as well as the areas that are completely shadowed (i.e. invisible to all cameras) thus answering the first question. We can also project a user-specified pattern onto the scene from each camera's light source. By using a grid pattern aligned with a camera's pixel grid we can visualize the approximate imaging resolution for any visible element of the scene, providing an answer to the second question. See Figure 15 for examples.

Figure 15: Multi-camera visibility simulator. Left: the cameras (floating cylinders at the left) are treated as flat-field colored lights. Center: the same view with projected grids to show each camera's imaging resolution. Right: a close-up showing camera visibility overlap areas.

5.2 Stochastic Planning Tool

We are also working on a stochastics-based tool to help us estimate and visualize 3D reconstruction *uncertainty* throughout the acquisition volume, for a particular candidate set of cameras.

We call our current prototype *Artemis* after the Greek goddess of hunting. Similarly to fluid or air-flow visualizations the idea is to make the invisible camera signal information visible to designers so that they can develop insight into the effects of design choices. Using Artemis the designer would look for "hot spots" or regions of sub-par information throughout the acquisition space, and then interactively adjust the design (camera choices, configurations, and parameters) to improve the performance.

Figure 16 shows a visualization of an actual seven-camera setup we have used for VDPC experiments. The small blue boxes represent the seven cameras. This example is noteworthy in that we learned something from it. In building the original seven-camera setup we analyzed the camera geometry to maximize the frusta overlap. From Artemis we learned that the information content (signal strength) dissipates significantly within the overlap region. As a result, we now realize that our reconstruction targets should be moved closer to the cameras—as close as possible. In retrospect this makes perfect sense. However despite our intuition and careful planning, the dissipation of the signal in the overlap region never entered our minds.

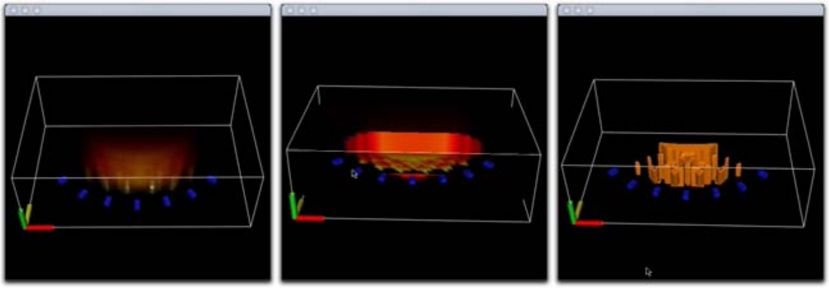

Figure 16: Artemis visualizations of a camera setup for scene reconstruction. Left: volume visualization. Middle: Y-slice and Z-slice. Right: isosurfaces. The small blue "actors" represent the cameras.

6 Future Work

While we are pleased with our progress and achievements during the past several years, we recognize that we are still in the relatively early stages of two long-term efforts. Many problems remain to be solved, and much work must be done to achieve the scale, fidelity, flexibility, and completeness that we envision.

For example, our reconstructions are relatively small with low resolution: a cubic volume approximately 40 cm on a side, with approximately three samples per cm. In addition we have not addressed visibility/occlusion problems. Our overall objective with respect to acquisition and 3D reconstruction is to *scale up the working volume* while maintaining sufficient *accuracy, resolution, and visibility.* Achieving this overall objective will require new methods that address the following specific issues.

- *Simultaneous wide-area and high resolution acquisition.* We effectively want millimeter-resolution 3D reconstructions over an entire room. The number of cameras needed for a brute-force approach would be impractical or impossible in terms of bandwidth and processing requirements.
- *Static occlusions.* Standing surgeons and fixed equipment can block critical camera views of important activity. One must choose cameras and camera placement carefully.
- *Dynamic occlusions.* Hands and instruments will pass in front of cameras, temporarily blocking their views. One must detect this and avoid using the occluded images.
- *Reconstruction, filtering, and representation.* Reconstructions from sparsely arranged cameras will contain noise. Millimeter-resolution reconstructions of dynamic events will result in massive amounts of data, well beyond the capabilities of current computers.
- *Robustness and quality.* Methods must work under dynamic

> lighting conditions, in the presence of scene motion, if cameras are bumped, etc., while producing results that are pleasing and accurate. In the spirit of the *prime non nocerum* dictate ("first, do no harm") the reconstruction process must not introduce unacceptable inaccuracies, and estimated uncertainties should be maintained with the data.

In addition to scaling up our capabilities in terms of the acquisition and reconstruction volume, we hope to increase the impact on the medical community by making complete IEBOOKs available on the web. The primary difficulty here is in determining which interaction techniques are appropriate, and how they can be implemented. We feel strongly that simply "dumbing down" the fully immersive interfaces is the wrong approach. Instead we want the best interfaces for each paradigm, and authoring tools that appropriately target each.

Much also remains to be done specifically with our NLM contract for "3D Telepresence for Medical Consultation." For example we are continuing to work on a permanent acquisition setup in our lab, including the planned installation of cameras and projectors on a used surgical lamp. And while we have performed a brief pilot user study as described in Section 4.7, we have the complete set of user studies ahead of us still.

We also hope for some near-term experiments involving real patients. We have procured four 1024×768 *stereo* color Bumblebee cameras from Point Grey Research, and are awaiting shipment of a five-node compute cluster with high-performance CPUs and ATI Radeon X800XT graphics accelerators. The servers will be housed in a hard-shell shock-resistant rack mount system so that we can transport it between the hospital and our lab as needed for experiments. Our plan is to implement the real-time VDPC described in Section 4.3 on that transportable cluster, and to use it to stream 3D+color from an ICU to our lab.

7 Acknowledgements

We want to acknowledge past and present collaborators on related work, including Jaron Lanier; Amela Sadagic at Advanced Network and Services; Ruzena Bajcsy at the University of California at Berkeley; Kostas Daniilidis at the University of Pennsylvania; and Marc Pollefeys, Kevin Jeffay, Dr. Anthony Meyer, and Dr. Ramon Ruiz at the University of North Carolina at Chapel Hill.

At Brown University we want to acknowledge Dmitri Lemmerman and Melih Betim for their assistance working with the VR-Cube$^{\mathrm{TM}}$ hardware and software; Mark Oribello for his audio-visual help; Eileen Vote and Olivia Linder for their models; Michael Kowalski for an early "splatting" implementation; and Bob Zeleznik and Noah Wardrip-Fruin for their interaction ideas.

At the University of North Carolina at Chapel Hill we want to acknowledge Jim Mahaney, John Thomas and David Harrison for their technical support; Myra Gwin-Summers for her continued assistance—and for the animal bones; Graduate Research Assistants Max Smolens for his work on the PDA scenario prototype (Section 4.6), and both Erica Stanley and Marcel Prastawa for work on the Artemis prototype (Section 5.2).

At Göteborg University and the University College of Borås we want to acknowledge graduate student Hanna Maurin for her work on the formal evaluation (Section 4.7).

This research was primarily supported by National Science Foundation (NSF) grant IIS-0121657: "Electronic Books for the Tele-immersion Age: A New Paradigm for Teaching Surgical Procedures," and 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] Jean-Yves Bouguet. Camera calibration toolbox for MATLAB^{TM}. `http://www.vision.caltech.edu/bouguetj/calib_doc/`.

[2] J. F. Brinkley. The digital anatomist project. In *Proceedings American Association for the Advancement of Science, Annual Meeting*, page 6, San Francisco, February 18–23 1994.

[3] H. Clark. *Using Language*. Cambridge University Press, Cambridge, UK, 1996.

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

[5] Robert T. Collins. A space-sweep approach to true multi-image matching. In *Proceedings of the 1996 Conference on Computer Vision and Pattern Recognition (CVPR '96)*, page 358. IEEE Computer Society, 1996.

[6] MD David G. Ellis, PhD James Mayrose, MD Dietrich V. Jehle, MD Ronald M. Moscati, and MD Guillermo J. Pierluisi. A telemedicine model for emergency care in a short-term correctional facility. *Telemedicine Journal and e-Health*, 7(2):87–92, 2001.

[7] Frank Dellaert, Sebastian Thrun, and Charles Thorpe. Jacobian images of super-resolved texture maps for model-based motion estimation and tracking. In *IEEE Workshop on Applications of Computer Vision (WACV'98)*, pages 2–7, Princeton, NJ, 1998. IEEE Computer Society.

[8] Paul Dourish, Annette Adler, Victoria Bellotti, and D. Austin Henderson Jr. Your place or mine? learning from long-term use of audio-video communication. *Computer Supported Cooperative Work*, 5(1):33–62, 1996.

[9] Marilyn J. Field and Institute of Medicine (U.S.) Committee on Evaluating Clinical Applications of Telemedicine. *Telemedicine : a guide to assessing telecommunications in health care.* National Academy Press, Washington, D.C., 1996.

[10] Henry Fuchs, Gary Bishop, Kevin Arthur, Leonard McMillan, Ruzena Bajcsy, Sang Lee, Hany Farid, and Takeo Kanade. Virtual space teleconferencing using a sea of cameras. In *Proceedings of the First International Symposium on Medical Robotics and Computer Assisted Surgery*, Pittsburgh, PA, 1994.

[11] P. D. Hanlon and Peter S. Maybeck. Multiple-model adaptive estimation using residual correlation kalman filter bank. In *Proceedings of the 1998 Conference on Decision and Control*, pages 4494–4495, Tampa, Florida, December 1998.

[12] Adrian Ilie. Ensuring color consistency across multiple cameras. Technical Report TR 04-12, University of North Carolina at Chapel Hill, Department of Computer Science, 2004.

[13] Adrian Ilie, Kok-Lim Low, Greg Welch, Anselmo Lastra, Henry Fuchs, and Bruce Cairns. Combining head-mounted and projector-based displays for surgical training. *Presence: Teleoperators and Virtual Environments*, 13(2):128–145, April 2004.

[14] Emil Kalman, Rudolph. A new approach to linear filtering and prediction problems. *Transaction of the ASME—Journal of Basic Engineering*, 82(Series D):35–45, 1960.

[15] Kyros Kutulakos and Steven Seitz. A theory of shape by space carving. *International Journal of Computer Vision*, 38(3):199–218, 2000.

[16] Alan Liu, Frank Tendick, Kevin Cleary, and Christoph Kaufmann. A survey of surgical simulation: Applications, technology, and education. *Presence: Teleoperators and Virtual Environments*, 12(6), 2003.

[17] Bruce D. Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision (ijcai). In *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI '81)*, pages 674–679, April 1981. A more complete version is available as Proceedings DARPA Image Understanding Workshop, April 1981, pp.121-130. When you refer to this work, please refer to the IJCAI paper.

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

[19] M.D. Michael G. Kienzle. Rural-academic integration: Iowas national laboratory for the study of rural telemedicine. Technical report, National Laboratory for the Study of Rural Telemedicine, 2000.

[20] C. Ntafos, Simeon. On gallery watchmen in grids. *Information Processing Letters*, 23(2):99–102, 1986. NP-complete cover problem in 3D. See email from Danette Allen.

[21] Gustavo Olague and Roger Mohr. Optimal camera placement for accurate reconstruction. *Pattern Recognition*, 35:927–944, 2002.

[22] G.M. Olson and J.S. Olson. Distance matters. *Human-Computer Interaction*, 15(2-3):139–178, 2000.

[23] Judith S. Olson and Stephanie Teasley. Groupware in the wild: lessons learned from a year of virtual collocation. In

Proceedings of the 1996 ACM conference on Computer supported cooperative work, pages 419–427. ACM Press, 1996.

[24] W. Orlikowski. Learning from notes: Organizational issues in groupware implementation. *The Information Society*, 9(3):237–252, 1993.

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

[26] Daniel M. Russo. Real-time highlighting techniques for point cloud animations. M.s. thesis, Brown University, Providence, RI USA, 2004.

[27] Bruce J. Schachter. Computer image generation for flight simulation. *IEEE Computer Graphics and Applications*, 1(4):29–68, October 1981.

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

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

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

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

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

[33] S. Tachakra, R. Newson, R. Wootton, and A. Stinson. Avoiding artificiality in teleconsultations. *J Telemed Telecare*, 7(Suppl 1):39–42., 2001.

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

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

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

[37] C. Tomasi and Takeo Kanade. Shape and motion from image streams: a factorization method - part 3 detection and tracking of point features. Technical Report CMU-CS-91-132, Computer Science Department, Carnegie Mellon University, Pittsburgh, PA, April 1991.

[38] Andries van Dam, Henry Fuchs, Sascha Becker, Loring Holden, Adrian Ilie, Kok-Lim Low, Anne Morgan Spalter, Ruigang Yang, and Greg Welch. Immersive electronic books for teaching surgical procedures. In Susumu Tachi, editor, *Telecommunication, Teleimmersion and Telexistence*, pages 99–132. Ohmsha, Ltd. (Tokyo, Japan), The University of Tokyo, Tokyo, Japan, December 2002.

[39] Greg Welch and Gary Bishop. An introduction to the Kalman filter. Technical Report TR95-041, University of North Carolina at Chapel Hill, Department of Computer Science, 1995.

[40] Greg Welch, Ruigang Yang, Sascha Becker, Adrian Ilie, Dan Russo, Jesse Funaro, Andrei State, Kok-Lim Low, Anselmo Lastra, Herman Towles, Bruce Cairns, M.D., Henry Fuchs, and Andy van Dam. Immersive electronic books for surgical training. *IEEE Multimedia*, 2004.

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

[42] Russ Zajtchuk and Richard M. Satava. Medical applications of virtual reality. *Commun. ACM*, 40(9):63–64, 1997.