A Survey of Surgical Simulation: Applications, Technology, and Education

Abstract

Surgical simulation for medical education is increasingly perceived as a valuable addition to traditional teaching methods. Simulators provide a structured learning experience, permitting practice without danger to patients, and simulators facilitate the teaching of rare or unusual cases. Simulators can also be used to provide an objective assessment of skills. This paper is a survey of current surgical simulator systems. The components of a simulator are described, current research directions are discussed, and key research questions are identified.

1 Background

Simulation is an integral part of surgical education. Surgeons develop and acquire skills through practice, and techniques are learned using animals, cadavers, volunteers, and patients. This approach has disadvantages. Animals have a different anatomy, cadavers cannot provide the appropriate physiological response, and there is a risk to patient safety while the caregiver is learning.

Recent advances in computer technology permit a new class of simulators to be developed. Computer-based surgical simulators use virtual patients. These simulators can generate realistic human anatomy and physiological responses (Kizakevich, McCartney, Nissman, Starko, & Smith, 1998) including certain types of pathology. For some medical encounters, students can practice on standardized cases and receive detailed feedback on their performance (Hubal, Kizakevich, Guinn, Merino, & West, 2000; Seymour et al., 2002). Patient safety is not compromised while the student is learning.

This paper is a survey of surgical simulation systems, techniques, and applications in medical education. The clinical motivation for surgical simulators is described in subsection 1.1. Subsection 1.2 is a broad survey of surgical simulators developed for research and education. Section 2 discusses research from different domains that are relevant to surgical simulation. Sections 3 and 4 outline the lessons learned with this new technology and summarize the discussion.

1.1 Clinical Motivation

Advances in medical technology and changes in the practice of modern medicine are forcing a reevaluation of the profession’s teaching methods. Surgical education uses the apprenticeship model, formalized as residency pro-
grams. In the United States, surgical residencies generally require five to seven years to complete (The Surgical Career Handbook, 2002; Johansen & Heimbach, 2003). Training in these programs can be characterized as training-by-opportunity. If a patient comes in with appendicitis, the resident learns the technical aspects of appendectomy. Ideally, over the course of a residency, the surgery resident participates in the care of enough patients with various diseases that the resident becomes competent in the academic and technical aspects of the chosen surgical specialty. However, there is no guarantee that this will be the case.

The problem is exacerbated by rising health care costs. Health insurance carriers are encouraging participants to choose outpatient over traditional inpatient surgery (Grant, 1992). Fewer patient contact hours are available for medical education. In addition, the economics of medicine has drastically reduced physician teaching time. Residents also face time constraints; regulations are now being adopted that limit both the number of hours that a resident can work per week, and the number of days that can be worked sequentially (Conyers, 2001; Corzine, 2002). With new procedures increasing and becoming more complex, it is clear that there are obstacles to effective medical education using the current model.

The certification of surgeons in recent years has consisted of a written exam and an oral exam. No formal board certification process currently includes a technical skills proficiency examination. Competency is judged subjectively, even though there is a consensus that objective measures are required.

Surgical simulators can potentially address many of these issues. Rather than training-by-opportunity, simulators can generate scenarios of graduated complexity. Simulators can be used to ensure that minimum standards are met prior to performing the next level of case complexity in the operating room, or progressing to the next level of training. New and complex procedures can be practiced safely on a simulator before proceeding to the patient. Simulators can decouple the physician-student time dependency. Students can practice on their own schedule. Practice sessions can be stored for later review by the physician.

1.2 Current Work

Surgical simulators have been developed for a wide range of procedures. They can be broadly classified in terms of their simulation complexity. Three classes are considered: needle-based, minimally invasive, and open surgery.

1.2.1 Simulators for Needle-Based Procedures. As the name implies, needle-based procedures use needles, catheters, guidewires, and other small-bore instruments. The Immersion Medical CathSim Vascular Access Simulator (Ursino, Tasto, Nguyen, Cunningham, & Merrill, 1999) was developed to train nursing students in the proper technique for starting an intravenous line. The system used a custom-developed haptic interface device to simulate the needle and catheter. The device can report three degrees of freedom (DOF) orientation data, and provide one-DOF haptic feedback along the direction of needle insertion. A desktop personal computer controlled the device, and real-time visual and haptic feedback was provided during simulation. Liu et al. used the same hardware configuration to develop needle-based trauma procedures such as pericardiocentesis and diagnostic peritoneal lavage (Liu, Kaufmann, & Tanaka, 2001; Liu, Kaufmann, & Ritchie, 2001).

Needle-based simulators generally have limited visual and haptic realism. They are useful for teaching relatively straightforward procedures with well-defined algorithms. Their simplicity makes them relatively inexpensive. They are suited for training widely performed procedures at low cost. Intravenous catheterization falls under this category. In situations where opportunities for practice are limited or where current methods using animal models are not optimal, needle-based simulators can also have a role. Pericardiocentesis and diagnostic peritoneal lavage fall in this category.

1.2.2 Simulators for Minimally Invasive Surgery. Minimally invasive procedures use specially designed instruments. The instruments are introduced into the body via small incisions. Visual feedback is obtained via inserted scopes, cameras, or fiberoptic devices,
and a video display monitor is used to show the image. Because the entry portal is small, these instruments have a limited range of motion. For example, laparoscopic instruments are constrained to pivot about the entry port on the abdominal wall. Other instruments are designed to work in confined areas. For example, only the tips of bronchoscopes can be flexed under user control. In both these cases, haptic feedback is muted due to scaling gaskets (such as in laparoscopic instruments) or the length of instrument within the patient’s body (as with bronchoscopes).

The limited range of motion and haptic feedback, use of specialized tools, and video displays facilitate simulator development. Simulated laparoscopy is an active focus of research. The LASSO project (Szekely, Brechbuhler, Dual, et al., 2000) is an integrated development effort to construct a laparoscopic simulation platform. The abdominal cavity was modeled using data from the Visible Human initiative. Organ surface features were generated using a combination of texture analysis/synthesis, procedural texturing, and L-systems-based methods for growing vascular networks. Real-time deformation, haptic, and rendering performance was achieved using a purpose-built, 64-node parallel processor. The Karlsruhe endoscopic surgery trainer (Kühnapfel, Çakmak, & Maass, 2000) is based on the KISMET environment for virtual surgery development. A simulated laparoscopic cholecystectomy procedure was developed on this system. The simulator used an SGI Octane/MXE workstation with two 250 MHz Mips R10000 CPUs to achieve a visual update rate of twenty frames/sec. A PC-based system recorded instrument positions and joint angles, communicating the information to the SGI via a serial interface. The system did not provide force feedback. A similar SGI platform was used to develop VESTA (Tendick et al., 2000), a laparoscopic simulator developed for understanding, training, and assessing surgical skills. Unlike the Karlsruhe simulator, VESTA provided force feedback using modified Phantom haptic interface devices (Massie & Salisbury, 1994). Commercial laparoscopy trainers include products from Surgical-Science (www.surgical-science.com) and Mentice (www.mentice.com). Both systems are PC based, using a non-force reflecting laparoscopic interface from Immersion Medical. In the Mentice system, trainees learned hand-eye coordination by manipulating spheres and other geometrical objects in an abstract environment, whereas the Surgical-Science simulator used a simplified rendition of the abdominal cavity as a practice environment. Students manipulated vessel-like structures and could cause bleeding due to careless handling. Other simulated minimally invasive procedures include endoscopy.

Bro-Nielsen et al. described a PC-based bronchoscopy simulator (Bro-Nielsen, Tasto, Cunningham, & Merrill, 1999). In addition to realistic visual effects, the system used a haptic interface designed to provide realistic force feedback during scope insertion. The system has since been expanded to include colonoscopy and flexible sigmoidoscopy.

Among minimally invasive surgery simulators, those for laparoscopy and endoscopy are the most advanced. The interior anatomy generally contains sufficient detail and realism for educational purposes. Commercially available laparoscopic trainers can teach basic skills such as camera navigation, grasping, suturing and knot tying, and cauterization. Laparoscopy simulators have been clinically validated to improve performance in the operating room (Seymour et al., 2002). Despite recent advances, limitations still exist. Surgical effects, such as bleeding, blood pooling, and tissue tearing, presently have limited realism. Real-time tissue and organ deformation are generally limited to specific organs or simple structures such as arteries, ducts, and other tubular structures. For these and other reasons, the goal of simulating medically relevant procedures from start to finish has not yet been reached.

### 1.2.3 Simulators for Open Surgery.

Open surgery requires larger incisions in the body. The surgeon often has direct visual and tactile contact with the region of interest. The visual field, range of haptic feedback, and freedom of motion are considerably larger compared to minimally invasive procedures. Open surgery is thus more difficult to simulate. Early work by O’Toole, Polayer, and Krummel (1999) included a simulator for vascular anastomosis. The system used an SGI Octane for visual rendering and a PC-based system generated haptic feedback using dual PHANToM hap-
tic devices. A dedicated 100 Mbps ethernet connection provided synchronization between the PC and the SGI. The system measured the operator’s performance, and could distinguish between broad skill levels. Bro-Nielsen, Helfrick, Glass, Zeng, and Connacher (1998) described a prototype abdominal trauma simulator with limited haptic feedback. The system was PC based. Haptic effects were rendered using a PHANToM. Bielser and Gross described the issues involved in simulating incisions and skin retractions (Bielser & Gross, 2002). Webster, Zimmerman, Mohler, Melkonian, and Haluck (2001) described a suturing simulator for incisions and wounds. The system used a dual-processor, PC-based system and a PHANToM desktop haptic device. Stereographic images provided a 3D rendering of the suture site.

Open surgery remains the holy grail of surgical simulation. Considerable advances in haptics, real-time deformation, organ and tissue modeling, and visual rendering must be made before open surgery can be simulated realistically.

2 The Components of a Surgical Simulator

A computer-based surgical simulation system draws on multiple disciplines. It has both technical and cognitive aspects. The technical components include a virtual-patient model and specialized input and output devices. The model must display physical properties consistent with a live patient. Thus, soft tissues should deform with contact pressure and should have the same texture and consistency as live tissue. The visual representation of the region of interest must be consistent with intraoperative views, and tissue perfusion and appearance must be consistent with actual scenes. Creating a virtual model of the relevant human anatomy draws on research from disparate fields, such as computer science and bioengineering.

A simulator engages the surgeon through multiple sensory channels. The surgeon’s actions are tracked and replicated in the virtual operating environment. Visual and haptic effects such as bleeding and tissue resistance are rendered on the appropriate hardware. These devices draw on research on visual and haptic displays.

An effective teaching simulator integrates its technical components with medical and educational content designed to impart specific skills or knowledge to the user. A simulator’s cognitive components draw on research on learning theory, performance measurement, and surgical knowledge.

The following sections discuss five main components of a surgical simulator: deformable models, collision detection, visual and haptic displays, tissue modeling and characterization, and performance and training.

2.1 Deformable Models

Tissue is elastic, and so the accurate modeling of human organs and tissues requires deformation to be considered. Methods of deformable modeling must consider the requirements of surgical simulators. The model must respond rapidly to interactive manipulation. It should closely approximate the behavior of tissues as they are being stretched or cut, and deformations should appear realistic when rendered. In situations where haptic feedback is used, the model should also provide a solution for computing reaction forces.

Deformable models can be broadly classified as being kinematically or physically based. Kinematic models do not consider the effects of the object’s mass, forces, or other physical properties during deformation. These models include splines, patches (Foley, van Dam, Feiner, & Hughes, 1990), and freeform deformations. The chain-mail algorithm (Gibson, 1997) describes an approach for propagating the displacement of volumetric elements to achieve deformation and this algorithm has been used for simulating arthroscopic knee surgery (Gibson, Samosky, & Mor, 1997). In general, kinematic models cannot incorporate physical properties easily, and they are not widely used in surgical simulation.

Physics-based models can incorporate material properties. Mass-spring and finite-element models are by far the most common methods used in surgical simulation. A mass-spring model consists of a network of point masses connected by spring dampers. A spring damper is represented by an idealized spring function and a
A mass-spring model can be fairly large and complex without sacrificing real-time response to user interaction. Bro-Nielsen et al. (1998) used a spring model for modeling the skin of the abdomen. In addition to deformation, the simplicity of a mass-spring construction permitted real-time cutting of the abdomen. The model was used for both visual and haptic rendering. The Karlsruhe endoscopic surgery trainer (Kühnapfel et al., 2000) extended the basic mass-spring model just described to include external and internal forces in addition to that exerted by spring dampers. Nedel and Thalmann (1998) employed a system of mass springs to simulate the deformation of muscles as they undergo contraction. The model incorporated angular springs to account for curvature and torsion constraints in muscles. Montgomery et al. (2002) described a surgical simulation software library that uses mass springs for modeling deformation. The library has been used to develop a number of different simulations (Montgomery, Bruyns, et al., 2001; Montgomery, Heinrichs, et al., 2001; Bruyns, Montgomery, & Wildermuth, 2001).

Mass-spring models have limitations. Realistic deformations require careful planning of spring-damper interconnections. Delingette (1998) described a condition that can lead to numerical instability. For a dynamic mass-spring model with $n$ nodes and total mass $m_{\text{total}}$, then $\kappa = m_{\text{total}}/n\pi^2(\Delta t)^2$, where $\Delta t$ is the simulation time step and $\kappa$ is the critical stiffness beyond which the system of equations is divergent. Thus, very hard objects such as bone require a small time step for numerical stability, and a small time step requires more computational power to run at interactive speed.

Not all tissue properties can be realistically modeled using mass springs. Tissues display nonlinear elasticity, and organs consist of inhomogeneous material. Specifying a mesh that will deform realistically to arbitrary surgical manipulation can be difficult. In addition, specifying realistic stiffness values is not easy. Radetzky, Rünger, Teistler, and Pretschner (1999) used a combination of fuzzy logic and neural networks for determining stiffness and other spring parameters. Fuzzy logic components permit a clinician’s domain-specific knowledge to be incorporated. Neural networks permit the model to be trained offline using slow but accurate models or from observations of actual organs.

### 2.1.1 Finite-Element Models

Finite-element (FE) models differ from mass-spring models in significant aspects. Mass-spring models represent regions of interest as point masses topologically connected by spring dampers. The initial formulation is discrete. Deforming the model changes the level of potential energy in the model. In contrast, FE models permit a continuous formulation relating deformation to energy. For example, Gibson and Mirtich (1997) described the formulation of a model based on strain energy, given as $\frac{1}{2} \int \sigma^T \epsilon \, dV$, where $\sigma$ and $\epsilon$ are material stress and strain, respectively. In practical applications, closed-form solutions for most formulations do not exist. FE models achieve a discrete approximation by dividing the region of interest into volumetric elements. For surgical simulation applications, a tetrahedral element is most common. Each volumetric element connects only to neighboring elements via shared nodes. The deformation (and thus energy) for points within an element is interpolated from the elements’ nodes. For example, Bro-Nielsen and Cotin (1996) described a method for computing displacements within a tetrahedral element by linear interpolation.

Discretizing the volume of interest produces a system of equations that can be expressed as $M\ddot{X} + CX + KX = F$, where $X$ is a composite vector of node displacements; $M$, $C$, and $K$ are the mass, damping, and...
stiffness matrices, respectively; and \( F \) is a composite vector of external forces.

FE models compute deformation over the entire volume instead of at discrete points. FE methods permit tissue properties to be more accurately modeled. Wu, Downes, Goktekin, and Tendick (2001) described a nonlinear FE model that is more representative of actual tissue behavior, where stress and strain properties are nonlinear. Szekely et al. (2000) modeled tissue properties as a set of four partial differential equations,

\[
\begin{align*}
\nabla \cdot (\mathbf{f}) &= \rho \mathbf{u} \quad \text{(momentum)}, \\
\n\nabla \cdot (\rho \mathbf{u}) + \rho &= 0 \quad \text{(continuity),} \\
\sigma &= f_1(\varepsilon) \quad \text{(constitutive law),} \\
\varepsilon &= f_2(\mathbf{u}) \quad \text{(strain formulation),}
\end{align*}
\]

where \( \mathbf{f} \) is the vector of volumetric forces, and \( \mathbf{u} \) is the material displacement. Picinbono et al. described a nonlinear, anisotropic formulation, permitting large tissue deformations to be modeled more accurately (Picinbono, Delingette, & Ayache, 2001). For similar accuracy reasons, FE models of the heart (Le Grice, Hunter, & Smaill, 1997; Nielsen, Le Grice, Smaill, & Hunter, 1991) have been used as part of a comprehensive analytic cardiac simulation (Smith et al., 2002).

The primary disadvantage of FE models is computational complexity. For example, the LASSO laparoscopy simulator (Szekely, Brechbuhler, Dual, et al., 2000) used a nonlinear FE formulation for organ modeling. To achieve a realistic update speed, custom-built hardware with a high degree of parallelism was required.

Other methods of improving computation speed have been developed. Bro-Nielsen and Cotin (1996) described a condensation method that computed deformation for surface elements without having to compute inner elements. Cotin, Delingette, and Ayache (1999) described a method of exhaustively precomputing “elementary” displacements for each nonfixed node. During run time, linear combinations of precomputed displacements were used to approximate actual deformations. Speed increases of several orders of magnitude were reported. Wu et al. (2001) described an adaptive method for reducing the number of elements. A hierarchy of progressively coarser elements was built from the original set. Regions of the model undergoing little or no deformation were computed using coarse elements, and regions with significant deformation were computed with finer elements. The author described several subdivision criteria. They include stress concentration, stress gradient, and degree of displacement.

A shortcoming of various speed-increasing methods is their inability to handle changes in tissue properties and topology, such as those due to cauterization and cutting. Cotin et al. (2000) describes a hybrid method using fast, precomputed deformations for regions where no cutting was expected, and a slower method for regions that will be cut.

### 2.2 Collision Detection

The interaction between surgical tools, tissues, and organs requires contact loci to be determined. Collision detection must be performed efficiently at interactive frame rates. This section surveys collision detection algorithms relevant to surgical simulation.

Most techniques adopt a two-level approach: first, a computationally inexpensive method bounds regions of intersection, then slower methods determine the exact collision loci. Bounding methods are discussed in subsection 2.2.1. Subsection 2.2.2 discusses exact computation methods.

#### 2.2.1 Collision Bounding

Spheres are the simplest bounding method. Objects that can potentially collide are bounded by spheres. Unless bounding spheres intersect, collision is not possible between the objects. Spheres are rotationally invariant, and intersections can be efficiently computed. They can be easily updated to account for object deformation. Spheres are most suitable for objects that are approximately spherical, but they are inefficient for long, thin objects such as surgical instruments.

Besides spheres, bounding boxes have been used. Axis-aligned bounding boxes (AABBs) (Cohen, Lin, Manocha, & Ponamgi, 1995) are bounding boxes whose sides are parallel to the coordinate axes. Intersection of AABBs occur if their projections onto all coordinate axes overlap. The size of the AABB varies with the object’s orientation. Long thin objects can have disproportionally large AABBs.

In contrast, an oriented bounding box (OBB) is aligned with the Eigen vectors of the object. OBBs are
independent of the object’s orientation. Using the separating axis theorem (Gottschalk, 1996; Gottschalk, Lin, & Manocha, 1996), intersection between two OBBs can be determined in at most fifteen arithmetic operations.

Static partitioning is a time-efficient bounding method. The volume of interest is partitioned into static subvolumes, which are typically boxes. A collision occurs only if more than one object occupies the same box. Both VESTA (Tendick et al., 2000) and Cotin’s work (Cotin et al., 2000) are examples of simulators using this method. This method is very efficient, executing in expected constant time and is easily capable of handling haptic update rates. It handles moving and deformable objects efficiently. Collisions can be checked at the same time. Its primary disadvantage is the large amount of memory required.

A single bounding level can lead to excessive false positives. Hierarchies of bounding tests have been implemented. Sphere trees (Hubbard, 1993, 1996) combine a hierarchy of bounding spheres with the simplicity of testing spheres for intersection. Different methods of generating the hierarchy were proposed, with a medial-axis approach (Blum, 1967) generating the most efficient hierarchy. Hierarchies of AABB trees (van den Bergen, 1997) and OBB trees (Gottschalk et al., 1996) have also been described. The LASSO laparoscopy simulator (Szekely, Brechbuhler, Dual, et al., 2000) uses a bounding box hierarchy for real-time collision detection.

### 2.2.2 Collision Refinement

When bounding volumes intersect, it is necessary to determine whether intersection actually occurred, and the loci of intersection. For polyhedral models, efficient methods include those by Lin and Canny (1991), Gilbert, Johnson, and Keerthi (1988), and Cameron (1997).

An ingenious hardware-based collision detection method has been developed for laparoscopic procedures (Lombardo, Cani, & Neyret, 1999). The authors note that for time steps \( t \) and \( t + 1 \), the OpenGL (Shreiner, 1999) viewing frustum can closely approximate the volume swept out by a moving laparoscopic instrument. If this frustum is rendered, any polygons visible must have collided with the instrument during the time interval. The authors implemented the algorithm using standard OpenGL function calls, thereby taking advantage of hardware used to accelerate graphics performance. Compared to software-based algorithms, the hardware-based method was up to 150 times faster.

Work on collision detection comes mainly from research in computer graphics, and issues specific to surgical simulation remain open. Surgical simulation often requires both haptic and visual rendering. Realistic haptics require update rates that are typically two orders of magnitude higher than that for visual rendering. Collision detection algorithms must be correspondingly faster. Detecting self-collisions, such as suture knots, can be challenging. Collisions in surgical simulation are not restricted to points, but can be large, two-dimensional regions. Open surgery simulation is one such example. The surgeon’s hands can be in contact with large areas of multiple organs. An efficient algorithm is necessary to determine the collision surfaces. Tissues are pliant, and organs can deform significantly when handled. Many collision bounding algorithms require precomputed data structures to work efficiently. Deformable models may require the data structures to be updated, resulting in suboptimal constructs. Performance is subsequently degraded (van den Bergen, 1997).

### 2.3 Visual and Haptic Displays

Visual and haptic displays are the means by which surgeons interact with simulators. Haptics refers to manual interactions with environments and is concerned with being able to touch, feel, and manipulate objects in the environment (Srinivasan & Basdogan, 1997). Tactile feedback is sensed by receptors close to the skin, especially in the fingertips. The term haptics includes kinesthetic feedback as well, which arises from position and force receptors in the muscles, tendons, and joints. Human tactile sensing is based on specialized receptors in the fingertips. This includes thermoreceptors that respond to a change in skin temperature, nocioreceptors that convey the sensation of pain, and mechanoreceptors that respond to mechanical action such as force,
vibration, and slip (Burdea, 1996). The hairless skin covering the palm and fingertips have five major types of receptors: free receptors, Meissner corpuscles, Merkel’s disks, Pacinian corpuscles, and Ruffini corpuscles. As an example, the Merkel’s disks form 25% of the receptors in the hand. They have a disk-like nerve ending with a receptive field diameter of 3–4 mm, a frequency range of 2–32 Hz (Johansson, Landstrom, & Lundstrom, 1982), a spatial resolution gap detection of 0.87 mm and a grating detection of 1 mm (Johnson & Phillips, 1981). The perceptual capabilities of the human tactile system needed for teletaction systems consisting of a tactile sensor, a tactile filter, and a tactile display are quantified by Moy, Singh, Tan, and Fearing, (2000). The results show that 10% amplitude resolution is sufficient for a teletaction system with a 2 mm elastic layer and 2 mm tactor spacing.

2.3.1 Tactile Feedback. Tactile feedback is useful for presenting information about texture, local compliance, and local shape. In many medical applications the physician’s sense of touch is critical, such as during palpation of the skin to check for suspicious masses.

Burdea (1996) divided tactile feedback interfaces into surface texture and geometry feedback, surface slip, and temperature feedback. Most researchers have focused on devices for surface texture and geometry feedback, which includes micro-pin actuators. The use of tactile interfaces for conveying task-related vibrations in teleoperation and virtual environments is described by Kuntarinis and Howe (1995). The paper also describes the kinds of tasks in which high-frequency vibratory feedback is important and gives design guidelines for the implementation of these interfaces.

An eight-by-eight element tactile capacitive array sensor for detection of submillimeter features and objects has been constructed (Gray & Fearing, 1996). The entire sensor array is smaller than the normal human spatial resolution of 1 mm. Each square element is less than 100 μm on a side with similar spacing between elements. The array is small enough to be placed on a catheter or endoscopic to provide tactile feedback during surgical procedures.

Despite its potential for enhancing realism, tactile feedback is not widely used in current surgical simulations due to the lack of good hardware for this purpose.

2.3.2 Force Feedback. Force feedback has a longer history than tactile feedback, with its origins in the nuclear industry for the handling of radioactive material. Several commercial devices usable for surgical simulation have become available. One of the most commonly used devices is the PHANToM from Sensable Technology, which is now available in several models. PHANToM is an acronym for Personal Haptic Interface Mechanism and evolved from haptic research at the MIT Artificial Intelligence Laboratory (Massie & Salisbury, 1994). Another popular force feedback device is the Laparoscopic Impulse Engine from Immersion Corporation, which also enables surgical tools to be tracked and manipulated in three-dimensional space. The device interfaces with a computer via a PCI card and the development kit supports Windows-based and Silicon Graphics computers.

A key issue in integrating force feedback devices into a surgical simulation system is the update rate required for high fidelity. Although the visual display can be updated at 30 Hz, the haptic interface update rate should be around 1,000 Hz for stability reasons and to obtain a responsive interface. This can be accomplished by a multirate simulation with a high-bandwidth force feedback loop as described by Cavusoglu and Tendick (2000).

Despite ingenious adaptations, haptic devices such as the PHANToM have limited utility for surgical simulation. Such devices were originally designed to render reaction forces about a point. Many surgical procedures require more-complex haptic feedback. For example, surgeons experience haptic feedback as objects are clamped or when staples are inserted during a laparoscopic procedure. Thus, current research has focused on the development of specialized devices for simulating specific classes of procedures, such as endoscopy (Tasto, Verstreken, Brown, & Bauer, 2000).

2.3.3 Visual Displays. A key feature of any surgical simulation system is the visual display. Visual dis-
plays include head-mounted displays, stereoscopic monitors, environmental displays, and retinal displays.

The use of virtual environments to enhance rather than replace real environments is referred to as augmented reality. To obtain an enhanced view of the real environment, head-mounted displays (HMDs) can be used. HMDs for augmented reality can be divided into two categories: optical and video (Rolland & Fuchs, 2000). With optical see-through displays, the real world is seen through half-transparent mirrors placed in front of the user’s eyes. With video see-through displays, the real-world view is captured with two video cameras mounted on the head gear, and the computer-generated images are overlaid with this view.

Essential to augmented reality systems are tracking devices to determine the position and orientation (pose) of the user’s head. The computer can then update the displayed image to reflect the current head pose. Augmented reality requires accurate trackers for good correspondence between real and virtual objects.

Tracking systems can be classified as inertial, acoustic, magnetic, optical, or mechanical. A survey can be found by Meyer, Applewhite, and Biocca (1992). For surgical simulation and training, optical and magnetic trackers are commonly used. A portable augmented reality surgical training tool was created using a see-through Sony Glasstron HMD, an Ascension Technologies electromagnetic tracker, and a PC-based computer (Montgomery, Thonier, Stephanides, & Schendel, 2001).

Fuchs and colleagues have been developing augmented reality systems for minimally invasive procedures (Fuchs et al., 1996). They have combined an HMD, a tracking device, and high-end computer graphics to develop a prototype system to aid in ultrasound-guided breast biopsy.

Despite advances, limitations remain in the use of HMDs for augmented reality. Tracking errors and latencies can result in misregistered and unsynchronized images. Many HMDs have a limited field of view, and the resolution of HMD displays may be insufficient for detailed work such as neurosurgery simulation.

Stereoscopic displays are designed to give the user a perception of depth. The most common approach is to generate a stereo-pair image, in which each eye is presented with a slightly different image of the scene. The images are similar to that perceived by each eye when viewing an actual 3D scene. When viewed together, the brain fuses the image pair, producing an illusion of depth.

One popular product is CrystalEyes from StereoGraphics Corporation which is a lightweight, liquid crystal shutter eyeglass. Alternating left- and right-eye images are displayed rapidly on a CRT display monitor. The display is synchronized with the eyeglass, which presents the alternating images to the appropriate eye. Due to the speed of the sequence, users perceive a continuous stereo-pair image. A recent innovation in this area is the development of autostereoscopic displays, which can generate a 3D image without special glasses. The display consists of an LCD screen with an integrated optical overlay. When the user views the display at the correct distance, the optical overlay directs light such that alternating LCD elements project onto each eye. When stereo-pair images are displayed, viewers perceive a distinct image in each eye.

Although stereoscopic display technology is improving, additional work remains. In addition to image disparity, human visual perception relies on cues such as accommodation and convergence for depth information. Stereoscopic displays generally do not provide these. As a result, conflicting cues can cause some viewers to fail to see stereoscopic images in such displays.

The Virtual Workbench (Poston & Serra, 1996) is an approach for unifying visual and haptic display. It has been used for surgical simulation and pre-surgery planning. In the Virtual Workbench, a PHANToM haptic interface is integrated with a stereoscopic display. A mirror arrangement creates a virtual image that is registered with the working volume of the haptic device. The surgeon’s natural hand-eye coordination is preserved. A volume-based preoperative planning system was developed for neurosurgical tumor applications and clinically evaluated in sixteen cases (Serra et al., 1998).

A novel visual display has been developed that obviates the need for an image screen (Johnston & Willey, 1995). Instead, coherent light from low-powered lasers is scanned directly on the retina. When used in a see-through HMD, this approach permits the generation of high-contrast images that are visible outdoors in bright daylight.
2.4 Tissue Modeling and Characterization

Soft tissue has extremely complex mechanical behavior. The stress-strain relationship is highly nonlinear. Unlike typical engineering materials like metals, very large deformations are possible, such as 40% for skin. Tissue behavior is viscoelastic (that is, the stress-strain relationship depends on the rate of deformation), inhomogeneous (varies through the tissue volume), and anisotropic (varies with direction). Properties also vary substantially depending on the species, age, and sex of the source, as well as the in vivo state of the tissue, such as loading and muscle activation.

The properties also change after removal from the body due to factors such as desiccation, blood loss, and a different stress state. Most sources of tissue mechanical properties use ex vivo samples from animals or human cadavers (Abe, Hayashi & Sato, 1996; Duck, 1990; Yamada, 1970). The advantage of ex vivo testing is that precise control of testing factors, especially sample geometry, is possible. In vivo testing, although less accurate, measures properties of tissue in its natural state.

Research groups have obtained in vivo data for simulation from a variety of animal and human tissues in tension (Brouwer et al., 2001) and compression (Brown, Rosen, Moreyra, Sinanan, & Hannaford, 2001; Carter, Frank, Davies, McLean, & Cuschieri, 2001; Ottensmeyer, 2001; Vuskovic, Kauer, Szekely, & Reidy, 2000). Sensors can also be mounted on instruments to measure interaction forces during typical tasks such as cutting or grasping (Brouwer et al., 2001; Rosen, Hannaford, MacFarlane, & Sinanan, 1999).

Tissue becomes stiffer as it stretches, and soft tissues often produce an exponential stress-strain or load-displacement curve (Figure 1). Due to viscoelastic effects, as tissue is stretched repeatedly less load is typically necessary to produce the same elongation. Data obtained from ex vivo tissues is typically “conditioned” by repeated cycling until a consistent curve is produced. However, tissue in vivo is likely in a different state. This is another motivation for measuring tissue properties in vivo. Methods for extracting parameters from exponential stress-strain curves and viscoelastic time constants from stress-time data can be found in Fung (1993).

A key question in modeling tissue behavior for simulation is the level of accuracy required. Human ability to detect differences in compliance between objects is very poor, with a just-noticeable difference of about 17%. However, we are very sensitive to changes in compliance or force that occur at higher temporal frequencies (Dhruv & Tendick, 2000). Research is necessary to determine what visual and haptic cues surgeons use, and how sensitive they are, to detect changes in tissue consistency due to damage, hidden structures, or disease.

Another important research question is how tissue damage caused by excessive force imparted by surgical instruments can be predicted. Well before tissue yield, damage to vessels can occur. This damage can lead to loss of function or tissue death (Morimoto et al., 1997). A simulation should be able to determine when the user is using too much force for tissue health, but there is little information on damaging loads. Although force information can be obtained from sensors mounted on...
instruments, it is also necessary to assess the damage caused by the applied forces. Unfortunately, this data is difficult to obtain because experiments require keeping animals alive for 24 hours or more to determine the course of injury. This demands ethical experimental protocols that minimize the possibility of pain in the animals.

2.5 Performance and Training

The goal of simulation is training the skills and teaching the knowledge necessary to successfully perform a procedure. An important distinction is that of ability versus skill. Ability is relatively stable capability or aptitude “that underlies (or supports) performance in a number of tasks or activities” (Schmidt & Lee, 1998); skill, however, is learned or trained, and may depend on a range of underlying abilities (Patrick, 1992). Abilities and components of skill can be cognitive as well as physical or perceptual-motor (Fleishman & Quaintance, 1984; Gagne, 1977).

The skills that are necessary for the performance of a procedure, and the abilities that underlie them, can be determined using task analysis. There are many varieties of task analysis (Patrick, 1992), but two examples are hierarchical decomposition and the critical incident technique. In hierarchical decomposition, the procedure is broken up into component steps, tasks, subtasks, and motions that can be analyzed (Cao et al., 1999). The goal of the critical incident technique is to identify key events in a procedure that can lead to success or failure. In a somewhat similar approach, a number of researchers have attempted to characterize events and behaviors that lead to errors in complex domains such as aviation or nuclear power plants. Common patterns emerge among errors at the levels of skill-based, rule-based, and knowledge-based performance (Reason, 1990). These patterns can also be observed in errors in surgery, such as bile duct injuries in laparoscopic cholecystectomy (gallbladder removal) (Gantert et al., 1998). For example, frequency gambling—in which one is biased toward an action that is appropriate in the most common situation but inappropriate in others—may lead to misidentification of variations in biliary anatomy.

Establishing and implementing metrics of performance is an important aspect of simulation. The form of metrics can vary with the knowledge or skill to be assessed. Declarative knowledge is explicit knowledge of facts, such as anatomic landmarks during a procedure or physiological effects of surgery. This knowledge can be assessed easily via a quiz or recognition tasks. Procedural knowledge is explicit knowledge of how to perform a procedure, such as the sequence of navigation of landmarks or the rules of proper use of an instrument. It can be expressed verbally, although it may depend on nonverbal (such as visual or haptic) information. Traditionally it is tested verbally, but it could be assessed instead in simulation by testing the user’s proper performance of the intended procedure.

Of course, much of surgical skill is nonverbal, relying on perceptual-motor abilities or nonverbal reasoning such as 3D spatial visualization (Eyal & Tendick, 2001). Examples of purely perceptual skills include visual recognition of a landmark in a cluttered environment or tactile recognition of tissue condition. A key concern in training and evaluating such skills in simulation is the level of realism necessary for proper training; much research is needed in this area. One advantage of training perceptual skills in simulation is that the information available in the real environment can be augmented in simulation, for example with a bird’s-eye view that is unavailable in endoscopy.

Skills involving both perception and motion in surgery can range from basic motor skills such as coordinated two-handed manipulations of tissue, to complex motor skills such as suturing, to skills that require higher-level cognition such as spatial reasoning or diagnostic reasoning. The first step in assessing or training these skills in simulation is recognizing the user’s actions using methods such as hidden Markov models (Rosen et al., 1999). To train motor skills, it is possible to provide visual or haptic guidance (Feygin, Keehner, & Tendick 2002). It is best, however, to use guidance primarily at the early stages of learning so that the student does not become overreliant on it (Schmidt & Lee, 1998).

Metrics for perceptual-motor skills include time, accuracy, or task-based criteria. Time is often not the best metric for surgery, because the fastest surgeon is not necessarily the best. Measures of accuracy can include
position, trajectory, force, or unintended contacts. Task-based criteria are successful results and avoidance of undesired events for an individual procedure.

An essential aspect of simulation is validation, or verifying the effectiveness of assessment or training in the simulation. There are many types of validation (Dick & Hagerty, 1971; Reznick, 1993). Content validity is the appropriateness of measures tested in the simulation to the task to be trained, such as determined by a task analysis. Construct validity is the extent to which an intended trait is measured. For example, senior residents should perform better than their junior colleagues on a test of surgical skill. Concurrent validity is the correlation of performance in the simulation with the real environment. Predictive validity measures this correlation as well, but as a prediction of future performance in the real environment. Consequently, this is relevant to preassessment, as for example the prediction of a medical student’s future performance as a surgeon. The most important type of validity, however, is the measure of training transfer, or the degree to which training in the simulation leads to improved performance in the real environment. So far, relatively few studies have shown transfer, although a number of experiments are underway.

3 Discussion

Computer-based surgical simulation is relatively new. Its multidisciplinary nature has drawn interest from many areas. Research is progressing vigorously at many institutions. Surgical simulators hold great potential for improving medical education. Preliminary clinical studies have shown that simulators can be used successfully in some medical courses. Because difficulty levels are controllable, a uniform learning experience is possible. Simulators also permit rare or unusual cases to be practiced routinely. Although initially expensive to acquire, simulators can be cost effective in the long term. Simulators do not require special storage, feeding, or maintenance facilities. They also do not require special lab facilities. In addition to skills training, simulators have potential applications in certification and recertification, selection for training, and selection for completion of training.

Despite the potential of such simulators, acceptance by the medical community has been slow. Physicians are familiar and comfortable with the current teaching model. They also remain largely unaware of simulation’s potential. One reason has been the lack of clinical studies. There is little information comparing the training efficacy of simulators with current teaching models. More validation studies are needed to increase adoption of simulation technology by the medical community.

4 Conclusion

This paper presented a survey of surgical simulation research. The authors described the interdisciplinary nature of the field and attempted to highlight current research across a broad cross section. Although considerable advances have been made, many unanswered questions remain. Key issues include the understanding and realistic modeling of tissue and organ properties, visual and haptic realism, realistic interactions (such as collisions) between physical objects, and understanding and measuring the learning process in surgical skills acquisition. This paper also discussed issues regarding the acceptance of simulation technology by the medical community.

Acknowledgments

Dr. Tendick was supported in part by the National Science Foundation under grants CDA-9726362 and BCS-9980122. The research work of Dr. Cleary was funded by U.S. Army grant DAMD17-99-1-9022. The content of this manuscript does not necessarily reflect the position or policy of the U.S. government.

References


Fuchs, H., State, A., Livingston, M. A., Garrett, W. F., Hi-


Johansen, K., & Heimbach, D. M. *So you want to be a Surgeon . . . A medical student's guide to surgery residencies*. American College of Surgeons Online Publications.


Szekely, G., Brechbuhler, C., Dual, J., Enzler, R., Hug, J,


