Mixed-Reality Simulation of Minimally Invasive Surgeries

Gerard Lacey
Trinity College Dublin

Donncha Ryan, Derek Cassidy, and Derek Young
Haptica

Our mixed-reality platform helps train surgeons in minimally invasive surgery and objectively assesses their performance. The platform uses multicamera stereo inside a patient manikin to measure the 3D positions of unmodified surgical instruments. It uses this information to drive a mixed-reality, computer-mediated learning system and provide objective measures of surgical skill.

Traditional surgical education is based on the apprentice model, where students learn within the hospital environment. In this setting, learning opportunities are secondary to the patients' clinical needs. Recent public and academic scrutiny,[1,2] however, has focused attention on the risk to patients while surgeons climb the learning curve as they acquire new skills.

For many years, researchers have identified the airline industry's simulation training and assessment model as the route to greater patient safety and proficiency-based progression in surgery.[3] Recent results provide evidence for the superiority of simulated training and objective assessment for laparoscopic cholecystectomy[4] and stenting of the Carotid artery.[5] (See the "Laparoscopic Surgery" sidebar for details on this procedure.) Simulation in medical training isn't confined to virtual reality simulation, however; Roger Kneebone[6] broadly defines *simulation* as the spectrum of teaching aids from inanimate models to VR equipment and argues that we must integrate simulation into the learning framework for it to be effective.

In this article, we describe the design, development, and evaluation of the ProMIS simulator, a mixed–reality platform for training surgeons in minimally invasive surgery, which we developed in close partnership with surgeons in leading institutes around the world. We created a human-patient simulator in which cameras track real surgical instruments' 3D positions within the simulation. Furthermore, we extended this basic architecture to facilitate the simulation of hand-assisted procedures.

This article also describes the development of objective and independently verified measures of surgical skill—namely, path length and economy of movement. Such assessment tools are one of the key components in raising a surgeon's proficiency prior to giving them operating privileges on patients.

Prior to its use in surgical training, the ProMIS system underwent multiple independent validation studies to test the system's ability to teach and measure surgical skills to the standards required by the Society of American Gastrointestinal and Endoscopic Surgeons (Sages). The system is now in use in more than 30 leading surgical training centers in North America, Europe, Asia, and Australia. It's also used by the Johnson and Johnson company Ethicon Endo-Surgery for training its surgeons and staff in hand-assisted laparoscopic colectomy.

State of the art in laparoscopic surgical training

The most basic device for training in laparoscopic skills is a *box trainer*, where a neoprene lid models the abdomen and students manipulate inanimate objects inside the box. Box trainers let students become familiar with concepts such as operating the laparoscope, the fulcrum effect of the abdominal wall, using real surgical instruments, and the difficulties of physical manipulation. Structured training and assessment programs based on the box trainer where human observers assess skills, such as the McGill Inanimate System for Training and Evaluation of Laparoscopic Skills (MISTELS)[7] and the Sages' Foundations in Laparoscopic Surgery (FLS),[8] have demonstrated improved surgical proficiency. However, these assessment tools require a significant observer workload and vigilance to ensure that the objective standards of assessment are maintained.

A common alternative to the human assessment of surgical skills is to use VR and mixed-reality systems that incorporate automated

1070-986X/07/$25.00 © 2007 IEEE

Published by the IEEE Computer Society

Authorized licensed use limited to: Maynooth University Library. Downloaded on January 30,2023 at 16:27:29 UTC from IEEE Xplore. Restrictions apply.

assessments. (Fiona Carter and her colleagues[9] provided a detailed review of the different simulators on the market in 2004.) VR systems typically use custom interface devices fitted with the handles of surgical instruments. Although nonhaptic systems have significant clinical validation,[4] there's some confounding evidence to suggest that haptics might be important at some points in the learning cycle.[10]

The advantages of the VR systems are their automated performance assessment as well as their ability to vary the simulated patients' anatomies and simulate random events that occur in surgery, such as bleeding and the fogging of the laparoscope. One challenge to using simulation training in surgery is the perception that it might train surgeons in habits that are inappropriate to the operating room. In addition, where assessment is based on simulated tasks, students might try to "game" the system, by discovering ways to increase their score without necessarily developing the intended skills.

Tackling these challenges, ProMIS is the system to use augmented reality (AR) for laparoscopic surgical training. (See the "Related Work in Surgical Training" sidebar on the next page for a discussion of other approaches.) This application combines the best features of box trainers: real instruments, accurate tactile feedback, low system complexity, automated measurement of surgical skills, display of anatomical variations, and random surgical events possible in VR systems.

Laparoscopic Surgery

Laparoscopy is a type of minimally invasive surgery where surgeons use long, narrow surgical instruments via small incisions in the abdomen, hence the common name *keyhole surgery*. Surgeons create the space to operate by inflating the abdomen with gas, typically carbon dioxide. They then view the surgical field by inserting a camera into one of the incisions. To perform these tasks, surgeons have a number of psychomotor challenges: they must learn to operate by looking at a TV monitor, reverse their hand movement directions to compensate for the fulcrum effect of the abdomen, and relearn the skills of physical manipulation to compensate for losing their sense of touch and the reduced degrees of freedom of the laparoscopic instruments.

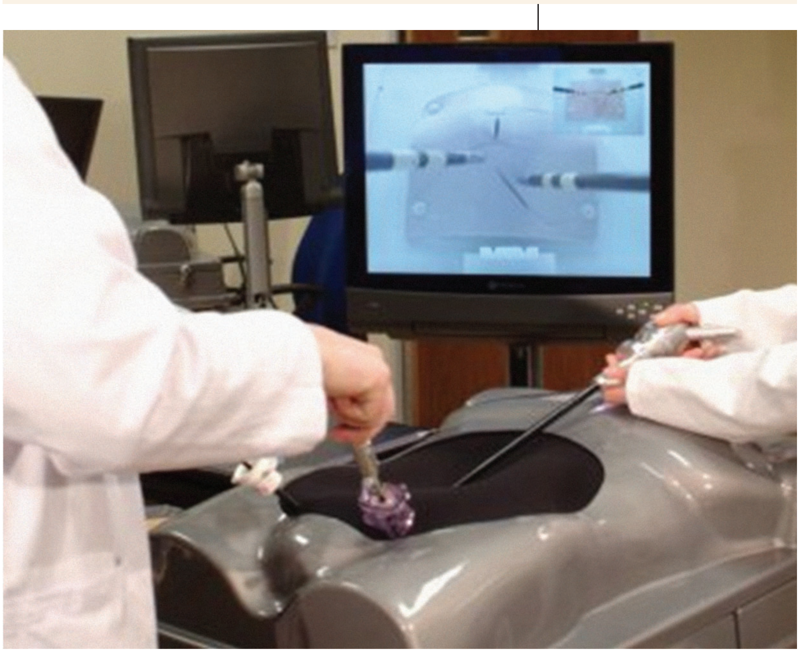

Figure 1. The ProMIS simulator. In this suturing training session, one trainee is operating and a second is holding a laparoscopic camera.

ProMIS simulator design

Our objective when designing ProMIS was to let trainees use real surgical instruments within the simulation and have them interact with both physical objects and graphical content using the same interface. Additionally, we didn't want to modify the instruments by changing their physical characteristics (such as weight or cutting or grasping mechanisms) or constrain users by asking them to attach wires to their hands or instruments.

The primary innovative step in ProMIS was to build a stereo vision system entirely within a patient manikin to track laparoscopic surgical instruments (see Figure 1). This approach doesn't require significant instrument modification, allows us to add new instruments easily, and lets us track up to five instruments simultaneously. Instrument design is a dynamic area, and vision tracking lets users change the instrument type and manufacturer without having to modify the system.

With ProMIS, users perform physical training tasks on custom trays that are mounted on the bodyform's base. Users can see their instruments working on the task by means of a video camera mounted inside the bodyform. We routed the camera feed through a PC and created the AR content by composting graphics onto this live video feed. We made VR tasks possible by replacing the video feed with a purely graphical scene. A single PC, typically a laptop, performs the real-time tracking of the instruments and generates the 3D graphical scenes.

Tracking algorithm

Students insert instruments into the bodyform from the top via a neoprene skin in the abdomen. The bodyform's internals are lit by an internal lighting system, and the walls on the inside of the bodyform are a light color. In this controlled environment, it's possible to segment

Authorized licensed use limited to: Maynooth University Library. Downloaded on January 30,2023 at 16:27:29 UTC from IEEE Xplore. Restrictions apply.

Related Work in Surgical Training

Researchers have demonstrated the use of augmented reality (AR) in various dexterous manipulation tasks such as assembly,[1] aircraft maintenance,[2] and electrical power maintenance.[3] In medical applications, researchers have widely proposed AR for many procedures[4] such as needle biopsy,[5] spinal surgery,[6] breast biopsy,[7] tumor ablation,[8] and laparoscopic surgery.[9] Within the medical training context, AR has been proposed for training in liver surgery,[10] patient clinical examination,[11] and forceps delivery.[12]

The ability to track infrared targets attached to the surgical instruments has been commercially available for some time from companies such as Medtronic, Brianlab, and Stryker. However, these systems are designed for surgery and their high cost prohibits their use in a training system. Also, in the ProMIS system, it would have been difficult for an external optical-tracking system to maintain a line of sight with all instruments given the common situation where many people are crowded around the training simulator (typically, three students plus an instructor).

Researchers have also used magnetic tracking in a training context. Nick Taffinder and his colleagues[13] implemented a skills assessment device based on magnetically tracking surgeons' hands. However, limitations on the number of instruments that could be used, the modification of the instruments, and magnetic interference limitations prevented using this system in the ProMIS simulator.

References

1. A.C. Boud et al., "Virtual Reality and Augmented Reality as a Training Tool for Assembly Tasks," *Proc. IEEE Int'l Conf. Information Visualization*, IEEE CS Press, 1999, pp. 32-36.
2. T. Haritos and N.D. Macchiarella, "A Mobile Application of Augmented Reality for Aerospace Maintenance Training," *Proc. 24th Digital Avionics Systems Conf.* (DASC 2005), vol. 1, IEEE Press, 2005, pp. 5.B.3-5.1-9.
3. C. Nakajima and N. Itho, "A Support System for Maintenance Training by Augmented Reality," *Proc. 12th Int'l Conf. Image Analysis and Processing* (ICIAP 03), IEEE Press, 2003, pp. 158-163.
4. J. Fischer et al., "Medical Augmented Reality Based on Commercial Image Guided Surgery," *Proc. Eurographics Symp. Virtual Environments*, Eurographics, 2004, pp. 83-86.
5. M. Rosenthal et al., "Augmented Reality Guidance for Needle Biopsies: An Initial Randomized, Controlled Trial in Phantoms," *Medical Image Analysis*, vol. 6, no. 3, 2002, pp. 313-320.
6. P. Brodeur et al., "A Points-to-Surfaces Matching Technique for the Application of Augmented Reality during Spine Surgery," *Proc. IEEE 17th Ann. Conf. Eng. in Medicine and Biology*, vol. 2, IEEE Press, 1995, pp. 1197-1198.
7. Y. Sato et al., "Image Guidance of Breast Cancer Surgery Using 3-D Ultrasound Images and Augmented Reality Visualization," *IEEE Trans. Medical Imaging*, vol. 17, no. 5, 1998, pp. 681-693.
8. S. Nicolau et al., "An Augmented Reality System to Guide Radio-Frequency Tumour Ablation," *Computer Animation and Virtual Worlds*, vol. 16, no. 1, 2005, pp. 1-10.
9. J. Marescaux et al., "Augmented-Reality-Assisted Laparoscopic Adrenalectomy," *J. Am. Medical Assoc.*, vol. 292, no. 18, 2004, pp. 2214-2215.
10. G. Welch et al., "Immersive Electronic Books for Surgical Training," *IEEE MultiMedia*, vol. 12, no. 3, 2005, pp. 22-35.
11. F.D. McKenzie et al., "Augmented Standardized Patients now Virtually a Reality," *Proc. 3rd IEEE and ACM Int'l Symp. Mixed and Augmented Reality* (ISMAR 04), IEEE CS Press, 2004, pp. 270-271.
12. T. Sielhorst, T. Blum, and N. Navab, "Synchronizing 3D Movements for Quantitative Comparison and Simultaneous Visualization of Actions," *Proc. 4th IEEE and ACM Int'l Symp. Mixed and Augmented Reality*, IEEE CS Press, 2005, pp. 38-47.
13. N. Taffinder et al., "Objective Measurement of Surgical Dexterity—Validation of the Imperial College Surgical Assessment Device," *Minimally Invasive Surgery and Allied Techniques*, vol. 7, no. 1, 1998.

the instruments by background segmentation and search the top third of the images for narrow shafts. Because the instruments' shafts are smooth, they can be segmented reliably using edge-detection algorithms.

During a lesson, the instruments' tips were often occluded by the training task, making accurate depth-of-insertion measurement difficult without some additional information. We resolved this issue by placing a small label at a specific distance from the instrument's tip—this is the only modification to the instruments required. We could have also measured the instrument's rotation by examining a pattern on the label, but in the final analysis, instrument rotation didn't prove material to measuring surgical skill. Figure 2 shows the label on the instrument.

Calibrating the tracking cameras

Calibrating a camera means finding a set of parameters that describe the image formation's mechanics. This information helps the system perform real-world measurements. We calibrated the ProMIS system using the method developed by Zhengyou Zhang.[11] During the assembly procedure, the assembly operator is guided through the procedure using an interactive application. To calibrate the cameras' positions with respect to each other and the bodyform, we placed the checkerboard target into the task-tray mounting slot.

Authorized licensed use limited to: Maynooth University Library. Downloaded on January 30,2023 at 16:27:29 UTC from IEEE Xplore. Restrictions apply.

Tracking validation

The tracking system tracks objects in a 30 × 30 × 15 cm volume, which is located 30 cm from the baseline of the stereo system. We used wide-angle video lenses to effectively image the volume. The resulting theoretical precision for a 320 × 240 pixel image was 1 to 2.4 mm per pixel depending on the distance from the camera. Because the objects we're tracking are linear, it's possible to interpolate across pixel boundaries using standard techniques[12] and achieve a repeatable root mean square (RMS) accuracy of 0.5 mm.

The focus of the tracking is the central area of the tracking volume where we fixed the interchangeable training tasks to the tray. We rigidly mounted the cameras and the drawer to the same base plate to help the system produce repeatable measurements.

To verify the accuracy and tune the system's performance, we used several standard tests, namely,

- static repeatability tests,
- gross-motion tracking accuracy tests, and
- micromotion tracking accuracy tests.

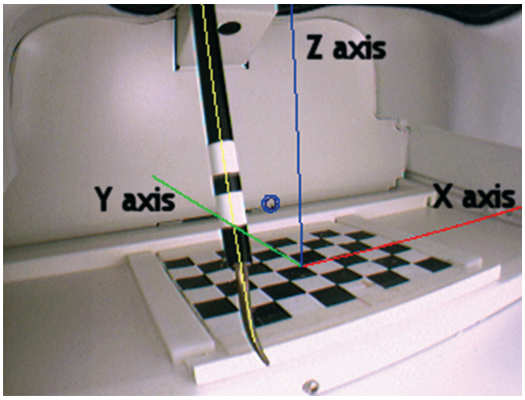

Figure 2. View of the tracking cameras. The image shows a surgical instrument as seen by one of the tracking cameras; a checkerboard calibration target is located where the surgical task tray is fixed to the base of the unit.

We performed static repeatability tests by placing the instrument in a number of defined positions within the tracking volume. We measured the instrument's position repeatedly to quantify the effect of noise on the system measurements. These tests showed that the RMS error in the instrument tip's position estimation was 0.5 mm.

Gross-motion tracking accuracy tests involved the instrument being translated parallel to one axis across the tracking volume. We achieved this using a manual photographic gantry system with a graduated scale.

Figure 3 shows the results' overall stability; some variation is noticeable in the results because of the slight misalignment of the axes of the gantry and the tracking system and vibration due to the experiment's manual nature.

We used micromotion tracking accuracy tests to ensure the subpixel tracking's accuracy across pixel boundaries. We attached the instrument to a manual micropositioning stage. This arrangement allowed the instrument to be moved in

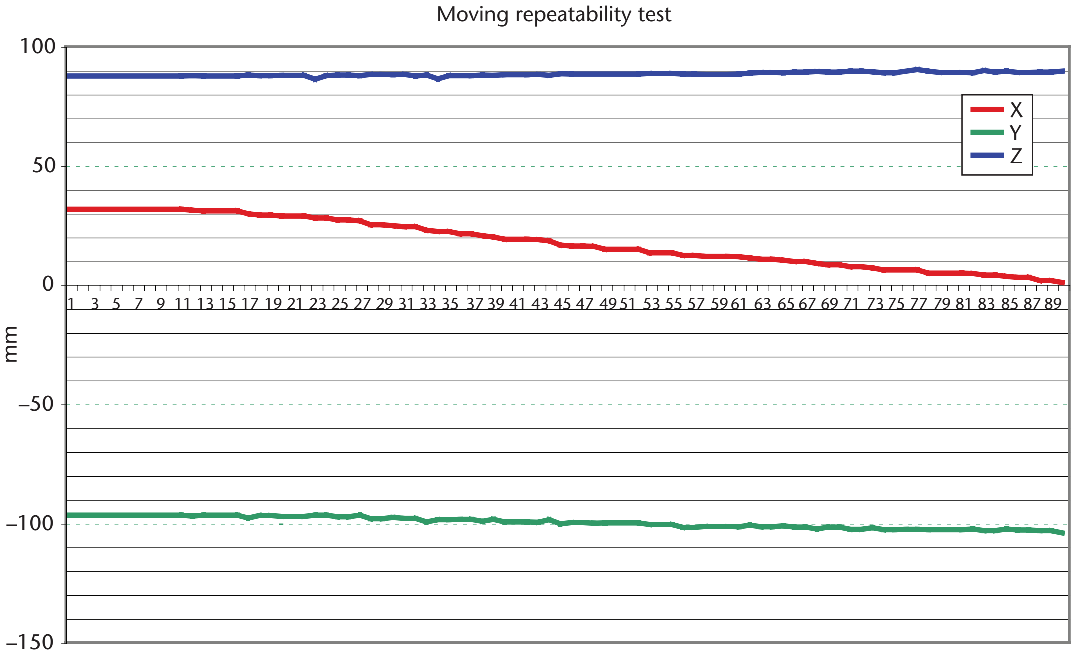

Figure 3. Graph of the gross motion tests showing the stability of tracking as the instrument is moved through the tracking volume.

Authorized licensed use limited to: Maynooth University Library. Downloaded on January 30,2023 at 16:27:29 UTC from IEEE Xplore. Restrictions apply.

Figure 4. ProMIS hand-assisted laparoscopic colectomy platform. Peter Geis (a medical doctor), with one of the authors, Derek Young, is using the hand accesses device to manipulate objects within ProMIS.

small increments, at which point we repeated the static repeatability tests. The resulting error was of the same order of magnitude (0.5 mm RMS) as the static test, and variation in the tool tip position was mostly in the *Z* direction because of the step of locating the icon on the instrument shaft, which was adversely affected by some artifacts caused by the Bayer color mask on the sensor.

The vision system demonstrated stable tracking over the entire volume with a 0.5-mm RMS error at 30 Hz. This compares favorably with magnetic-tracking devices because ProMIS wasn't adversely affected by metal instruments or by instruments of different diameters, lengths, or manufacturers. It wasn't necessary to attach wires or sensors to the instruments, which let users feel the weight of their instruments and have full natural movement. We only required users to affix a simple adhesive label to the instruments at a specific offset from the tip to achieve accurate tracking. This step typically took 30 seconds per instrument using the instrument marking kit we provided.

Multiple instrument tracking

We built the system to track up to five instruments. We chose five as the upper limit because this corresponds to two instruments for a lead surgeon, two for an assistant, and a camera for a second assistant. In the worst case, these instruments will be identical; we used geometric constraints to resolve this potential ambiguity in matching—that is, the instruments pivot about entry points in the neoprene creating **a** geometric constraint in the stereo line-matching problem.

An entry point is the 3D point at which a line intersects a plane. In ProMIS, the plane we used is horizontal to the physical skin where the instruments enter the body form. We register entry points as each new instrument is inserted into the body form.

Hand tracking in the HALC procedure

In laparoscopic colectomy procedures, the surgeon removes a piece of the large intestine from a patient. This is a technically challenging procedure that uses multiple laparoscopic instruments, graspers, staplers, and a large circular stapler that is inserted anally. Several features of this procedure require tactile feedback and would be very challenging to simulate using VR. One approach to this procedure is to insert the surgeon's hand into the body via a wound protector. We adapted the ProMIS platform to simulate this procedure, called a hand-assisted laparoscopic colectomy (HALC). A key aspect of this adaptation is that the task tray is now much larger, incorporating a physical model of the colon and several anatomical structures around it. In addition to the instruments, we must track the surgeon's hand to extract potential performance metrics.

Figure 4 shows the ProMIS HALC platform. The only constraint on the hand tracking is the surgeon's typical, purple surgical gloves. Because we're only tracking one hand in the scene and it's within a limited range of possible color, we use the camshift algorithm[13] to segment the hand from the background. In the current system, we use the location and properties bounding ellipse to calculate the hand's metrics.

Instructional design and mixed-reality interaction

The ProMIS system is primarily a training device, and as a result, we paid a great deal of attention to the instructional design. We identified and divided the key skills of laparoscopic surgery into specific tasks such as camera navigation, instrument handling, sharp dissection, clipping, and so on.

ProMIS task design

Since the development of minimally invasive surgery, surgeons have used box trainers and a series of graduated tasks to learn and develop their psychomotor skills (such as bead transfer, peg insertion, knot tying, and peeling a grape) in addition to the more formal training programs. Because of its design, ProMIS can incorporate the

Authorized licensed use limited to: Maynooth University Library. Downloaded on January 30,2023 at 16:27:29 UTC from IEEE Xplore. Restrictions apply.

clinically validated box trainer and VR tasks, such as camera navigation or diathermy. We blended AR with physical tasks to instruct students, give error warnings, and create the random events typical in surgery that the students must deal with in a timely fashion.

Lesson structure and tools

The ProMIS lesson structure is consistent across all the lessons; initially, the simulator demonstrates a task to a trainee via a video clip, and then the student practices that task. During the practice, phase guidance helps students with the task's steps via AR animations or picture-in-picture video.

At the end of each task, trainees get quantitative and qualitative feedback on their performance. The qualitative feedback shows trainees a VR replay of their task that traces their instrument movements. This focuses attention on the efficiency of their movement because the traced path clearly shows any wild or out-of-control movements. The quantitative feedback includes their scores on ProMIS metrics, in the context of the proficiency score the instructor expected for this task. In addition to objective scores, this feedback flags specific errors such as letting an instrument drift out of the surgical field.

Following the summative assessment, we provide students with a self-assessment questionnaire. This encourages trainees to reflect on their performance and focus on areas for improvement.

As students progress through a series of tasks requiring increasing levels of skill and dexterity, the system records their performance and makes it available to the instructor. The system records videos of each training session, which instructors can graphically annotate to provide students with specific feedback. This interface lets each institution set its own proficiency score, which typically reflects the trainees' level and the expected proficiency target.

Delivering augmented reality

Many of the ProMIS lessons overlay graphical content on the trainees' video camera. Although some of this is simple picture-in-picture video, the majority is AR registered to the image's contents, creating the illusion that it's a natural part of the image at a specific 3D depth. We created AR within the system by blending the live video feed with selected objects from the VR system. Systems such as the AR toolkit[14] track icons and matte the graphical objects onto the video image. Our approach is to register the tracking cameras, body form, and the graphics engine to one coordinate reference frame and track all the moving objects in the scene.

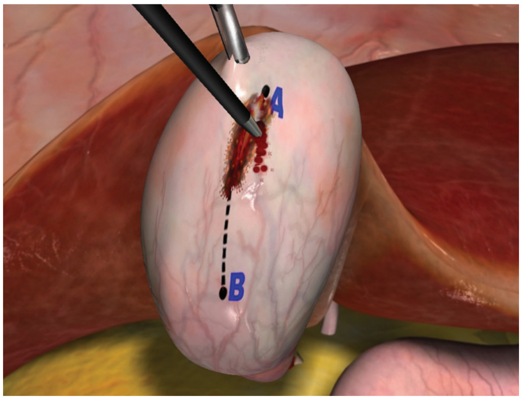

Figure 5. Example VR task (electrocautery) that requires little haptic feedback. Trainees must cut along the line A to B without touching any other surface of the VR organs.

To achieve the correct perspective projection, we calibrated the user camera in a similar manner to the tracking cameras. We used these parameters to set the virtual camera's parameters in the graphics engine.

Both the task tray and moving instruments' fixed objects are fully modeled within the graphics engine. We don't render these models when creating AR, however; they're in the z-buffer so other graphical objects rendered by the graphics engine will be occluded when the instrument comes between the user and the object, creating the illusion of the object at that depth. The instruments' 3D models update 30 times a second to follow the position of the real instruments using the position information the tracking system provides.

VR and mixed-reality interactions

The camera-based tracking system provides the instruments' 3D position, which lets us create a pure VR simulation. This approach is particularly suited to training in tasks that require little haptic feedback such as pure hand–eye coordination or clinical skills such as laparoscope orientation or electrocautery, as Figure 5 shows. If some physical interaction is necessary, as in some gynecological procedures, a registered physical model can blend real haptics with virtual images.

Tracking the instruments in the mixed-reality scene lets users naturally and directly interact with virtual objects. In simple training tasks, students receive immediate feedback about whether they're moving correctly when objects change colors. For example, if students try to touch a target, this target will change color after their instrument tip has been in contact with the object for a specified time period; alternatively, if their

Authorized licensed use limited to: Maynooth University Library. Downloaded on January 30,2023 at 16:27:29 UTC from IEEE Xplore. Restrictions apply.

Figure 6. Augmented-reality (AR) example. The trainee is using a real harmonic scalpel to use ultrasound to stem the AR bleeding.

instrument drifts out of the surgical field and into other tissues, the screen will flash red.

To provide students with adequate surgical context, we built certain visual effects into the simulation, including smoke (from electrocautery), bleeding, and water (from irrigation). Figure 6 gives an example of a user interacting with AR bleeding. In this image, the surgeon is using a real harmonic scalpel to use ultrasound to stem the AR bleeding.

To provide a link between the “patient” orientation and the simulation’s behavior, we equipped the body form with a tilt sensor so that as the students change the patient’s tilt (a common procedure in surgery) the AR organs respond by sliding out of the surgical field. In advanced procedures such as HALC, students must identify certain vessels (such as veins, arteries, and urethra) by touching the surrounding tissues. (The students then avoid these vessels because cutting them is a serious error.) To achieve this type of simulation, we developed an animatronic system to create the pulsation effect in tubes embedded within plastic organs. Thus, we effectively included subtle but important haptic cues into the simulation.

To provide training both prior to and during a training session, we presented the user with multimedia learning content using a video pop-up encapsulated within the live surgical field or as an AR animation with instructional audio. In contrast to in vivo AR where the simulation aims to be as realistic as possible, we found that instructional AR was most effective when presented as a cartoon.

Assessing surgical skills

The question “What makes a good surgeon?” has been on the minds of doctors and patients probably since the beginning of medicine. The current focus of research and practice is on creating more objective measures of medical skills. Specific to surgery, many researchers are actively seeking objective measurements, validated training tasks, and proof-of-skills transfer to the operating room.

Measuring surgical skill

One of the reasons for tracking surgical instruments was to produce an objective measure of surgical skills in real and virtual tasks. The objective assessment of surgical skills in both open and laparoscopic surgery is an active research area. A research team in Toronto led by Richard Resnick[15] developed and validated the Objective Structured Assessment of Technical Skills (OSATS) paradigm in which human observers rated videos of subjects performing bench tasks using task-specific checklists. The method also uses an additional global rating step that includes general parameters such as respect for tissues, instrument handling, and task flow. This paradigm is labor intensive for the expert reviewers.

A research team from Imperial College London[16] developed a method for producing objective scores of surgical skills using a magnetic motion tracker to record the movement of the surgeon’s hands. This work defined a quantitative measure for some of the subjective measures in OSATS. This work combined the objective assessments with end-product assessments such as knot-break strength, cutting latex sheets in error, and the leaking of sutured tubes.

In ProMIS, we track the tip of each instrument and analyze the path to extract a number of base metrics—that is, the time it takes to complete the task, overall distance traveled by the instrument tip, and the instrument’s economy of motion. This last metric is based on the observation that trainee surgeons often overshoot their target point or find that they must adjust their orientation upon reaching the goal point. Expert surgeons are more accurate in their movement and rarely need to adjust their orientation on arriving at the goal because the necessary path planning has become second nature to them. Consequently, it’s possible to identify surgeons’ skill levels by measuring the number of direction changes as they work through the tasks.

In addition to the base metrics, we measured other performance aspects in the training tasks, including whether the tasks were completed in the correct sequence, instruments’ tips drifted out of the laparoscope’s field of view, and both hands worked together equally. To help detect such

Authorized licensed use limited to: Maynooth University Library. Downloaded on January 30,2023 at 16:27:29 UTC from IEEE Xplore. Restrictions apply.

errors, we decomposed the training tasks into a series of steps. We ensure trainees correctly complete these steps by associating each step with an appropriate 3D location and motion pattern of the instruments.

Because ProMIS is an AR simulation, we can use traditional assessments such as examining a task's end product—for example, by looking for knot-break strength and punctures to latex membranes—in addition to the motion-analysis-based assessments. This provides substantial validity to both the training and assessment and is a key advantage of the mixed-reality approach.

Surgical skill performance metrics in ProMIS

The fundamental information the system gathers is the XYZ position of the instrument tips. Additional information is possible such as the relative position between the instrument tips and each other, distance from parts of the task tray, and important factors such as whether the instrument tip is within the frustum of the user camera. We use many of these latter measurements to detect procedural student errors. The fundamental measurements of skill—the time it takes to complete a task, path length, and smoothness—are independent of the surgical task's nature. (These performance metrics are similar to those that Vivek Datta et al.[17] used to assess surgical skills using hand motions.)

Specifically, we measure the time from when students start their lesson to when they remove their instruments. The path length is the sum of the position data's intersample distances. We calculate smoothness using this basic algorithm:

1. Smooth the position data to remove hand tremors.
2. Take the first derivative of step 1 data.
3. Calculate the absolute value of step 2 data.
4. Take the derivative of step 3 data.
5. Count the number of times the step 4 data crosses the x axis to calculate the smoothness score.

To count only significant movements, we apply a significance factor in step 5—that is, we only count *zero crossings*, where the absolute value of the intersample distance is above a 200 mm per second^{-3} threshold.

Validating surgical skill measurements

To validly assess surgical skills and discriminate proficient surgeons from novices, we needed a consistent measurement of instrument position and a task design that challenges the surgeon to use their skills in a manner similar to the operating room. When assessing a training scenario's validity, we can use face or construct validity or measure the transfer of skills from the simulator to the operating room. *Face validity* is determined by asking a number of surgeons if a task feels like, or looks like, a situation that they would face in the operating room. There's a tendency to concentrate on the assessment's visual aspects rather than the behavior of simulated tissues or instruments; consequently face validity has the potential to create a misleading impression of a simulation's educational value. *Construct validity* measures the ability of a training task to reliably tell the skill levels of experts from novices. This is an important stage in assessing a simulator's educational value because it verifies that we can use the task to examine a student's skill level.

The final form of assessment—*assessing skills transfer* to the operating room—assesses the degree to which learned skills transfer to actual use in the operating room. Such studies, sometimes called VR-to-OR studies, compare traditional training methods in surgery against novel ones such as simulation. The metric for comparing performance is the number of errors that the students make in the operating room.

In the case of ProMIS, constructing validity on a task-by-task basis was clearly the key type of assessment required. To increase the medical community's confidence in our metrics for assessing surgical skills, it was critical that we have the ProMIS platform validated by independent labs. We partnered with a number of institutions to support studies on the validity of ProMIS as a training device. In recent years, numerous centers have also independently verified the metrics in ProMIS and expanded its use for robot surgery and novel training tasks. (All these studies have been reported elsewhere in more detail, but we summarize the results here.)

First independent validation study

David Broe and his colleagues carried out the first study in the Adelaide and Meath National Children's Hospital (AMNCH) Surgical Unit in Dublin, Ireland.[18] This study evaluated three ProMIS tasks:

Authorized licensed use limited to: Maynooth University Library. Downloaded on January 30,2023 at 16:27:29 UTC from IEEE Xplore. Restrictions apply.

Table 1. Results of the second validation study.[19]

Measure	Novices	Experts	*p* value
Time	421.5 ± 128	123.9 ± 36	0.0001
Path length	408.1 ± 138	149 ± 53.9	0.0001
Smoothness	248.7 ± 92.3	67.7 ± 28.3	0.0001

- *camera navigation*, a VR task that trains users in finding specific anatomy within the abdomen;
- *sharp dissection*, a physical task in which trainees are presented with a double-layer latex balloon and are required to cut shapes out of the top layer without puncturing the lower layer; and
- *laparoscopic needle handling*, a task in which trainees are asked to pass a needle and thread via four hoops using laparoscopic needle drivers.

These tasks were used to assess the skill levels of 20 surgical residents from novices to experts. No prior training was allowed to reduce familiarity bias. The expert level was set for those individuals who had a Certificate of Completion of Specialist Training (CCST) or a consultant (attending) level with more than 100 laparoscopic procedures. The researchers divided the subjects into five groups similar to the American Post Graduate Year (PGY) 1 to 5 classification: group 1 consisted of interns ($n = 7$); group 2, senior house officers ($n = 6$); group 3, midgrade registrars ($n = 1$); group 4, higher surgical trainees ($n = 3$); and group 5, experts ($n = 3$). They also compared the ProMIS metrics with an alternative, previously validated, assessment measure. Consequently, Broe and his team developed an OSATS protocol for each level of each task. Two experts, with an excess of 100 laparoscopic procedures and previous experience of OSATS, scored videos of the subjects in a double-blind fashion.

The procedure's first step was to verify that the OSATS scoring system was reliable by measuring the inter-rater reliability using Cronbach alphas. This measure showed 0.88 for the task-specific checklist and 0.93 for the global score of overall performance, proving OSATS was a reliable score of subject performance.

The camera-navigation task showed mixed-construct validity and a plateau effect for experienced subjects. This is because camera navigation is one of the first tasks surgical trainees learn and it was difficult to challenge the skill levels of all subjects on such a simple task.

Sharp dissection showed good construct validity throughout, but again, the Broe team noted a plateau effect among experienced subjects.

Laparoscopic needle handling showed good construct validity and an excellent correlation with increasing levels of experience.

The study concluded that the camera-navigation task was too simple to discriminate skills because surgeons acquire this skill quickly. The sharp-dissection and needle-handling tasks showed higher correlation with skill level. Because the dissection task showed a plateau effect, it's only useful for assessing the skill levels of subjects up to the PG3 level. Since the needle-handling task didn't show a plateau effect, assessment of these complex skills would normally be confined to later years.

Second independent validation study

Kent Van Sickle and his colleagues carried out the second validation study at the E-STAR Center at Emory University in Atlanta, Georgia.[19] In this study, the research team developed a suturing task where subjects were asked to pass a laparoscopic needle between two paired circular targets in a latex sheet.

This study assessed five experts and five novices for baseline perceptual, visiospatial, and psychomotor abilities using Minimally Invasive Surgical Trainer (MIST-VR) and pictorial surface orientation and cube comparison tests. The Van Sickle team found no significant differences between the intrinsic abilities that they were measuring. (They wanted to ensure that their results were not skewed by having an outlier among their subjects.)

On the suturing task, the expert group outperformed the novice group on all the metrics, as Table 1 shows. The researchers compared mean performance (*p* value) using the Mann-Whitney *U* test and showed with strong certainty the ability to discriminate skill levels. Van Sickle's team repeated the trials three times with the experts showing greater consistency in their performance than the novice group. Both groups showed improvement in each successive trial, but this didn't reach a statistically significant level.

Third independent validation study

Imperial College London's Surgical Unit carried out the third study.[20] This study had 32 subjects of differing laparoscopic experience perform

Authorized licensed use limited to: Maynooth University Library. Downloaded on January 30,2023 at 16:27:29 UTC from IEEE Xplore. Restrictions apply.

three tasks on the simulator. The researchers segmented the group into novice surgeons (*n* = 12, with no laparoscopic procedures), trainees (*n* = 11, with 5 to 50 laparoscopic procedures), and experts (*n* = 9, with more than 100 laparoscopic procedures).

The three tasks selected from the ProMIS curriculum were object positioning, clip and cut, and sharp dissection. The performance metrics were analyzed using nonparametric tests, and significant differences were observed in the performance of all three groups for all parameters on object positioning ($p = 0.001$) and sharp dissection ($p = 0.001$). For the clip and cut exercise, there was a plateau effect—that is, the researchers saw no significant difference between experts and trainees ($p = 0.489$), although novices were significantly worse ($p = 0.005$).

Successive independent validation studies

Since the publication of the first three studies in 2004, there have been multiple validation studies performed in leading centers around the world.

Matt Ritter and his colleagues had 60 subjects (divided into novice, intermediate, and experienced surgeons) perform five trials of the Fundamental of Laparoscopic Surgery (FLS) peg transfer task.[21] They scored the subjects using ProMIS and FLS scores.

For each of the five trials, experienced subjects outperformed intermediates, who in turn out performed novices. In addition, significant differences were seen between the groups using FLS scores ($p < 0.001$), ProMIS path lengths ($p < 0.001$), and ProMIS smoothness ($p < 0.001$).

FLS scores correlated highly with ProMIS scores, indicating the concurrent validity of the ProMIS metrics and that using ProMIS can significantly reduce the personnel requirements of assessing FLS skills while maintaining the objectivity. Anthony McCluney and his colleagues performed a similar study using five laparoscopic novices and five experts and reached similar conclusions.[22]

Erika Fellinger and her colleagues compared the ProMIS simulator with the Medical Education Technologies Incorporated (Meti) VR simulator using 73 subjects at the 2005 Sages meeting.[23] Subjects (who classified themselves as nonexpert and expert) were asked to perform two iterations of laparoscopic suturing and intracorporeal knot tying.

The research team collected performance data and used a six-question survey to define impressions of task realism, relevance, and execution. The task completion rate was 80 percent for the Meti device and 93 percent for ProMIS. Experts performed better than nonexperts for all performance measures on both devices. Post-task survey scores for ProMIS were significantly higher for perceived realism, reflection of clinical ability, and overall educational value.

Vimal Narula and his colleagues used ProMIS to objectively assess task-performance data with the Intuitive Surgical DaVinci robot instrumentation.[24] Their study showed that all the tasks were performed faster and with more precision using the robotic technology than using standard laparoscopy.

Laurel Vuong and her colleagues undertook a study with 21 subjects (six medical students, 14 surgical residents, and one expert surgeon) performing peg transfer, pattern cutting, pre-tied loop placement, extracorporeal, and intracorporeal knot tying.[25] Their results showed that automatic and objectively measured motion derivatives can be associated with experience level.

Michael Pellen and his colleagues undertook an 83-subject study (with 17 experts, 38 novices, and 28 basic surgical trainees) to assess performance on three trials of the sharp-dissection task.[26] Experts performed all three tasks significantly faster, smoother, and with more economy of movement ($p < 0.05$). Experienced participants performed sharp dissection more accurately ($p < 0.01$), although the study showed no difference in balloon puncture.

Conclusion

Because of ProMIS's use of mixed reality, we can combine plastic simulated tissue with AR graphics to provide students with accurate haptic and visual feedback. The mixed-reality simulation represents a significant advantage in terms of fidelity, flexibility, and complexity over a pure VR simulation. This mixed-reality architecture also facilitates the delivery of various multimedia training content such as video, audio, and animation.

Camera-based tracking is a powerful and unobtrusive means of objectively measuring performance and delivering AR learning content. We intend to extend the ProMIS system's architecture to cover a wider array of clinical and industrial training tasks. We'll also support the final stage in the simulator's assessment to demonstrate that the skills learned on ProMIS result in lower errors in the operating room. **MM**

Authorized licensed use limited to: Maynooth University Library. Downloaded on January 30,2023 at 16:27:29 UTC from IEEE Xplore. Restrictions apply.

References

1. R. Smith, “All Changed, Changed Utterly,” *British Medical J.*, vol. 316, no. 7149, 1998, pp. 1917-1918.
2. L.T. Kohn, J.M. Corrigan, and M.S. Donaldson, *To Err Is Human: Building a Safer Health System*, Committee on Quality of Health Care in Am., Inst. of Medicine, Nat’l Academy of Sciences, 2000.
3. R.M. Satava, “Virtual Reality Surgical Simulator,” *Surgical Endoscopy*, vol. 7, no. 3, 1993, pp. 203-205.
4. N.E. Seymour et al., “Virtual Reality Training Improves Operating Room Performance: Results of a Randomized, Double-Blinded Study,” *Annals of Surgery*, vol. 4, no. 236, 2002, pp. 458-464.
5. A.G. Gallagher and C.U. Cates, “MD Approval of Virtual Reality Training for Carotid Stenting: What This Means for Procedural-Based Medicine,” *J. Am. Medical Assoc.*, vol. 292, no. 24, 2004, pp. 3024-3026.
6. R. Kneebone, “Simulation in Surgical Education: Educational Issues and Practical Implications,” *Medical Education*, vol. 37, no. 3, 2003, pp. 267-277.
7. G.M. Fried et al., “Proving the Value of Simulation in Laparoscopic Surgery,” *Annals of Surgery*, vol. 3, no. 240, 2004, pp. 518-528.
8. J.H. Peters et al., “Development and Validation of a Comprehensive Program of Education and Assessment of the Basic Fundamentals of Laparoscopic Surgery,” *Surgery*, vol. 135, no. 1, 2004, pp. 21-27.
9. F.J. Carter et al., “Consensus Guidelines for Validation of Virtual Reality Surgical Simulators,” *Surgical Endoscopy*, vol. 19, no. 12, 2005, pp. 1523-1532.
10. C.Y. Ro et al., “The Impact of Haptic Expectations on Initial Lapsim Performance: Prior Laparoscopic Experience Does Not Predict Performance,” *Proc. Sages 2005*, Soc. Am. Gastrointestinal and Endoscopic Surgeons, 2005.
11. Z. Zhang, “A Flexible New Technique for Camera Calibration,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, 2000, pp. 1330-1334.
12. R.B. Fisher and D.K. Naidu, “A Comparison of Algorithms for Subpixel Peak Detection,” *Advances in Image Processing, Multimedia and Machine Vision*, Springer-Verlag, 1996, pp. 168-169.
13. G.R. Bradski, “Real Time Face and Object Tracking as a Component of a Perceptual User Interface,” *Proc. 4th IEEE Workshop on Applications of Computer Vision* (WACV 98), IEEE CS Press, 1998, pp. 214-219.
14. H. Kato and M. Billinghurst, “Marker Tracking and HMD Calibration for a Video-Based Augmented Reality Conferencing System,” *Proc. 2nd Int’l Workshop on Augmented Reality* (IWAR 99), IEEE CS Press, 1999, pp. 85-94.
15. H. Faulkner et al., “Validation of and Objective Structured Assessment of Technical Skills for Surgical Residents,” *Academic Medicine*, vol. 71, no. 12, 1996, pp. 1363-1365.
16. N. Taffinder et al., “Objective Measurement of Surgical Dexterity—Validation of the Imperial College Surgical Assessment Device,” *Minimally Invasive Surgery and Allied Techniques*, vol. 7, suppl. 1, 1998, p. 11.
17. V. Datta et al., “The Use of Electromagnetic Motion Tracking Analysis to Objectively Measure Open Surgical Skill in the Laboratory-Based Model,” *J. Am. College of Surgeons*, vol. 193, no. 5, 2001, pp. 479-485.
18. D. Broe et al., “Construct Validation of a Novel Hybrid Surgical Simulator,” *Surgical Endoscopy*, vol. 20, no. 6, 2006, pp. 900-904.
19. K. Van Sickle et al., “Construct Validation of the ProMIS Simulator Using a Novel Laparoscopic Suturing Task,” *Surgical Endoscopy*, vol. 19, no. 9, 2005, pp. 1227-1231.
20. J. Hance et al., “Evaluation of a Laparoscopic Video Trainer with In-Built Measures of Performance,” *Proc. 13th Int’l Congress and Ann. General Meeting Soc. Laparoendoscopic Surgeons*, Soc. Laparoendoscopic Surgeons, 2004.
21. E.M. Ritter et al., “Concurrent Validity of Augmented Reality Metrics Applied to the Fundamentals of Laparoscopic Surgery (FLS),” *Proc. Sages 2006*, Soc. Am. Gastrointestinal and Endoscopic Surgeons, 2006, pp. 104-105.
22. A.L. McCluney et al., “Validation of the ProMIS Hybrid Simulator Using a Standard Set of Laparoscopic Tasks,” *Proc. Sages 2006*, Soc. Am. Gastrointestinal and Endoscopic Surgeons, 2006, p. 194.
23. E. Fellinger et al., “Complex Laparoscopic Task Performance on Two New Computer Based Skills Training Devices,” *Proc. Sages 2006 Education/Outcomes*, Soc. Am. Gastrointestinal and Endoscopic Surgeons, 2006, pp. 190-191.
24. V.K. Narula et al., “A Computerized Analysis of Robotic Versus Laparoscopic Task Performance,” *Proc. Sages 2006 Scientific Sessions*, Soc. Am. Gastrointestinal and Endoscopic Surgeons, 2006, pp. 248-249.
25. L.N. Vuong, S. Schwaitzberg, and C.G. Cao, “What Can Motion Derivatives Tell Us about Skill Performance?” *Proc. Sages 2006 Scientific Sessions*, Soc. Am. Gastrointestinal and Endoscopic Surgeons, 2006, p. 105.

Authorized licensed use limited to: Maynooth University Library. Downloaded on January 30,2023 at 16:27:29 UTC from IEEE Xplore. Restrictions apply.

26. M. Pellen, J. Barton, and F. Horgan, “Laparoscopic Skills Acquisition: Is Psychometric Motion Analysis a Valid Assessment Tool?” *Proc. Int’l Congress of the European Assoc. for Endoscopic Surgery* (EAES 2006), 2006.

Gerard Lacey is a lecturer at Trinity College Dublin and a founder and former CTO at Haptica. His research interests include computer vison, robotics, and augmented reality. He has a PhD in computer science from Trinity College Dubilin. He is a member of the IEEE.

Donncha Ryan is a project manager at Haptica. His research interests include graphics, development tools, and augmented reality. He has a bachelor’s degree in computer science and math from Trinity College Dublin.

Derek Cassidy is a senior software engineer at Haptica. His research interests include computer vision and robotics. He has an MSc in computer science from Trinity College Dublin.

Derek Young is the director of surgical affairs at Haptica. His research interests include medical device design and innovation. He has a bachelor’s degree in mechanical engineering from Dublin Institute of Technology.

Readers may contact Gerard Lacey at Gerard.Lacey@cs.tcd.ie.

For further information on this or any other computing topic, please visit our Digital Library at http://www.computer.org/publications/dlib.

IEEE MultiMedia

OCTOBER–DECEMBER 2007

Advertiser / Products	Page Number
Ateme SA	93
The Croquet Consortium	93
Digital Foci	92
Digital Rapids	92
I-Movix	92
Sorenson Media	93

FUTURE ISSUE

January–March 2008

Upward Mobility and Media Streaming

Multimedia is about much more than being a social climber–it’s about improving our communications and mobility, as well as melding multiple technologies together to enrich our daily lives. In this issue, we present multiple systems that offer enhancements to the field, including how to dynamically transcode video in mobile environments, automatically digest commercial clips from TV streams, and greatly improve efficiency in video stream filtering.

Advertising Sales Offices

Sandy Brown
10662 Los Vaqueros Circle
Los Alamitos, California 90720-1314
USA
Phone: +1 714 821-8380
Fax: +1 714 821-4010
sbrown@computer.org

For production information, conference, and classified advertising, contact
Marian Anderson
10662 Los Vaqueros Circle
Los Alamitos, California 90720-1314
Phone: +1 714 821-8380
Fax: +1 714 821-4010
manderson@computer.org

http://www.computer.org

Authorized licensed use limited to: Maynooth University Library. Downloaded on January 30,2023 at 16:27:29 UTC from IEEE Xplore. Restrictions apply.