Immersive Electronic Books for Teaching Surgical Procedures

Andries van Dam, Brown University
Henry Fuchs, University of North Carolina at Chapel Hill
with Sascha Becker, Loring Holden, Adrian Ilie, Kok-Lim Low, Anne Morgan Spalter, Ruigang Yang, and Greg Welch

Abstract

This paper reports on early progress with the use of immersive virtual reality technology for trauma surgery training. We discuss our technical goals and the application area, and then describe our work to date. The need to create a system that can be used by overworked, highly time-constrained surgeons and surgical trainees has affected many of our decisions, from the type of displays used to the focus on time navigation (to let trainees experience important moments and skip well-understood ones) to the use of easily-learned traditional 2D interfaces (vs. more demanding innovative 3D interfaces) for some of the interaction methods. This three-year research project, which is just entering its second year, is supported by a National Science Foundation Information Technology Research grant for collaborative research between groups at Brown University and the University of North Carolina at Chapel Hill.

1. Introduction

Immersive 3D virtual environments (IVR) provide a promising medium for combining real-world geometry and data and simulated experience. IVR can bring together the flexibility of abstract intellectual structures such as story-telling narratives, pedagogical techniques, and even “time travel,” with accurate real-world data gathered at different levels of detail from both 2D and 3D sources. While there are many potential application areas for such novel technologies—including maintenance and repair, virtual prototyping, and archaeological reconstruction—in this collaborative research project between Brown University and the University of North Carolina at Chapel Hill (UNC) we focus on a particularly socially compelling and technologically challenging driving application: teaching procedures for trauma surgery. We focus on a class of trauma injuries that are both potentially lethal and also particularly difficult to treat, namely blunt liver traumas.

Today, the pace of surgical innovations has increased dramatically, as have the societal demands for safe and effective practices. The mechanisms for surgical training and re-training suffer from inflexible and overconstrained timing for both instructors and trainees, as well as limited available content – that is, there are a limited number of operations which are suitable as teaching cases, a finite number of surgeons who are talented educators, and most importantly, an increased risk of complications associated with slowing down to teach during trauma surgeries. Traditional videotaped instruction has long been available to help teach surgeons new procedures, but these videos are only marginally effective because of their fixed point of view and lack of both depth perception and interactivity. The experience of watching a video is simply not sufficiently close to experiencing the actual procedure.

We are developing a new approach to teaching surgical procedures that combines immersion with time navigation in a hypermedia environment – *immersive virtual reality books (ivrBooks).* Our goal is to allow surgeons and surgeon trainees to witness and explore a past surgical procedure as if they were there, with the added benefit of commentary from the original surgeon or another instructor, as well as integrated 2D and 3D illustrations, annotations, and relevant medical metadata. The data gathered for a surgical experience will include the patient's biochemistry and vital signs, various types of diagnostic scans, real-time audio, etc.. The trainees will be able to freely and naturally walk around a life-sized, high-fidelity 3D graphical reconstruction of the original time-varying events, pausing or stepping forward and backward in time to satisfy curiosity or allay confusion. Existing tools for time-varying 3D scene capture and the necessary authoring tools are still crude at best; interaction metaphors and “cinematic” techniques for this fundamentally new medium are as yet unexplored.

The system has three main components:

1. Real-time capture of the original live procedure, followed by geometry reconstruction from the captured data,
2. Creation of the ivrBook, and
3. Use of the ivrBook by a trainee.

In the latter two phases, users can move about the (virtual) operating room as if they were there, stepping back to see the overall context or moving up close to see the details. During ivrBook creation, the surgeon, working with a media specialist, is be able to annotate the virtual 3D scene with analyses, illustrative explanations, imported 3D models, diagrams, and other pre- and post-operative material. During ivrBook "reading," trainees will be able to experience the level of detail appropriate to their background and goals (see Figure 1, which is an illustration of a "reading" experience in a CAVE™, Brown's IVR environment).

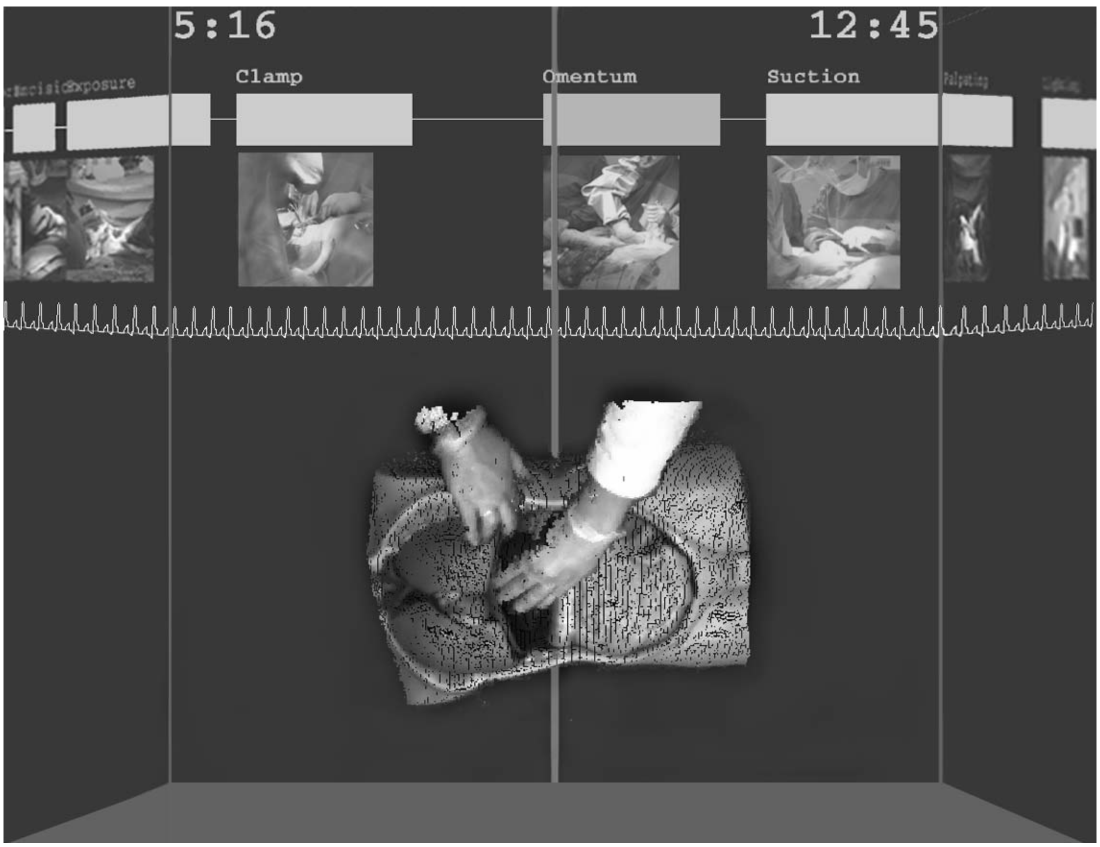

Figure 1: ivrBook of a liver operation. This scene shows a reconstruction of a surgeon manipulating the internal organs of a surgical dummy, a timeline with "now" in the center, and a heartbeat recording.

If successful, the surgical ivrBook training system will offer most of the positive aspects of traditional media such as textbooks, articles, and videos, while avoiding limitations such as the use of static 2D illustrations to represent 3D time-varying geometry, and the fixed camera angle(s) of a video. In many instances, a well-designed ivrBook will offer teaching opportunities that even real-life observation of a surgical procedure would lack. For example, the ivrBook user can adopt any viewpoint, including those obscured in the real-life procedure, and the ivrBook can provide varied levels of detail, as appropriate for each user. Users can access rich information sources including text, animations, a patient's chart, and the surgeon's narration. Our goal is to have the ivrBook be quick to produce and disseminate (produced within a day or two and requiring only a few hours of effort by the surgeon and the media specialist), offering trainees a chance to "be there" and engage in an effective, learner-centered learning experience that does not endanger the patient.

2. Related Work

A variety of approaches have been attempted in acquisition and reconstruction of 3D room-sized environments. Among the earliest is one of ours, [1], which proposed a “sea of cameras” around a room such as a surgical suite, to capture, in 3D, the real-time activity for remote consultation. Some of our early results from collaborations with Bajcsy’s group at Pennsylvania and Kanade’s group at CMU showed, in a virtual environment, faces and castle models reconstructed from multiple video images[2]. Subsequently, Kanade’s group [3, 4] has reconstructed moving individuals within a space enclosed by some 50 video cameras. In most of these systems, the extraction of 3D surfaces used stereo correlation-based methods, and the reconstructed renderings used polygons fitted over the surface points. More recently, the Blue-C project at Zurich, [5] seeks to put 3D reconstructions of remotely collaborating humans inside virtual environments, using a surrounding environment of wall-sized rear-projection displays, combined with a similarly-surrounding ring of cameras. Among their many intended applications is surgical training. For rapid acquisition and 3D reconstruction of human figures, Matusik et al [6] presented an efficient method using the intersection of silhouettes (“visual hulls”) from several cameras. For teleconferencing applications, a similar approach described in [7], and [8] focuses on reconstructing individuals into a common virtual environments. With a more general, but computationally expensive approach, Kutulakos [9] has demonstrated impressive volumetric space-carving methods. Agrawal and Davis [10] present a sophisticated probabilistic approach that might reconstruct objects with specular and translucent surfaces. In all these systems, to our knowledge, the example reconstructions to date have been discrete objects, not entire room scenes --with walls, furniture, people, and the detailed clutter of real life. This level of complexity is far beyond the design specifications of simpler, silhouette-based methods, and presents a considerable challenge to even the most general of these approaches.

In this early first prototype of the ivrBook authoring and 'reading' environment, we have built directly on Xerox PARC's Perspective Wall [11] work, applying it to the IVR environment for presenting the timeline and vital signs data on the back wall of the CAVE. As we proceed with our research we will be building on Catherine Marshall's work on hypertext annotation [12], Lynda Hardman's time-based hypermedia work with the Amsterdam Hypermedia Model [13], and Nitin Sawhney's hypervideo work in accessing and manipulating multiple conversational streams in a hypermedia video [14].

3. Capture of the Live Procedure: Scene Acquisition and Reconstruction

UNC has been concentrating on two aspects of the ivrBooks research: (1) improved scene acquisition and reconstruction, and (2) improved displays for both the surgeons and trainees using the ivrBook.

While in the distant future we envision an operating theater covered with a “sea of cameras” our present acquisition research primarily involves the use of a one-meter square multi-camera setup meant to mimic a small number of cameras mounted on a rigid surgical lamp. We have constructed a one meter cube “camera cube” (see Figure 2) that consists of an aluminum frame with eight downward-facing cameras mounted on the top face. We are currently using PointGrey Firewire™ cameras with built-in (hardware-based) synchronization circuitry. We have connected these cameras to four PC-based image acquisition servers. With this setup we can collect eight 15 frame-per-second synchronized color image streams, on line, in real time, while writing the data to disk. Our rationale for the carefully constructed and controlled setup is to restrict the set of problems we must face at one time. If we cannot achieve acceptable results here, it would be hard to imagine (or argue) that we would in a real operating theater.

Figure 2. Camera cube.

In terms of algorithms for acquisition and reconstruction, the larger UNC "Office of the Future" team [15] has been pursuing two paths of investigation. Both approaches are aimed at improving the 3D reconstruction by harnessing more resources for this computationally-intensive problem. One effort is part of a related project involving UNC, the University of Pennsylvania, and the Pittsburgh Supercomputing Center (PSC). In that effort, the scene reconstruction task is distributed among the more than 3,000 1Ghz processors of the largest computer at PSC. This effort is aimed (at the moment) more at harnessing powerful computers to achieve real-time tele-immersion.

A second effort at UNC, which is the research topic of Ph.D. student Ruigang Yang, is more directly targeted at our ivrBook project. A key early observation with respect to this work is that in some circumstances there can be significant benefit to a view-dependent *image-based* approach. The typical approach to the novel view problem is to attempt to reconstruct a full 3D model of the scene, and then render the 3D model. This is often called a "forward" approach because it follows the typical rendering paradigm, where a piecewise continuous 3D model is rasterized into a 2D image. Instead Yang began exploring an image-based approach aimed at estimating the most *likely* pixel colors for the 2D image corresponding to a desired 6D pose (3D position and orientation). Because this involves querying from the desired image into the (implicit) 3D scene, it can be thought of as an "inverse" approach. Forward approaches typically have the advantage that once a 3D model has been obtained, one can use conventional graphics hardware to render that model from any desired view, and that rendering is independent of (decoupled from) the scene acquisition and modeling. However if one can quickly and reliably estimate the pixel colors for the 2D images corresponding to moving viewpoint, then in some cases one does not need the full 3D scene description. And yet for our ivrBook, we actually *want* a full 3D model. As it turns out, with a small change in Yang's algorithm we are able to obtain a full 3D description, with better results that a typical forward approach. We describe this new approach briefly here.

The basic approach makes use of the Pixel Shader technology in the latest graphics processing units to calculate the values for many pixels in parallel. For a desired new view C_n (the red dot in Figure 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 an 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. In Figure 4, we show a number of images from different depth planes. Note that each of these images contains the projections from all input images, and the area corresponding to the intersection of objects and the depth plane remains sharp. 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.

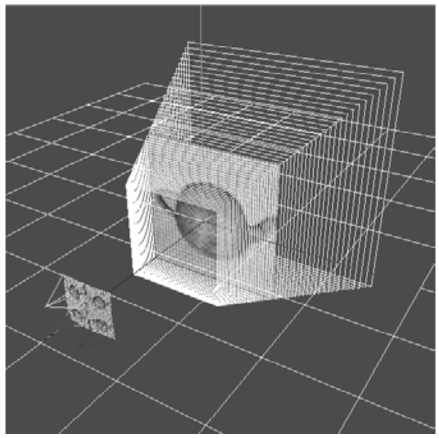

Figure 3: A configuration where there are five input cameras. The red dot represents the novel view point. The 3D space is discretized into a number of parallel planes.

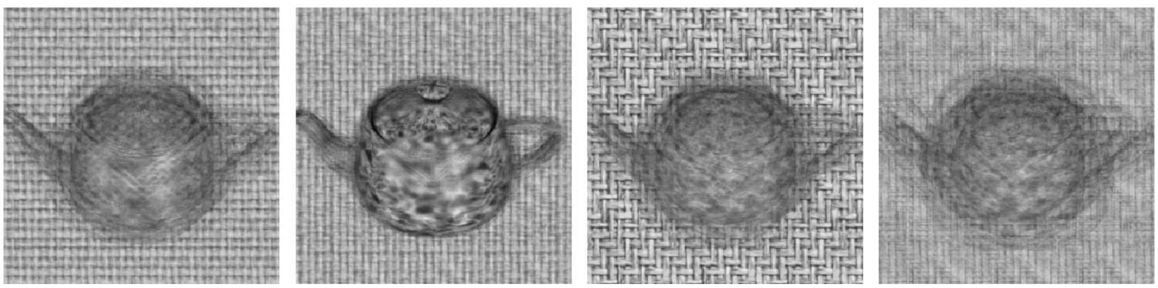

Figure 4: A synthetic scene with a teapot and a background plane is discretized into 50 planes. Left to right: depth plane images 0, 14, 43, and 49.

The concept of sweeping a plane through a discretized space is not new—it has been used in a number of computer vision algorithms for scene reconstruction [9, 16, 17, 18]. However the basis of our approach in [19] is to combine scene reconstruction and view synthesis into a *single* step using the programmable Pixel Shader technology typically available on modern graphics hardware.

While our current approach to scene reconstruction works well for scenes with diffuse objects, the scene in the OR, and even our training model, will certainly *not* be diffuse in most cases of interest to our ivrBooks project, in particular efforts aimed at surgical training. Blood and other bodily fluids associated with surgical procedures result in surfaces with very specular appearance (see Figure 5). We are presently extending our approach to include much more sophisticated probabilistic models. The results (thus far) are improved 3D models in the presence of surface areas of both constant color and specular highlights. The price is increased computational complexity, and a loss of real time, on line computation. However, for our ivrBook effort, this is an acceptable tradeoff, one that will arguably improve as computing power (including graphics computation) continues to improve.

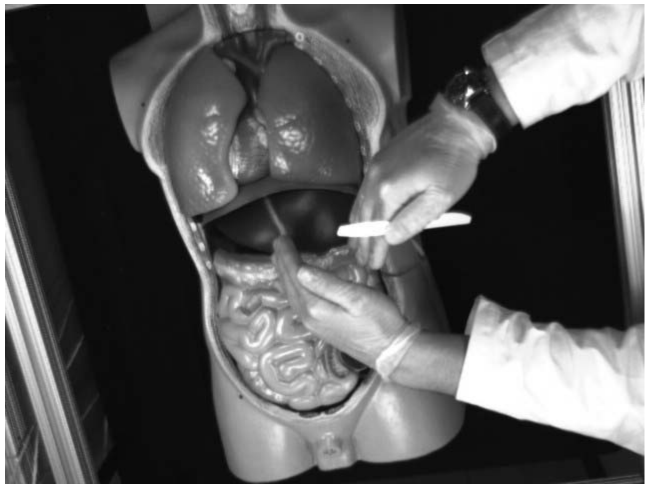

Figure 5: Specular highlights are visible in this image from a surgical training mockup and would be even worse in a real surgical situation.

Our reconstruction method, like many others, will generate a point cloud. It can be turned into a triangular mesh, for example, using the volume method from Curless and Levoy [20]. Or it can be further processed for filtering or simplification using a point-based approach, such as the one proposed by Pauly and Gross [21].

4. IVR for the ivrBook

Many challenges remain in the area of acquisition and reconstruction, but, as has occurred in rendering, we expect new algorithms and ever faster hardware to provide continuous incremental improvements. The challenges that lie in the field of relevant display technology are more daunting on several fronts. To begin with, it is not clear that display hardware configurations can be incrementally improved. And, while scene acquisition and reconstruction is being done with of-the-shelf components, display research may well call for improvements in technology that we have no control over, such as the quality of LCD displays.

Our current efforts focus on the problem that no existing display seems to satisfy the ivrBook's requirements, namely to be able to walk around in a room-sized environment, observe detail in nearby regions (such as the surgical site), and observe activity in surrounding areas (the other individuals in the surgical suite). Current displays fall short either because they offer only a narrow field of view, such as with HMDs, or they restrict trainees to spaces smaller than a real OR, such as a CAVE. Although many techniques for spatial navigation have been developed to address the small size of a CAVE, we feel that real physical motion through the space is vital for trainees to understand all the events taking place. As a proof of concept and a means of investigation, we have implemented a simple prototype of a hybrid display system (see Figure 6) that combines a high-resolution stereoscopic HMD display with monoscopic table-top and wall-based projections, as described in [22]. The wall projections are provided to give trainees a sense of context and are meant be seen in peripheral vision only. The table-top projections are provided to give trainees some sense of motion down under/below the HMD. This narrow role for the projected images means that display configurations much less expensive than CAVEs can be used.

Figure 6: The hybrid display consists of a head-mounted display and several wall-based projections.

For our first prototype of the hybrid display, we created a simplified synthetic model of an operation room with a virtual patient lying on the surgical table and three virtual surgeons or nurses moving around near the table. Sometimes, one of the surgeons/nurses would extend her arm towards the center of the table, mimicking a real nurse handling a surgical tool to a surgeon during a real surgery. When we tested the prototype ourselves, we would attempt to act in a way a surgical trainee would when using an actual training system. We tried to concentrate on the virtual patient's abdominal area while attempting to maintain awareness of the movements of the virtual surgeons around us.

For comparison, each of us had opportunities to try two different setups, one with the use of both the HMD and projectors (the hybrid display setup), and another with the use of the HMD alone. For the HMD-only setup, the projectors were switched off and the laboratory was kept dark. After having experienced the two setups, every one of us felt that with the hybrid display setup it was much easier to maintain awareness of the movements of the virtual surgeons/patient while concentrating on the patient's abdominal area. On the other hand, with the HMD-only setup, it was almost impossible to know the positions of the virtual surgeons without frequently turning our heads to scan the surroundings.

For a more rigorous and objective evaluation of the hybrid display we conducted a formal human subject experiment involving 25 subjects randomly recruited from the UNC-Chapel Hill student population. Again, our goal was to compare the effectiveness of our hybrid display (HMD-projector) with an HMD-only display. In keeping with the visual requirements of our surgical training application, our primary criterion for the experimental tasks was that they should force the subject to concentrate on a central up-close task while simultaneously maintaining awareness of peripheral events. The overall purpose of the experiment was to study the effect of two different displays on a user's performance in tasks that simultaneously require (1) the ability to visually and mentally concentrate on a central static virtual object, and (2) the ability to be visually aware of changes in his/her surroundings in the virtual environment.

We developed two hypotheses for these conditions: (1) hybrid display users are more visually aware of changes in their virtual surroundings, and (2) users of the hybrid display can visually and mentally concentrate better on the central static virtual object. Hypothesis 1 was supported by our experiments and hypothesis 2 was not supported by our results. We believe this lack of support for hypothesis 2 probably lies in an inappropriate choice of tasks for our experiment. Our central up-close task involved solving a puzzle, and people's ability to do this might vary significantly across the general population. In retrospect, we realize that a simpler central task, such as monitoring some changing digits for a “magic number,” would likely have less variance. Details of this experiment can be found in [23].

5. Creation and Use of the ivrBook

In our current work, the scenario for real-world use of the surgical ivrBook would be similar to the following detailed scenario in which a surgeon operates on a patient with blunt liver trauma:

A middle-aged man gets into a car accident on a snowy night. He isn’t wearing his seatbelt, and his abdomen collides with the steering wheel. He’s able to call for help on his cell phone immediately; paramedics arrive and take him to the local hospital with a trauma center. He’s in luck, as much as a car accident victim can be, in that a world-class trauma surgeon is on call that night. The victim is transported directly to the OR, where a plethora of cameras records the scene from many angles. The surgeon scrubs up and goes to work; within an hour, the patient has been stabilized.

Some time after surgery, the surgeon visits an Immersive Virtual Reality installation to annotate the 3D scene reconstruction into an ivrBook. A media specialist, who captures the surgeon’s commentary and performs other tasks in a media specialist application (see Figure 7), joins the surgeon. The media specialist uses the authoring software in collaboration with the surgeon, who must understand the concepts of ivrBook construction, but need not understand the software.

Figure 7: The Media Specialist Application. The panes are described in the text below. Note that for the sake of clarity we've used photographs of actual operations rather than reconstructed images.

A timeline projected on the back wall keeps the surgeon oriented in time. The past extends to the left, and the future, to the right. "Now" is represented by a green line in the center of the timeline (see Figure 7). This timeline becomes the backbone of the ivrBook. Vital signs such as heart rate and blood pressure can be represented during the procedure, providing a good indicator of the patient's overall stability.

The surgeon begins to navigate through time using a simple rotary input device. He can control the velocity of time motion, then stop time by tapping. In another mode, he can control his position in time directly, with a clockwise rotation moving time forward, and a counter-clockwise rotation moving time backward.

The typical first step in annotation would be to find the major events of the surgery, and mark them with regions of time we call "clips." The surgeon uses a tapping gesture to indicate the beginning and end of each clip, e.g., the region of time in which the surgeon made the abdominal incision.

The media expert and surgeon work together to identify each clip with a name and a description, but also with a snapshot, which is an iconic identifier for the clip (see Figure 8). Note that all the 2D images in this paper are stand-ins; ultimately they will be from the reconstructed geometry.

Figure 8: The New Clip Pane, showing an snapshot as iconic identifier, a textual description, and some simple metadata, to be inserted in the timeline.

This image appears in the 2D user interface and also in the timeline, adjacent to the clip. The surgeon asks the media specialist to record the title "Exploring the Abdomen" for the clip and adds a description such as "I'm reaching into the abdomen to identify other possible sources of bleeding." The surgeon and media expert repeat this process to identify many regions of time (see Figure 9) breaking the procedure down into a series of steps and providing commentary for each one. Each of these clips appears in both the 2D timeline (see Figure 10) and the IVR timeline.

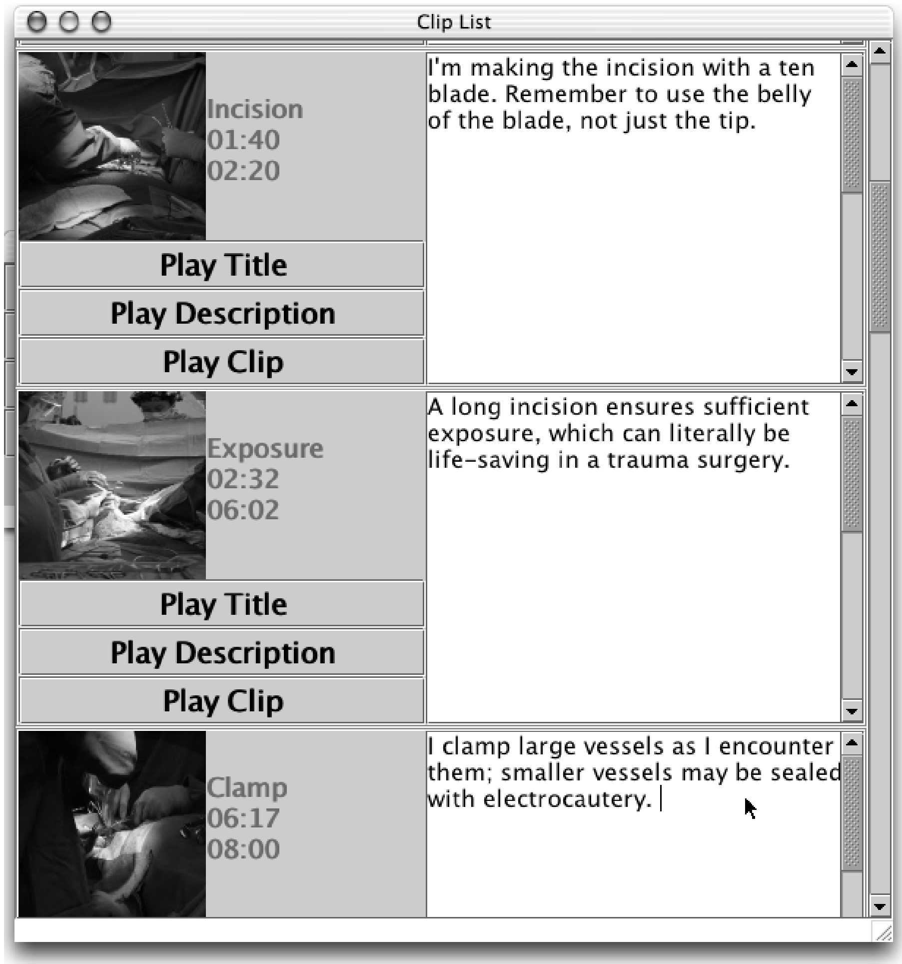

Figure 9: The Clip List Pane, showing a scrolling list of all of the "clips" identified by the surgeon and media specialist.

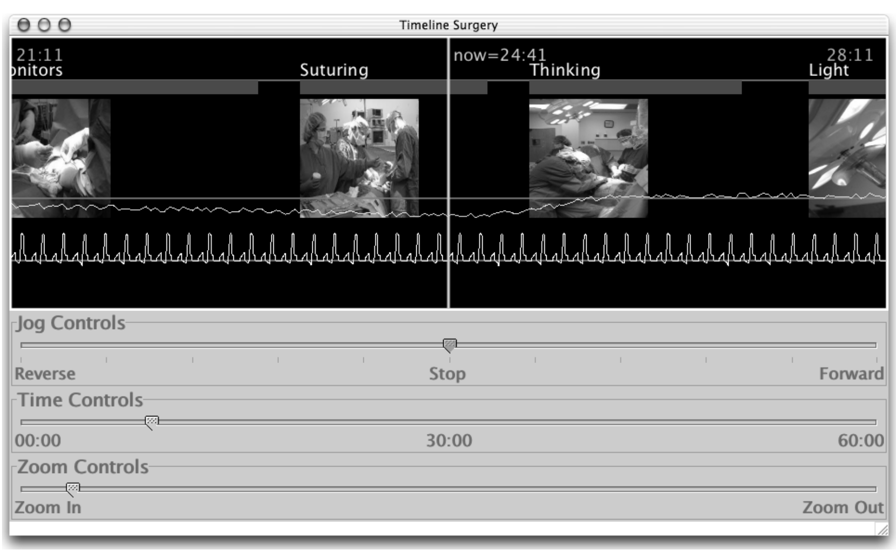

Figure 10: The Timeline Pane, showing various controls. The jog and time controls may be controlled either by the media specialist or by the IVR user. The timeline visible to the IVR user is always in sync with the timeline visible to the media specialist. Each clip is represented as a rectangle, extending over a region of time, plus an image from the surgery at that time.

In addition to observing the surgeon’s actions and hearing explanations of individual actions (in order to learn technique), surgical trainees must learn how the surgeon thinks and makes decisions—the clinical decision-making process—because after training they will have to make life-and-death decisions under tremendous time pressure. In effect, teaching surgery is more about the decision-making than it is about imparting of technical skills to the trainee. "It's 3/4 decision-making and 1/4 technical skills" according to Dr. Cairns, one of our project's trauma surgeons. By having essentially on-demand experiences that are not as quite as good as being there, and better in other ways, trainees will get far more experience on which to base their intuition than they could ever hope to accumulate through practice in formal training in an OR. Just like other forms of technological surgical learning tools, this does not replace but augments. Surgical intuition must be built up from experience, and the ivrBook provides such experience beyond what an individual can experience by being present for actual surgeries.

The ability for surgeons to add detailed commentary combined with the ivrBook controls that give a trainee the choice of what level of commentary to hear make the ivrBook a far more flexible pedagogical tool, and makes better use of the trainee’s time, than a narrated linear video (which can only be controlled with normal VCR controls). Today's entertainment DVDs provide multiple voiceovers with the main film, the ability to view individual scenes, and various types of supplemental footage explaining and annotating the film. While this sort of content goes beyond the purely linear, it hardly provides the kind of hypertextual branching narratives necessary to represent the complex thought structures of expert surgeons. Our notion of the ivrBook expands this type of medium—providing branching narrative with elision and spatial immersion.

To return to the authoring process, the final step in constructing a clinical decision-making guide is to associate parts of the text explanation with particular regions of time. The media specialist does this under the direction of the surgeon by dragging and dropping clips onto highlighted text (see Figure 11).

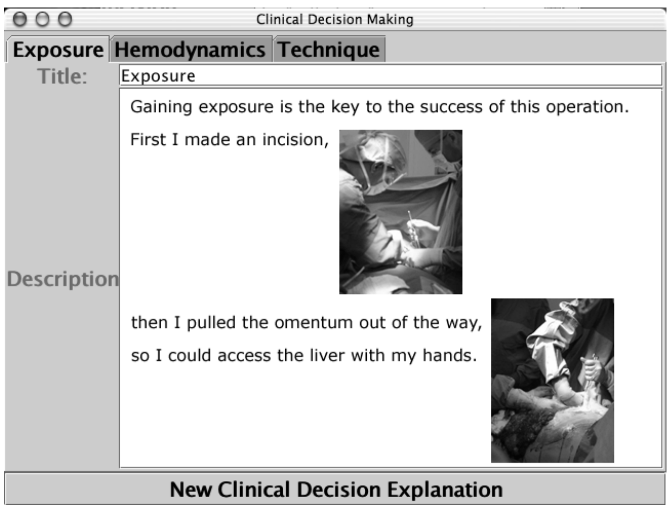

Figure 11: This interface allows the user to associate a thought process with individual clips, which are atomic elements of the surgical record.

Most operations are on the order of hours, so a compressed explanation of high points would be much easier for an overworked surgical trainee to appreciate than an entire operation. Using this application, the surgeon has ample time to teach without endangering his patient. He can also capture a number of simultaneous, distinct decision processes (i.e., the various actions which he must take in order to keep the patient alive and stable, in conjunction with, but separate from, the particular decisions about how to handle the liver injury itself).

Unlike someone learning from a video, an ivrBook trainee could experience not just clips from a single static position and angle, but any part of the surgery from any position and angle. For instance, he could observe the thoracotomy from the point of view of the surgeon, at the side of the table, or from above. He could see the angle of approach the surgeon makes with his scalpel. Also, the trainee would be able to place these highlight clips within the context of the entire operation, rather than see them only as isolated moments in time.

As Dr. Cairns says, “We’re not trying to replace in-person training; we’re not trying to replace surgical mannequins. We’re trying to create a new modality of learning that will *augment* all the existing ones.”

6. Influence of Practical Considerations on Interface Research

Unexpectedly, one of the most challenging problems in our research on interfaces for creating and using ivrBooks has been aligning our research visions with the practical requirements of the system. Our previous immersion work centered on collaborative design among remote participants, and interaction was chiefly embedded in the 3D spaces seen by each person. Each participant could interact with the object being designed, and manipulate it in 3-space. Protocols for object-locking and constraints on editing were also signaled in 3D. While this approach worked well for the design task, and while 3D interaction is one of the known strengths of our research group, we were forced to abandon many of our plans to use 3D interaction in the first phases of this surgical ivrBooks project.

For example, in the important area of annotation, one of our original goals was to have surgeons indicate some annotations during the real 3D surgery. We quickly found, however, that this had the potential to distract the surgeon and that if any non-surgical hand or head or eye movements were involved, these would actually reduce the training value of the recording/reconstruction because these gestures should, of course, not be learned or followed by the trainees. Annotation during surgery is also problematic for the same reasons that surgeons do not do much expository teaching during surgery; they cannot afford the distraction, at least not during critical sections of the surgery. Thus, teaching during surgeries tends to be catch-as-catch-can, not following a planned pedagogic design.

Delaying the surgeon's annotation until after the surgery reduced spontaneity, but still seemed acceptable. In fact, it follows the traditional rhythms of dictating case notes after the fact, presenting a case at rounds, or writing a case report up for a journal. We had at first envisioned creation of 3D text in the scene that trainees could view and reposition, and we planned to build 3D interaction tools that would enable the surgeon to place anatomical models or organs into the reconstructed scene, often rendered non-realistically, as an aid to pedagogical explanations and contrasts with other possible scenarios or outcomes. Our approach to relating patient anatomy to reference models was to begin with creating a library of standard anatomical features, perhaps derived from the Visible Human Project [24] or the Digital Anatomist Project [25], and to apply illustration techniques that made the patient anatomy look more like reference drawings. Annotation of a surgery would then consist in part of placing features from this library in the reconstructed data, and selectively painting obscured or confusing regions of the reconstructed data. We also envisioned the surgeons using 3D mark-up tools to create diagrams within the reconstructed surgery, drawing lines and marking both regions of an image and parts of the reconstructed patient's anatomy to indicate visible anatomical clues for the underlying structures.

When we began to work with surgeons and medical experts, we discovered that, although many of these ideas had merit as research projects, the likelihood of an overworked surgeon learning to use our custom 3D interaction methods was negligible. Although it was painful to put our innovative goals on a back burner, we readjusted our focus toward learning more about the actuality of hospital life and the schedules and demands of both surgeons and surgical residents. We came to realize that we needed a system that optimized the surgeon's time and required little to no learning curve.

We decided to take most of the annotation controls out of the 3D realm and to further offload tasks from the surgeon by having a media specialist work with the surgeon to input the annotations and link them to proper places and times in the reconstruction. We have developed the 2D media control application mentioned in the previous section, which links to the 3D reconstruction. This Java application communicates over TCP/IP sockets with a custom Brown-developed 3D toolkit, GLUE, via ASCII commands. We are currently using JDK 1.3.1 and a Macintosh-specific use of Apple Quicktime for audio.

The 2D media controller represents the immersive 3D experience in shorthand, using a traditional video-editing time line and screen captures from the 3D application. The media specialist application edits clips like a video editor but also deals with metadata (such as title, comments, and viewpoint), not just start and stop time. We are also editing clips that are spatial in nature, not just 2D video. The end result (the ivrBook) can be browsed non-linearly as well as in the programmed sequence of clips. The media controller thus combines aspects of non-linear video editing, hypermedia authoring, and 3D applications—combining features from all these areas to reach a coherent goal that, to the best of our knowledge, has not been pursued before.

Again in consultation with working surgeons, we decided that although we could offload many tasks to a media expert working a GUI interface tool, if we took all control from the surgeon we would introduce unnecessary frustration and communication difficulties. We are currently giving the surgeon control over the passage of time in the reconstruction playback as well as the ability to "click" or mark a point in time to say "now." We have also introduced some supporting imagery into the 3D space, although not the 3D non-photorealistically rendered models we had anticipated. Instead, what seemed most valuable for annotation and training purposes was a 2D timeline like the one in the 2D annotation controller application and a constant read-out of vital signs, such as heart rate and blood pressure. These are both shown in 2D on the back wall of the simulated OR. They are easy to refer to but do not interfere with the visual experience of the reconstructed surgery.

7. Future Directions

7.1 Geometry Acquisition & Reconstruction

When our system is ready to capture live surgery, we expect to face several challenges particular to the application area created by restrictions on the placement of cameras in the operating room (OR). It is important not to introduce either lighting or objects, such as cameras, that would interfere with the surgical proceedings. We are investigating camera placement in a working OR. A permanent, calibrated array of downward-facing cameras on the ceiling and inward-facing cameras on the walls will provide the lion’s share of the raw data; these must be augmented with cameras aimed directly at the interesting parts of the procedure, which will tend to be obscured by people and tools. One possible location for these augmenting cameras may be in the surgical lights, which will probably also provide sufficient lighting for good image capture when combined with ambient overhead lighting. Our observations suggest that the area under the surgeon’s hands is rarely in total shadow. If some light is reaching that location, chances are that we can provide reconstruction of views from angles that would be very difficult for a trainee to have in the real OR, including, potentially, even under the surgeon’s hands. Another (ideal) location for cameras is on the surgeon’s head. This seems plausible from both a technical, cultural, and ergonomic standpoint, as surgeons already wear lamps, magnifying lenses, and cameras on their heads.

We also plan to address the problems of acquisition of specular objects, as described in the "Capture of the Live Procedure" section earlier in this paper.

7.2 Authoring & Pedagogic Design

As we gain experience with basic surgical annotation and illustration techniques, we will start to apply storytelling concepts to reconstruction sequences. In particular, we envision the use of layered path storytelling in which the same sequence of events can be shown multiple times, but with different cinematic cues each time to highlight different material. For example, a scene of a patient being wheeled into the trauma room might be viewed without embellishment to highlight the activity and confusion of the real incident. Then the same scene might be replayed with special lighting cues that focus on critical activity. A third replay of the scene might involve changing the pacing of the replay to bring attention to rapid or subtle movements. Ultimately, the branching narrative concepts that emerge from studying trauma surgery may be applicable to a broad scope of situation-based training.

A kind of telepathy takes place between a surgeon and her surgical teammates, usually less experienced surgeons themselves. Many of the standard surgical techniques require two people to work as one – for instance, one surgeon may use blunt scissors to dissect and hold a membrane, while another surgeon uses an electrocautery to cut that membrane. The surgeons work with their heads bent close together, communicating in near-whispers unintelligible to observers a few feet away. We envision equipping surgeons, and other OR personnel, with microphones and recording a time-stamped stream from each of them. Other interesting conversations take place between surgeon and anesthesiologist, surgeon and medical student, and other permutations. At the same time, some conversations may be all noise, no signal – medical students discussing their weekend plans, for instance. We envision selecting particular conversations (time span and participants) as areas of interest. In this way, the learner will be able to listen to “signal” without much “noise.”

We also hope to capture the geometry of the annotation step; that is, place the entire IVR installation in which the surgeon is annotating/authoring, itself, within the view of the sea of cameras. This way we can see what the surgeon is pointing to when she says, perhaps, “right here, I should have seen this polyp” or “I would have been more comfortable if I had rested my wrist here while I suture.” We can also use one operation’s captured geometry of a whole OR to teach multiple roles. For instance, an exploration of the abdomen following a car accident could be used to teach experienced surgeons a new operative procedure; to teach inexperienced surgeons principles of planning trauma surgeries; to teach anesthesiologists how to maintain hemodynamic stability in hemorraging patients; to teach medical students anatomy and suturing techniques; and to teach OR technicians how to lay out and identify various tools while maintaining sterile technique. In fact, the interests of each of these audiences probably overlap; the technicians and anaesthesiologists will be interested in the surgeon's thought process, since it will offer information not available to them during real procedures.

As we continue to focus on real-life use of the system we will spend much more time than would be necessary for most research projects on ensuring robust operation of both the IVR environment and the media specialist application. These must be easily used by people with no interest in or time for computer problems.

Most ORs have some way to display medical imaging, including X-rays, MRIs, and CT scans, plus real-time medical imaging devices such as Doppler flow monitors and ultrasonography. Our virtual OR can incorporate this imaging in a more natural way than can a lightbox on the wall, and the virtual OR can even show the relationship between the medical imaging and the "real" captured geometry.

Further, we want to represent the patient's real vital signs, which will almost certainly have been recorded frequently from the moment rescue personnel first reached the patient. We also want to make available in the ivrBook the patient's entire pre- and more importantly post-operative history. If a wound becomes infected, it would be nice if the surgeon could find the point where the sterile field was contaminated. Or if a scar heals badly, or a tumor re-grows, it would be beneficial to be able to identify the causes of these bad outcomes.

7.3 Display technologies

Our plans for display work include tackling the difficult problem of combining the stereo image in our hybrid display's HMD with the mono image projected on the walls and floor. For example, currently we cannot give a sense of the scene directly below the HMD floating out in space, which creates a difficult transition area in the visual field. This may be solved with a physical "blinder" that keeps the trainee from looking down and being surprised by seeing an image of, say, the operating table looking as if it lies flat on the floor instead of being properly positioned in 3D space. For the hybrid display, we are also considering methods for introducing the trainee's hands, feet, and body into the scene. Although only future studies can confirm this, we expect that a sense of one's own presence in the space of the OR will be important to the learning experience.

8. Conclusion

We have made early progress towards our eventual goal of building a richly annotated immersive environment allowing full spatio-temporal navigation for training in procedures which are difficult to observe. Research in time-varying geometry acquisition at UNC is moving toward the goal of capturing an entire operating room for a trauma procedure. New types of displays will let surgeons and trainees make the most of their time spent with ivrBooks. Interface systems for creating and using ivrBooks are addressing the practical needs of the ivrBook audiences and should result in tools usable by those outside the computing field.

Over the next two years, we plan to build a system that is truly useful for surgical training, and with which we can conduct formal user studies. We will then improve the technology and user interfaces based on the outcomes of such studies. After perhaps another round of user testing, we will work towards installing a system in a real operating room, and running field trials of the use of ivrBooks in surgical resident education.

9. Acknowledgments

Many thanks to Dr. Bruce Cairns, Josh Carroll, Andy Forsberg, Joe LaViola, Mark Oribello, Tom Roache, Rosemary M. Simpson, Noah Wardrip-fruin, and Andrew J. Young for invaluable support on the project, this paper, and on the video which accompanies it.

References

[1] Neumann, Ulrich and Henry Fuchs, "A Vision of Telepresence for Medical Consultation and Other Applications," *Proceedings of the Sixth International Symposium on Robotics Research*, Hidden Valley, PA, Oct. 1-5, 1993, pp. 565-571.
[2]Fuchs, Henry, Gary Bishop, Kevin Arthur, Leonard McMillan, Ruzena Bajcsy, Sang Lee, Hany Farid, and Takeo Kanade, "Virtual Space Teleconferencing Using a Sea of Cameras," *Proceedings of the First International Symposium on Medical Robotics and Computer-Assisted Surgery*, vol. 2, pp. 161-167, September 22-24, 1994, Pittsburgh, PA.
[3] P. Narayanan, P. Rander, and T. Kanade. Constructing Virtual Worlds using Dense Stereo. In *Proceedings of International Conference on Computer Vision*, pp. 3–10, June 1998.
[4] Sundar Vedulay, Simon Baker, Peter Randeryz, Robert Collinsy, and Takeo Kanade. Three Dimensional Scene Flow. In *Proceedings of International Conference on Computer Vision*, pages 722–729, Sept 1999.
[5] Staadt Oliver, Gross Markus, Kunz Andreas, Meier Markus, Gross Markus, "The Blue-C: Integrating Real Humans into a Networked Immersive Environment", ACM Collaborative Virtual Environments 2000.
[6] W. Matusik, C. Buehler, R. Raskar, S. Gortler, and L. McMillan. Image-Based Visual Hulls. In *Proceedings of SIGGRAPH 2000*, pages 369–374, New Orleans, August 2000.
[7] Li-Qun Xu, Bang Jun Lei and Emile Hendriks: Computer Vision for a 3-D Visualisation and telepresence Collaborative Working Environment", BT Technology Journal, Vol 20 No 1, January 2002.
[8] E. Trucco, C. Plakas, N. Brandenburg, P. Kauff, M. Karl, O. Schreer, Real-Time Disparity Analysis for Immersive 3-D Teleconferencing by Hybrid Recursive Matching and Census Transform", 8th IEEE Int. Conf. On Computer Vision; ICCV 2001, Proc. of Workshop on Video Registration, Vancouver, Canada, July 2001
[9] K. N. Kutulakos and S. M. Seitz, “A Theory of Shape by Space Carving,” *Int. J. Computer Vision*, vol. 38, no. 3, pp. 197–216, 2000. David Marr Prize Special Issue.
[10] Agrawal, Motilal, and Larry S. Davis. A Probabilistic Framework for Surface Reconstruction from Multiple Images, in *Proceedings of Conference on Computer Vision and Pattern Recognition*, 2001.
[11] J.D. Mackinlay, G.G. Robertson and S.K. Card. "The Perspective Wall: Detail and Context Smoothly Integrated" in Proceedings of ACM CHI 91, p. 173-180, 1991.
[12] Catherine C. Marshall. "Toward an Ecology of Hypertext Annotation" in Proceedings of ACM Hypertext '98, pp. 40-49, 1998.
[13] Lynda Hardman, Dick C.A. Bulterman, and Guido van Rossum. "The Amsterdam Hypermedia Model: Adding Time and Context to the Dexter Model" in ACM CACM, 37(9), pp. 50-62, 1994.
[14] Nitin Sawhney, David Balcom, Ian Smith. "HyperCafe: Narrative and Aesthetic Properties of Hypervideo" in Proceedings of ACM Hypertext 96, pp. 1-10, 1996.
[15] Ramesh Raskar, Greg Welch, Matt Cutts, Adam Lake, Lev Stesin, Henry Fuchs. "The Office of the Future: a Unified Approach to Image-based Modeling and Spatially Immersive Displays" in Proceedings of ACM SIGGRAPH '98, pp.179 - 188, 1998.
[16] R. T. Collins. "A Space-Sweep Approach to True Multi-image Matching" in Proceedings of Conference on Computer Vision and Pattern Recognition, pages 358–363, June 1996.
[17] S. M. Seitz and C. R. Dyer. "Photorealistic Scene Reconstruction by Voxel Coloring" in Proceedings of CVPR 1998, pages 1067–1073, 1997.
[18] R. Szeliski and P. Golland. "Stereo Matching with Transparency and Matting" in Proceedings of International Conference on Computer Vision, pages 517– 524, Sept 1998.
[19] Ruigang Yang, Greg Welch, Gary Bishop (2002). “Real-Time Consensus-Based Scene Reconstruction using Commodity Graphics Hardware” in Proceedings of Pacific Graphics 2002. Tsinghua University, Beijing, China (October 9-11, 2002).
[20] B. Curless, M. Levoy. "A volumetric method for building complex models from range images" in Proceedings of *ACM Siggraph '96*, pp 303–312, 1996.
[21] Mark Pauly and Markus Gross. "Spectral Processing of Point-Sampled Geometry" in Proceedings of *ACM Siggraph '01*, pp 379—386, 2001.
[22] Low, Kok-Lim, Greg Welch, Anselmo Lastra, and Henry Fuchs (2001). “Life-Sized Projector-Based Dioramas" in Proceedings of the ACM Symposium on Virtual Reality Software and Technology. Banff Cantre, Banff, Alberta, Canada (November 15-17, 2001), ACM SIGGRAPH, Addison-Wesley.
[23] Kok-Lim Low, Adrian Ilie, Greg Welch, and Anselmo Lastra (2003). “Combining Head-Mounted and Projector-Based Displays for Surgical Training” to appear in Proceedings of IEEE Virtual Reality 2003. Los Angeles, California. (March 22-26, 2003).
[24] The Visible Human Project website. http://www.nlm.nih.gov/research/visible/visible_human.html
[25]Digital Anatomist Project website. http://sig.biostr.washington.edu/projects/da/index.html