The American Journal of Surgery (2012) 203, 32–36

The American Journal of Surgery®

The Association for Surgical Education

A novel multimodal platform for assessing surgical technical skills

Ranil Sonnadara, Ph.D.[a,d,*], Neil Rittenhouse, M.H.Sc.[a,b], Ajmal Khan, B.A.Sc.[a,b], Alex Mihailidis, Ph.D., P.Eng.[b,c], Gregory Drozdzal[a], Oleg Safir, M.D.[a,d], Shuk On Leung, B.A.Sc.[a]

[a]Surgical Skills Centre, Mount Sinai Hospital, Toronto, ON, Canada; [b]Institute of Biomaterials and Biomedical Engineering; [c]Department of Occupational Science and Occupational Therapy; [d]Department of Surgery, University of Toronto, Toronto, ON, Canada

KEYWORDS:
Technical skills assessment;
Motion tracking

Abstract

BACKGROUND: Established methods for assessing surgical performance face limitations. Global rating scales and procedure-specific checklists are resource intensive and rely on expert opinions. Alternatives that use technology to track hand movements, such as magnetic and optical tracking systems, are generally expensive and ill suited to the surgical environment.

METHODS: The authors present a new platform that integrates a novel, low-cost optical tracking system, magnetic tracking technology and a videographic recording system to quantify surgical performance synchronously across all modalities. The validity of this platform was tested by examining its ability to differentiate between the performance of expert and novice participants on a basic surgical task.

RESULTS: Each modality was able to differentiate between expert and novice participants, and metrics were well correlated across modalities.

CONCLUSIONS: The authors have developed a platform for assessing surgical performance. It can operate in the absence of expert raters and has the potential to provide immediate feedback to trainees.

In surgical education today, there is increased awareness of patient safety, coupled with ever increasing demands on surgical faculty members. There has also been an exponential rise in the number of surgical techniques that must be mastered by surgical trainees. These factors have created a need for tools that can help trainees practice newly taught skills and that facilitate the assessment of surgical technical competence without the need for direct faculty involvement.[1,2] Current assessment strategies for technical skills in open surgery rely heavily on expert involvement, and there is growing interest in replacing traditional subjective assessment techniques with more objective methods.[1] Global rating scales (GRS) and procedure-specific checklists are routinely used to assess surgical performance in both real and simulated environments.[3] Although GRS scores have been shown to have high reliability and validity,[3,4] they rely on subjective judgments. GRS scores are also unable to provide detailed feedback, which limits their usefulness as a formative tool. Checklists have lower reliability and validity[3] but

This work was funded by the Network of Excellence in Simulation for Clinical Teaching and Learning.

* Corresponding author. Tel.: 416-586-4800 ext. 3850; fax: 416-586-8829.

E-mail address: ranil.sonnadara@utoronto.ca

Manuscript received April 18, 2011; revised manuscript August 27, 2011

doi:10.1016/j.amjsurg.2011.08.008

are less subjective. Both tools, however, require the presence of expert raters and are thus time-consuming and expensive to administer.

Recently, surgical educators have turned to technology to find new metrics for assessing technical performance on the basis of the movements made by surgical trainees as they execute surgical techniques. Parameters currently being explored include the number of movements (NM), rate of movements, total path length (PL), movement variability, time taken to complete the technique, as well as more general parameters, such as average velocity and peak velocity.[5,6] These parameters are easily measured for laparoscopic procedures using computer-based simulators that track the locations of virtual instruments at all times.[7,8] Capturing movement parameters for open procedures, in contrast, is more complex, requiring external technologies to capture the movements of either hands or instruments. These movements can then be deconstructed into basic performance metrics[9,10] or analyzed using gesture recognition.[11] The most commonly used methods of capturing movement data are video analysis, electromagnetic (EM) tracking, and optical tracking,[12] but each has its limitations in the open environment. Video analysis and optical tracking both require a clear line of sight in an environment that is typically highly obstructed, and EM tracking is highly sensitive to environmental noise. In surgical education, using video for assessing technical skills has largely been restricted to objective structured assessment of technical skills examinations,[4] in which examiners viewing the videotape are blinded to the identities of the candidates. However, the validity and reliability of video-based examinations remains an open question.[13–15] EM and optical tracking technologies have been used to track the position of the hands or instruments,[16] from which basic parameters such as PL, NM, and total time taken and more sophisticated parameters such as adherence to optimal path, average velocity, and peak velocity can be determined.[8] PL, NM, and total time taken have been found to correlate strongly with scores on objective structured assessment of technical skills exams and have hence been used as secondary measures of technical ability.[5,9]

In EM tracking systems, a varying magnetic field is generated by an alternating current or pulsed direct current source.[17,18] Both alternating current and pulsed direct current systems then use sensors that detect the fluctuating magnetic field to determine the position of the sensors relative to a known position. EM technology has been used to assess technical skills in a surgical environment by placing sensors on the dorsum of a surgeon's hand.[9] However, metallic objects near the sensor or field generator (such as some surgical instruments), and devices that generate magnetic fields of their own (such as cautery), can cause additional, unpredictable fluctuations in the magnetic field.[19] These fluctuations decrease tracking accuracy and can lead to false results.[20]

Optical tracking systems use a calibrated system of cameras to track the position, typically in 3 dimensions, of markers using a coordinate system that is common to all cameras. The position of any object that is visible to 2 cameras can be calculated at any time using stereoscopic vision techniques.[21] Commercial optical tracking systems are typically less susceptible to noise issues in comparison with EM tracking systems and are frequently used for motion capture and analysis in other domains.[22–24] They are typically very expensive, however, and require uninterrupted lines of sight between the markers and cameras, which makes optical tracking systems impractical to use in a realistic surgical training environment.

Motivated by the high costs associated with commercial optical tracking systems, efforts have been made to develop a lower cost alternative. Several groups have turned to gaming technology and have developed relatively low-cost infrared (IR) tracking systems using the cameras embedded in Wii remote controls (Nintendo, Inc., Kyoto, Japan) and a freely available camera calibration toolbox for MATLAB (The MathWorks, Inc., Natick, MA).[12,25]

The cameras embedded in Wii remotes have optical properties that make them particularly suited to environments that have short distances between the cameras and the position sensors. These properties may make such cameras more suitable for the surgical environment than other commercial systems that require much greater distances between the cameras and the objects of interest, because reducing the distance between the cameras and the markers minimizes the likelihood that markers will be masked from the cameras by other objects within the surgical field. Although each tracking modality has its limitations, a performance evaluation tool that uses multiple modalities (eg, optical, IR, and video) to synchronously collect information might free us from the limitations of each modality when used on its own.

Our laboratory has developed a custom-designed data collection platform using Wii remotes, a Polhemus Patriot EM tracking system (Polhemus, Colchester, VT), and video tracking, with an overall hardware cost of <CDN$6,000. Four Wii remotes were placed at the corners of a 4 × 2 ft operative field (see Fig. 1). The remotes were calibrated to define a tracking volume using a customized open-source calibration algorithm. The position of an IR marker within the calibrated volume could then be determined by a triangulation algorithm using information recorded from 2 Wii remotes[12,25] (4 remotes were used to maximize the likelihood that markers would remain visible to ≥2 cameras during the task). The Patriot system was used for EM tracking, and the coordinate systems of the EM and Wii tracking systems were aligned. In addition, 2 standard video cameras were used to record the task from 2 different perspectives. A custom software application suite was developed using the LabVIEW development environment (National Instruments Corporation, Austin, TX), which allowed for synchronized collection and analysis of data from all 3 modalities.

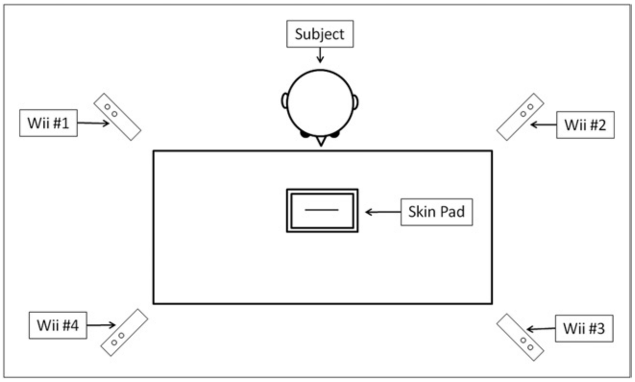

Figure 1 The experimental setup, showing the positioning of the Wii remotes around the operative field. Remotes are placed at the 4 corners of the field to maximize the likelihood that markers will remain visible to ≥2 of the remotes at all times.

Methods

This study was approved by the research ethics board at Mount Sinai Hospital, and informed consent was obtained from all participants. We explored the validity of our platform using a basic suturing task performed by 10 novices (undergraduate students with no prior training in suturing who were recruited from the Faculty of Arts and Science at the University of Toronto) and 10 experts (senior residents or practicing surgeons from the Faculty of Medicine at the University of Toronto). Both groups of participants were instructed to close a 10-cm incision on a standard silicone simulated skin pad using 8 simple interrupted sutures with 3 square knots per suture and a minimum of 3 throws on the first knot of each suture. Novice participants were first shown a 20-minute training video that taught them the required techniques and presented a series of unsupervised practice exercises. To enable tracking during the task, IR markers were taped onto the dorsa of participants' dominant hands, and magnetic markers were taped onto both hands. Hand position data were collected using both the EM and IR systems, and the NM, PL, and total time taken were calculated[5] using LabVIEW and MATLAB software. These data were analyzed using Student's *t* test for independent groups. Separate analyses were conducted for each variable. In addition, relations between the data from the IR and EM tracking modalities were examined using Pearson's product-moment correlation tests. Video recordings were also analyzed using blinded expert reviewers, who were asked to score performance using a procedure-specific GRS.

Results

Data from both IR and EM tracking modalities were normally distributed and revealed clear differences between expert and novice participants for NM and PL ($t[18] > 3.4$, $P < .01$; see Fig. 2). In addition, we observed strong correlations between the EM and IR systems for each variable ($R_{NM} = .96$, $R_{PL} = .95$). Analysis of the total time required to complete the task revealed similar results. Expert rating (GRS) data from the video recordings followed a similar pattern, with experts being scored higher than novices for all subscales ($t[18] > 4.8$, $P < .01$). Of particular interest are the scores for overall performance (novices, 2.9/5 [.74]; experts, 4.7/5 [.48]; $t[18] = 6.5$, $P < .001$) and for the quality of the final product (novices, 3.4/5 [.52]; experts, 4.7/5 [.48]; $t[18] = 5.8$, $P < .001$). There were moderate negative correlations between the GRS and EM/IR data ($R = -.49$ to $-.61$), most likely because of restricted range effects in the scores for our expert participants.

Comments

Our data show that we have produced an objective data acquisition platform that is capable of differentiating between expert and novice performers for a basic surgical task. In addition, we have shown that our low-cost, optical tracking solution is capable, at least for this task, of pro-

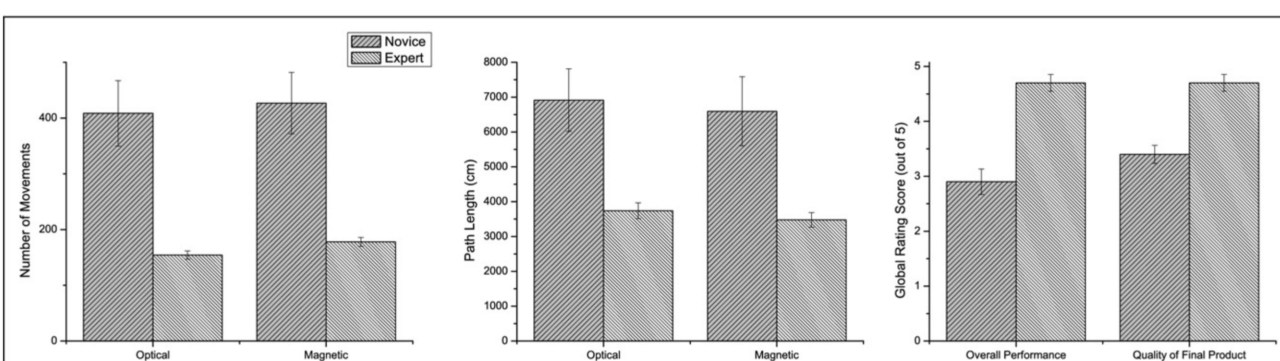

Figure 2 Mean data for NM for the dominant hand (left), PL for the dominant hand (center), and GRS scores (right) for 10 novice and 10 expert participants. Expert participants used fewer movements and traveled shorter distances when completing the suturing task than novice participants. Expert participants also received higher ratings on the GRS by blinded reviewers when video recordings of the suturing task were examined.

ducing data that are highly correlated with previously validated EM tracking tools. The moderate correlation with the expert rating scores shows potential for using this device for both formative and summative assessments. The low cost of our platform, coupled with the speed at which it can provide feedback, and its ability to assess performance in the absence of expert raters, means that it could potentially be of enormous value in supplementing existing assessment tools and as a tool for postgraduate surgical training.

EM tracking systems have been shown to be able to distinguish between novices and experts in areas such as laparoscopic cholecystectomy,[9,26] knot tying,[27,28] and labor epidural catheter placement.[29] It is conceivable that the optical tracking component of our system presented here could also distinguish between different levels of performance, making it a significantly lower cost alternative for training and assessment (the hardware cost of the optical tracking component of our platform was <$300, compared with ≥$5,000 for the EM tracking system).

With further development, our platform may also offer a lower cost alternative to existing laparoscopic simulators, which are typically extremely expensive. However, optical tracking requires a clear line of sight, and EM tracking is subject to interference from metal surgical instruments. Our system can certainly track hand motion on the basis of markers placed on the dorsa,[30] but it could not be used to accurately track instruments that are placed inside laparoscopic operating ports in its present form.

Our platform is also limited to tracking 4 markers simultaneously, which may limit the number and complexity of tasks that can be assessed. Furthermore, it is not clear that basic movement parameters are suitable for assessing performance in complex procedures (eg, a greater number of sutures, and thus more hand movements, may result in a more robust bowel anastomosis than a smaller number of sutures while producing scores indicative of novice performance), and more work is required to explore the best parameters for assessing performance in more complex tasks and surgical procedures. Preliminary work from our laboratory suggests that measures derived from gesture recognition algorithms, measures that assess the variability and accuracy of movements for a given task, and measures that incorporate information about the quality of the task (eg, the spacing of sutures) might offer more meaningful metrics of performance for these more complex tasks. These alternative metrics may also provide more useful feedback to trainees seeking to improve their operative performance than the raw NM and PL values and require further exploration.

The present study presents data from only 2 different levels of training (experts and novices). The sensitivity of our platform to finer differences in performance remains an open question. Nevertheless, our data provide compelling evidence that our platform is capable of differentiating performance, and providing feedback to trainees, for a basic surgical skill in the absence of expert raters.

Future work will include enhancing the IR tracking modality to enable it to track more objects simultaneously, exploring additional performance metrics, collecting data to enable us to model "typical" performance for different stages of training, as well as exploring the use of this platform in more complex surgical tasks and procedures and examining its sensitivity.

References

1. Fried GM, Feldman LS. Objective assessment of technical performance. World J Surg 2008;32:156–60.
2. Vassiliou MC, Feldman LS, Fraser SA, et al. Evaluating intraoperative laparoscopic skill: direct observation versus blinded videotaped performances. Surg Innov 2007;14:211–16.
3. Regehr G, MacRae H, Reznick RK, et al. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. Acad Med 1998;73:993–7.
4. Martin JA, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg 1997;84:273–8.
5. Datta V, Chang A, Mackay S, et al. The relationship between motion analysis and surgical technical assessments. Am J Surg 2002;184: 70–3.
6. Gockel I, Hakman P, Beardi J, et al. New perspectives in laparoscopic simulation: from students' skills lab to stress evaluation [article in German]. Zentralbl Chir 2008;133:244–9.
7. Schijven MP, Jakimowicz JJ. Introducing the Xitact LS500 laparoscopy simulator: toward a revolution in surgical education. Surg Technol Int 2003;11:32–6.
8. Duffy AJ, Hogle NJ, McCarthy H, et al. Construct validity for the LAPSIM laparoscopic surgical simulator. Surg Endosc 2004;19:401–5.
9. Dosis A, Aggarwal R, Bello F, et al. Synchronized video and motion analysis for the assessment of procedures in the operating theater. Arch Surg 2005;140:293–9.
10. Kundhal PS, Grantcharov TP. Psychomotor performance measured in a virtual environment correlates with technical skills in the operating room. Surg Endosc 2008;23:645–9.
11. King RC, Atallah L, Lo BPL, et al. Development of a wireless sensor glove for surgical skills assessment. IEEE Trans Inf Technol Biomed 2009;13:673–9.
12. Bouget J. Camera Calibration Toolbox for MATLAB 2010. Available at: http://www.vision.caltech.edu/bouguetj/calib_doc/index.html#parameters. Accessed October 23, 2011.
13. Scott DJ, Rege RV, Bergen PC, et al. Measuring operative performance after laparoscopic skills training: edited videotape versus direct observation. J Laparoendosc Adv Surg Tech 2000;10:183–90.
14. Beard JD. Setting standards for the assessment of operative competence. Eur J Vasc Endovasc Surg 2005;30:215–8.
15. Beard JD, Jolly BC, Newble DI, et al. Assessing the technical skills of surgical trainees. Br J Surg 2005;92:778–82.
16. Lee G, Lee T, Dexter D, et al. Methodological infrastructure in surgical ergonomics: a review of tasks, models, and measurement systems. Surg Innov 2007;14:153–67.
17. Birkfellner W, Watzinger F, Wanschitz F, et al. Systematic distortions in magnetic position digitizers. Med Phys 1998;25:2242.
18. Hummel JB, Bax MR, Figl ML, et al. Design and application of an assessment protocol for electromagnetic tracking systems. Med Phys 2005;32:2371.
19. Magnetic fact sheet and video: DC vs AC tracking. Available at: http://www.ascension-tech.com/docs/ASCWhitePaperDCvAC.pdf. Accessed June 14, 2011.
20. Rittenhouse N. Development and validation of an assessment model for open surgical procedures. Available at: https://tspace.library.utoronto.ca/handle/1807/25905. Accessed October 23, 2011.

21. Forsyth DA, Ponce J. Computer vision: a modern approach. New York: Prentice Hall; 2002.
22. Unnithan VB, Dowling JJ, Frost G, Bar-Or O. Role of mechanical power estimates in the O2 cost of walking in children with cerebral palsy. Med Sci Sports Exerc 1999;31:1703–8.
23. Nadeau S, McFadyen BJ, Malouin F. Frontal and sagittal plane analyses of the stair climbing task in healthy adults aged over 40 years: what are the challenges compared to level walking? Clin Biomech 2003;18:950–9.
24. Morea C, Langlotz F, Scheer C, et al. Development of an optoelectronic positioning device for serial direct digital images of oral structures. J Periodont Res 2000;35:225–31.
25. Svoboda T, Martinec D, Pajdla T. A convenient multi-camera self-calibration for virtual environments. Presence Teleoperators Virt Environ 2005;14:407–22.
26. Aggarwal R, Grantcharov T, Moorthy K, et al. An evaluation of the feasibility, validity, and reliability of laparoscopic skills assessment in the operating room. Ann Surg 2007;245:992–9.
27. Bann SD, Khan MS, Darzi AW. Measurement of surgical dexterity using motion analysis of simple bench tasks. World J Surg 2003;27:390–4.
28. Brydges R, Classen R, Larmer J, Xeroulis G, Dubrowski A. Computer-assisted assessment of one-handed knot tying skills performed within various contexts: a construct validity study. Am J Surg 2006;192:109–13.
29. Hayter MA, Friedman Z, Bould MD, et al. Validation of the Imperial College Surgical Assessment Device (ICSAD) for labour epidural placement. Can J Anesth 2009;56:419–26.
30. Dunkin B, Adrales GL, Apelgren K, Mellinger JD. Surgical simulation: a current review. Surg Endosc 2007;21:357–66.