

Collaborative eye tracking: a potential training tool in laparoscopic surgery

Andrew S. A. Chetwood · Ka-Wai Kwok ·
Loi-Wah Sun · George P. Mylonas ·
James Clark · Ara Darzi · Guang-Zhong Yang

Received: 24 August 2011 / Accepted: 15 December 2011 / Published online: 19 January 2012
© Springer Science+Business Media, LLC 2012

Abstract

Background Eye-tracking technology has been shown to improve trainee performance in the aircraft industry, radiology, and surgery. The ability to track the point-of-regard of a supervisor and reflect this onto a subjects' laparoscopic screen to aid instruction of a simulated task is attractive, in particular when considering the multilingual make up of modern surgical teams and the development of collaborative surgical techniques. We tried to develop a bespoke interface to project a supervisors' point-of-regard onto a subjects' laparoscopic screen and to investigate whether using the supervisor's eye-gaze could be used as a tool to aid the identification of a target during a surgical-simulated task.

Methods We developed software to project a supervisors' point-of-regard onto a subjects' screen whilst undertaking surgically related laparoscopic tasks. Twenty-eight subjects with varying levels of operative experience and proficiency in English undertook a series of surgically minded laparoscopic tasks. Subjects were instructed with verbal queues (V), a cursor reflecting supervisor's eye-gaze (E), or both (VE). Performance metrics included time to complete tasks, eye-gaze latency, and number of errors.

Results Completion times and number of errors were significantly reduced when eye-gaze instruction was employed (VE, E). In addition, the time taken for the subject to correctly focus on the target (latency) was significantly reduced.

Conclusions We have successfully demonstrated the effectiveness of a novel framework to enable a supervisor eye-gaze to be projected onto a trainee's laparoscopic screen. Furthermore, we have shown that utilizing eye-tracking technology to provide visual instruction improves completion times and reduces errors in a simulated environment. Although this technology requires significant development, the potential applications are wide-ranging.

Keywords Human/Robotic · Computing · Surgical · Training/Courses · Eye-tracking

Minimally invasive surgery (MIS) forms a significant proportion of operative procedures with well-recognized advantages that include reduced in-patient stay, reduced pain postoperatively, and reduced incision size, and morbidity [4]. However, reduced depth perception, two-dimensional (2D) representation of a three-dimensional (3D) field, and the "fulcrum effect" are some of the factors that make laparoscopic surgery more challenging for the trainee [5]. As a result, methods to improve training in MIS have been the subject of a great deal of research, with the development of new simulators and assessment tools [9]. With advances in minimally invasive surgery and medical robotics, additional techniques/devices that may assist the surgeon are an exciting area of research.

Eye tracking was first introduced more than 100 years ago for use in reading research [10]. Eye-tracking technology has been shown to improve the performance of trainees in a number of different environments. This utilizes the ability to display a supervisor's point-of-regard to a student as a training tool. In the aircraft inspection industry, which has an inherently high level of safety, it has been shown that scan-path-based feed-forward training of

A. S. A. Chetwood · K.-W. Kwok · L.-W. Sun ·
G. P. Mylonas · J. Clark · A. Darzi · G.-Z. Yang
Hamlyn Centre for Robotic Surgery, Institute of Global Health
Innovation, Imperial College, London, UK

A. S. A. Chetwood (✉)
Division of Surgery, Imperial College, London, UK
e-mail: Andrewchetwood@doctors.org.uk

novices by experts in the field leads to improved accuracy in the simulator environment [12]. Different search strategies have been identified between novice and experienced radiologists. Furthermore, eye-tracking feedback has been shown to improve performance of trainee radiologists when reviewing images compared with those for whom no feedback was provided [6].

A frustration during the acquisition of laparoscopic skills in the operating room can be the uncertainty as to exactly where the supervising surgeon wants the trainee to target. An example would be the precise placement of a surgical clip or the best position to retract tissue to achieve optimal tension for dissection. With the supervisor often wearing a surgical gown and gloves and unable to indicate on the monitor, confusion and delays can ensue. There is currently limited published work on the potential use of collaborative gaze-contingent eye tracking as a training tool in the surgical field. The possibility of a supervising surgeon being able to indicate to his trainee exactly where he/she should be focusing could have a significant impact on the performance of a collaborative task and the efficiency of the overall workflow. With the recent advances in the dual console DaVinci systems and remote collaborative surgery, the demand for accurately depicting the point-of-regard on the surgical display is increasing [17]. This may represent a novel use for gaze-contingent technology and enhance the communication modalities available to surgeons of all grades.

The purpose of this article was to develop an interface to project a supervisors' point-of-regard onto a subjects'

laparoscopic screen. We investigated whether a supervisor's eye-gaze could be used to improve subject accuracy during a surgical-simulated task. Furthermore, we investigated whether the use of collaborative eye-tracking could help to reduce mistakes and overcome language barriers in the operating environment.

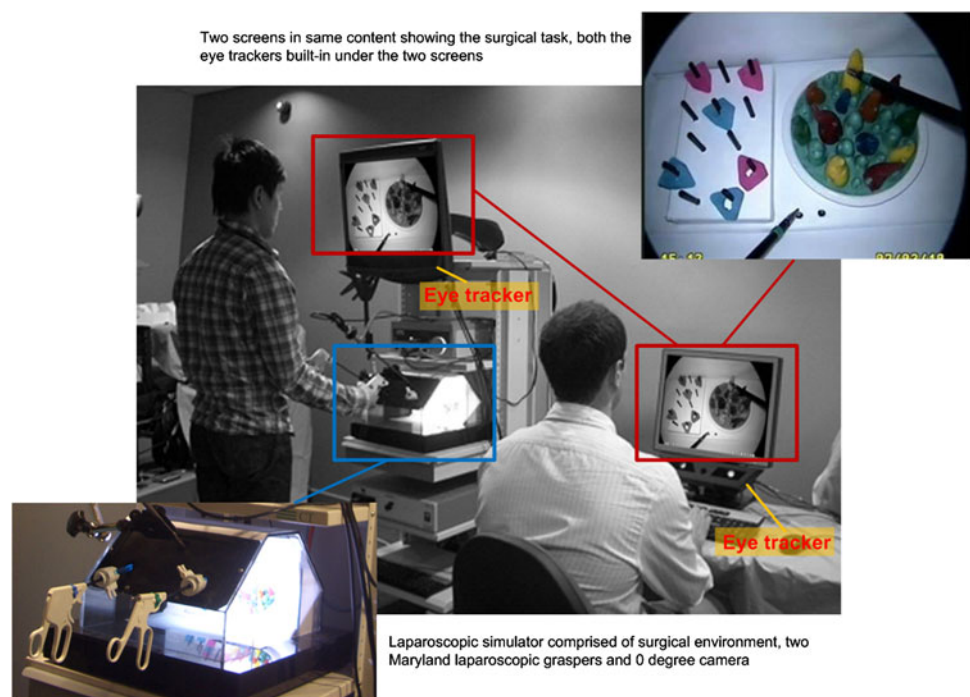
Methods

Experimental set-up

A laparoscopic simulator using a 0° camera, light source, and mobile display screen (Storz) was linked to a 17' monitor with inbuilt eye-tracking apparatus (Tobii Technologies AB, Sweden) with fixation sample rates of up to 50 Hz, 0.5° of visual angle accuracy, and working volume $30 \times 16 \times 20 \text{ cm}^3$. The image from the laparoscopic camera was displayed to the subject monitor and relayed to an additional monitor, also with eye-tracking apparatus, for the supervisor. A bespoke interface was developed whereby the point-of-regard of the supervisor, which was recorded by the eye-tracking apparatus, could be displayed real-time onto the monitor of the subject as a superimposed cursor (Fig. 1).

The surgical tasks involved subjects identifying ten different objects in an environment containing various shapes of different sizes and color, modified from the fundamentals of laparoscopic surgery training course (University of Washington) and DaVinci robotic training

Fig. 1 Experimental setup shows the simulated environment, subject, and supervisor



equipment. The subject had the use of two Maryland laparoscopic graspers and was able to use either hand freely. The camera was white balanced, focused, and in a fixed position with all of the objects clearly visible.

Twenty-eight subjects (23 females and 5 males) of varying nationalities and ages, comprising 7 surgeons and 21 non-clinicians, with laparoscopic experience ranging from nil to consultant and a variety of first languages were invited to take part in the study. After informed, written consent, the subjects were taken through the experiment and given a set time to familiarize themselves with the setup. Subjects underwent a nine-point calibration so that accurate subject-specific data for point-of-regard could be recorded. The subjects were instructed as to which object to target with spoken verbal instruction from the supervisor (V), eye cursor guidance as the supervisors' point of regard superimposed onto the monitor (E), or both (VE). The verbal instructions were worded in a simple, nonambiguous way, and the tasks were performed in a random order to prevent subjects from predicting the order of tasks. There was a single, English-speaking instructor who spoke clearly and slowly during the tasks. Tasks consisted of identifying ten objects with the Maryland graspers correctly and varied from simple instructions to complicated orders involving the use of color, size, and shape to assist in identifying the correct object, e.g., "Touch the blue object in between the two orange objects." Between tasks, subjects returned their instruments to a set position to standardize completion time and gaze-response measurement. Four variables were recorded and analyzed: (1) completion time; (2) gaze latency; (3) gaze convergence; and (4) number of errors. One-way ANOVA and Paired *t* test were used to compare four variables between groups and methods respectively (SPSS).

Data collection

Data were recorded from the start of instruction until the correct object had been touched with the graspers by activating specific keys recognized by the interface. During the experiment, the 2D gaze positions on the screen were recorded at the rate of 25 Hz to synchronization with the video frame rate. Time to complete the task and errors made also were recorded. In between each group of ten consecutive tasks with the same instruction, subjects were allowed a short break to prevent fatigue. During the data recording phase of the experiment the supervisor focused his gaze on the target throughout. *Off-line* analysis of the 2D screen image was undertaken to give accurate coordinates of the visual target in each subject test, negating the possibility of the supervisor momentarily looking away from the target object during the task.

We calculated the gaze latency (the length of time taken for the subject to focus on the target area from start of instruction) and the gaze convergence (the proportion of time during the task that the subject gaze was focused at the target area). The *off-line* calculation of the target location was used for this analysis. At the end of the experiment, participants completed a short questionnaire where demographics, language, and experimental feedback were obtained.

Results

The data collected were designed to enable accurate calculation of a number of indices that can be used to assess performance indirectly, including task completion time, number of errors (where the incorrect object was identified), gaze latency (length of time for the subject to identify correctly the target from the start of the supervisors instruction), and gaze convergence (how long the subject was correctly focusing on the target during the task). In addition, the subject questionnaire enabled subgroup analysis for English versus non-English as a first language and the collection of qualitative data, which can reflect useful information from subject feedback.

Before formal data analysis, it was evident from observation that subject performance was superior when eye-gaze guidance was utilized. Completion times were faster, less time was spent scanning the screen, and movements of the instruments were more precise with less hesitation. A useful way to depict this qualitative, observational data is via hotspot analysis whereby areas of focused subject attention are shown as bright colors, which can be shown over an image of the simulated surgical environment (Fig. 2). More complex analysis of the subjects' point-of-regard as recorded by the eye-tracker for a single task revealed a more focused, target-specific visual response when eye-guidance (E) was used compared with verbal guidance (V) (Fig. 3).

Verbal versus eye versus verbal + eye

Table 1 summarizes the performance indices for the three methods of instruction (V, E, and VE) and for English (group 1, $n = 14$) versus non-English (group 2, $n = 14$) as first language. Among 28 subjects, significantly increased completion times (6.1 s), gaze latency (2.3 s), gaze convergence (202 pixels), and increased numbers of errors were found when purely verbal (V) instruction was utilized compared with eye-tracking guidance (E + VE) ($p < 0.05$).

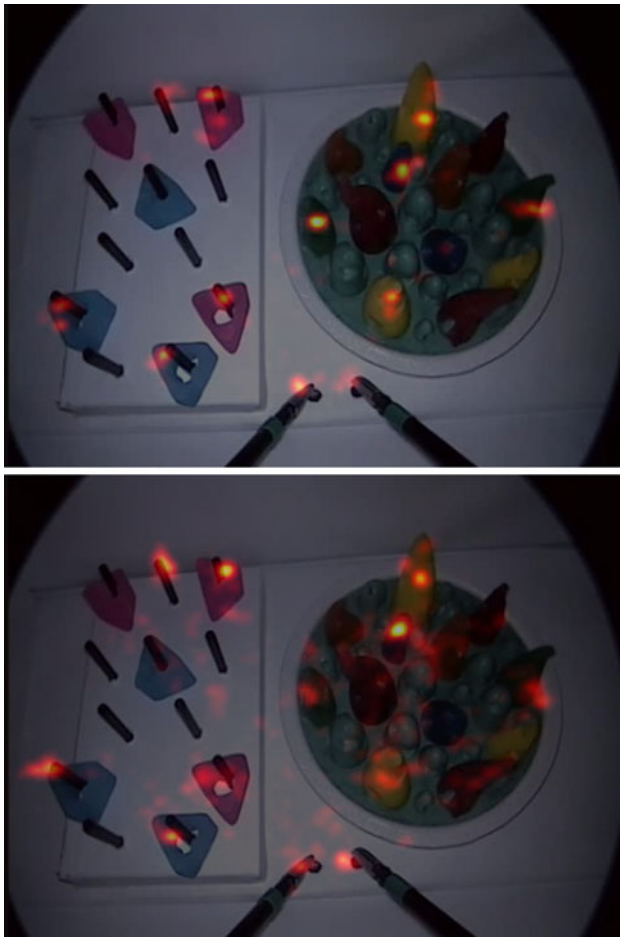


Fig. 2 Pictorial representation of subjects' point-of-regard. More precise areas of focus ("hotspots") are visible in the top image which represents (verbal plus eye) VE guidance. By comparison, the verbal guidance shows numerous areas of focus away from the target image reflecting scanning of the object field in an attempt to follow the spoken instruction (Color figure online)

Language: English versus non-English as first language

Table 2 shows the one-way ANOVA result between groups (English vs. non-English). Significant difference occurred between two groups on completion time for all three types of instruction (V, E, and VE) with p values of 0, 0.028 and 0.013 respectively. Verbal guidance resulted in shorter completion and convergence times in the English-speaking group (1) compared with the non-English group (2) ($p < 0.05$). However, there was no significant difference identified in completion times between E and VE instruction for the two groups (English vs. non-English).

Completion times

Overall completion times were significantly reduced when eye gaze instruction was used (E and VE) compared to when verbal instruction were given alone ($p < 0.05$; Fig. 4).

Gaze latency

The time period for the subject to correctly focus their gaze onto the target was significantly reduced when eye gaze instruction was used (Fig. 4).

Errors

There were no errors recorded when eye-gaze instruction was utilized (E and VE). However, multiple mistakes were encountered when instructions came from verbal guidance alone.

Gaze convergence

This reflects the proportion of the duration of the task where the subjects' point-of-regard was focused at the correct target. This was reduced when eye gaze instruction was employed.

Subject questionnaire

A short participant questionnaire was completed after the experiment. A Likert scale was used: 0 = strongly disagreed to 5 = strongly agreed. The subjects strongly supported the use of the eye-tracking guidance and reported that they found it more effective (mean = 4.3), convenient (4.8), intuitive (4.7), and practical (4.8). Fifteen participants preferred the use of eye-tracking guidance alone, whereas 13 found the combined verbal plus visual guidance the most effective.

Discussion

This article describes the development of a novel framework to enable a supervisor's point-of-regard to be shown on a subjects' laparoscopic screen. We have shown that using this method of visual guidance accuracy is improved and mistakes are reduced within the limits of a simulated surgical environment.

Eye movements appear to move from fixation to fixation via saccades, which are quick eye movement occurring in between. The recording and analysis of the so-called "scan-path" may provide information on the underlying cognitive function of an individual and has led to numerous research avenues and potential commercial applications, including advertising and website design [2]. The aircraft industry has shown that the projection of an experienced cargo hold inspector's gaze onto a trainee's field of vision improved the performance and the accuracy of the trainee [13]. Thus far, the medical applications have been primarily explored in radiology where search strategies of

Fig. 3 A single task, “Touch the blue object inbetween the two orange objects” (shown with red cross) is analyzed to show how the subject’s eye gaze around the target varies depending on type of instruction seen (eye guidance top row; verbal guidance bottom row). Furthermore, division between languages has been performed. A more focused point-of-regard is pictorially demonstrated when using eye-guidance is visible

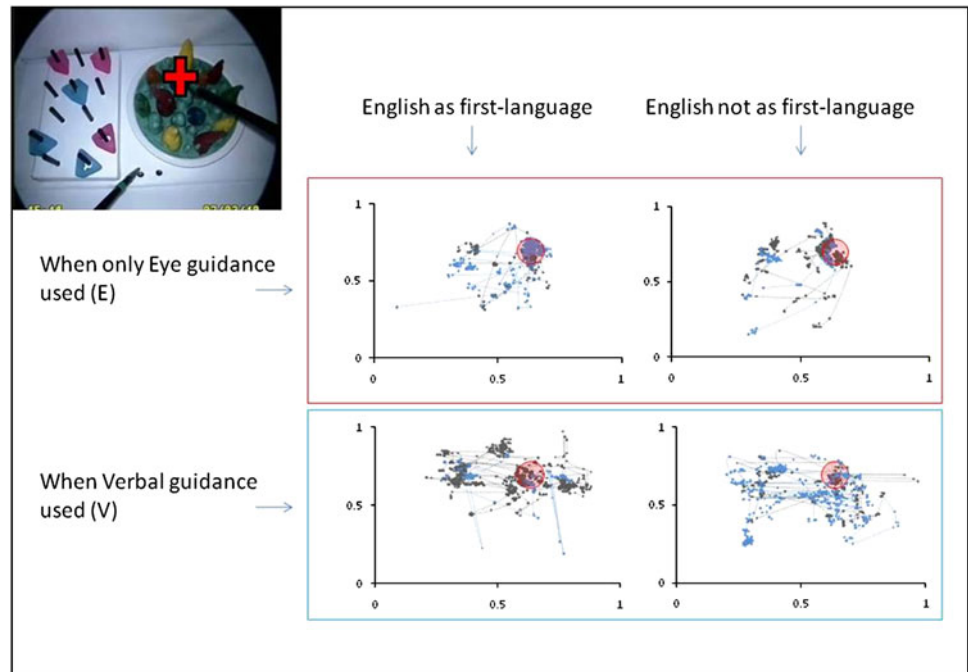


Table 1 Summary of values of four parameters for the three methods and two groups

	Verbal						Eye						Verbal + Eye					
	1		2		Total		1		2		Total		1		2		Total	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Completion time (sec)	5.3	0.7	6.8	1.2	6.1	1.2	2.5	0.4	2.9	1	2.7	0.8	3	0.7	3.8	1	3.4	1
Gaze latency (sec)	2.1	0.6	2.4	0.8	2.3	0.7	0.6	0.2	0.7	0.3	0.6	0.2	0.8	0.2	0.8	0.2	0.8	0.2
Gaze convergence (pixel)	175.3	31.2	228.7	67.8	202	58.5	43.4	16.9	56	25.2	49.7	22	58.4	23.5	76	35.7	67.1	31
No. of failures	0.5	0.8	1.1	1.3	0.8	1.1	0	0	0	0	0	0	0	0	0	0	0	0

Sample sizes for groups 1 and 2 and total are 14, 14, and 28, respectively

Table 2 One-way ANOVA result between groups

Between groups	*Significant at $\alpha = 0.05$ (two-tailed)		Verbal		Eye		Verbal + eye	
	F	p	F	p	F	p	F	p
Completion time (sec)	18.1	0*	2.3	0.142	5.4	0.028*		
Gaze latency (sec)	1	0.334	1.1	0.308	0.1	0.794		
Gaze convergence (pixel)	7.2	0.013*	2.4	0.133	2.4	0.137		
No. of failures	2.1	0.16	N/A	N/A	N/A	N/A		

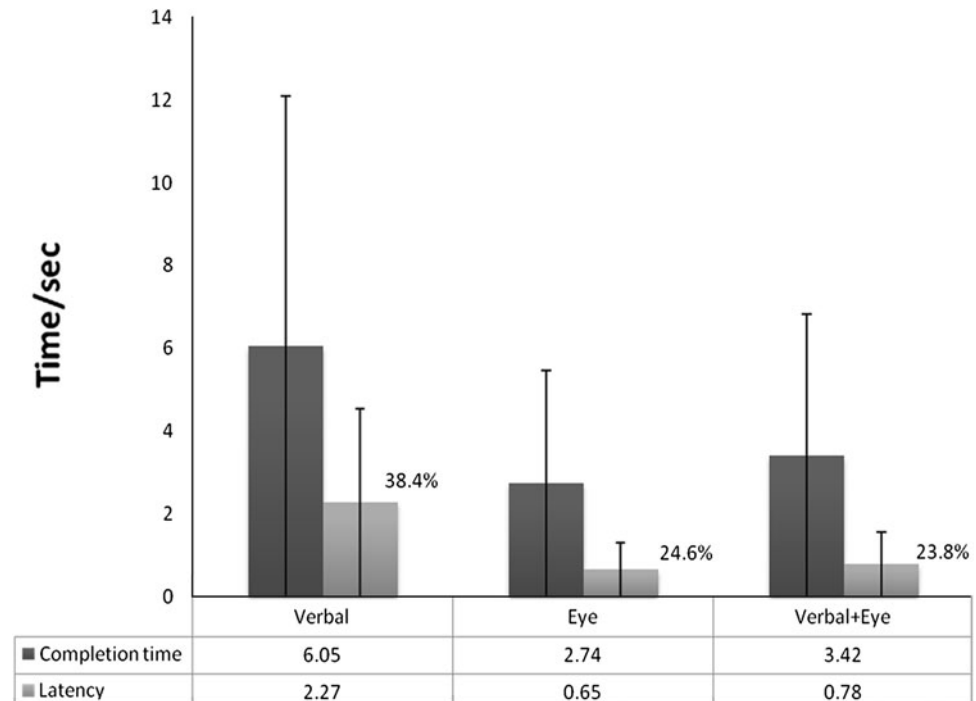
varying levels of radiologist have been studied, and techniques for optimal data presentation and training have been suggested [3, 6].

In a minimally invasive surgical simulator setting, experts have been shown to employ different eye movements and fixation strategies to novices, including less time spent focussing on the tool tip and faster movements. In addition, a recent study has demonstrated the use of eye tracking data to be able to distinguish between expert and nonexpert surgeons. Despite the small sample size the authors suggest this

method may provide a novel technique for assessing surgical skill [11]. Further work could reveal a role in eye movement behavior analysis for training and assessment in surgeons of differing skill, and there may be a role in the assessment of progression in trainee surgeons [7].

The technical aspects involved in designing a collaborative eye-tracking setup, which could be translated into the clinical environment are challenging. A major success of this study is the development of a robust interface, which enabled accurate and real-time projecting of the supervisors

Fig. 4 Completion times and eye-gaze latency for the different methods of instruction show reduced completion times and gaze latency times when eye-gaze guidance is utilized



gaze onto the subjects' screen. The DaVinci surgical platform (Intuitive Surgical, CA) has a fixed head position of the operator, which is advantageous for eye-tracking. The recent appearance of the dual console DaVinci Si system and the development of a prototype inbuilt eye-tracker brings the prospect of a collaborative surgery closer [17, 16]. Eye tracking has been developed as a possible tool to assist in providing "a third hand" for the surgeon. Early work from our group has demonstrated a novel gaze contingent framework where an additional tool can be controlled directly by the eyes of the surgeon [8]. In the computer gaming industry, eye-gaze has been developed as a possible control device during game-play and has enabled participants to feel more immersed in the virtual reality environment [15]. Experimental work integrating an eye-tracker with a laparoscopic camera to permit autonomous camera control also has been developed [1]. This enabled the surgeons' point-of-regard to be kept in the center of the screen. Eye-tracking in this collaborative surgical environment could aid communication between surgeons, improve training, and assist in operating surgical instruments.

Participant feedback presented from this study suggested that those subjects whose first language was not English found the eye-guidance particularly useful, although we do not have quantitative data to support this. However, results from language subgroup analysis (Table 2) suggest that eye-gaze instruction may assist in overcoming language barriers in this simulated task. Modern clinical teams are comprised of a variety of nationalities, and in the surgical environment where masks

prevent lip reading problems as a result of language barriers can occur. Collaboration between clinicians from different countries and the possibility of remotely operated collaborative surgery will bring these issues to the fore.

This study provides evidence of the potential for a novel way to utilize eye-tracking technology in a training role in laparoscopic surgery. However, it is limited in that we have not shown clear evidence of improved *performance*. Quantitative demonstrating of superior performance in a surgical setting is challenging. Multiple performance indices exist in an attempt to quantify "better performance." Time to complete a task is frequently used and in combination with error analysis may help to better reflect accuracy or efficiency when performing a simulated task, which time alone does not. Eye-gaze measures, including point-of-regard analysis and comparison to instrument position and/or gaze of an expert, have been used where authors demonstrated significant differences in gaze analysis between novices and expert in a laparoscopic training environment [11, 18]. Instrument tip position can be measured in a virtual trainer, but in a real-time laparoscopic trainer requires more complex tracking methods. Such devices, where instruments are connected to sensors and provide information about the position of the tip in X, Y, and Z axes, have been developed and could provide a future testing platform for this interface [14].

The principles of this study could be extended to other clinical specialities where doctors are using an image to guide a procedure. Examples could be for a consultant cardiologist to guide a trainee during angiography. Orthopedic

surgeons may be able to aid a trainee by glancing at the image-intensifier screen, as opposed to having a walk over and point. The potential implications are wide-ranging and truly multidisciplinary. The design of a truly user-friendly eye-tracker that could tolerate the rigors of an operating room environment is required. This step would enable true translational research into the clinical applications of eye tracking and the development as an additional tool to the surgeon with the goal to take this technology into a real-life surgical setting and to assess whether surgical performance is improved.

Conclusions

This article represents the development of a novel eye-tracking interface and its use to improve the accuracy and reduce errors of subjects performing surgical-simulated tasks. The possible clinical applications of this technology are wide-ranging and could improve efficiency and safety in the surgical setting.

Disclosures A. Chetwood, J. Clark, K.W. Kwok, L.W. Sun, A. Darzi, and G.Z. Yang have no conflicts of interest or financial ties to disclose.

References

1. Ali SM, Reisner LA, King B, Cao A, Auner G, Klein M, Pandya AK (2008) Eye gaze tracking for endoscopic camera positioning: an application of a hardware/software interface developed to automate Aesop. *Stud Health Technol Inform* 132:4–7
2. Ballard DH (1992) Hand-eye coordination during sequential tasks. *Philosophical transactions—royal society. Biol Sci* 337:331
3. Ellis SM, Hu X, Dempere-Marco L, Yang GZ, Wells AU, Hansell DM (2006) CT of the lungs: eye-tracking analysis of the visual approach to reading tiled and stacked display formats. *Eur J Radiol* 59(2):257–264
4. Fuchs KH (2002) Minimally invasive surgery. *Endoscopy* 34:154–159
5. Hogle NJ, Chang L, Strong VE, Welcome AO, Sinaan M, Bailey R, Fowler DL (2009) Validation of laparoscopic surgical skills training outside the operating room: a long road. *Surg Endosc* 23:1476–1482
6. Kundel HL, Nodine CF, Krupinski EA (1990) Computer-displayed eye position as a visual aid to pulmonary nodule interpretation. *Invest Radiol* 25:890–896
7. Law B, Atkins MS, Kirkpatrick AE, Lomax AJ (2004) Eye gaze patterns differentiate novice and experts in a virtual laparoscopic surgery training environment. In: proceedings of the 2004 symposium on eye tracking research & applications. ACM, San Antonio, Texas
8. Noonan DP, Mylonas GP, Darzi A, Yang GZ (2008) Gaze contingent articulated robot control for robot assisted minimally invasive surgery. In: 2008 IEEE/RSJ international conference on robots and intelligent systems, vol 1–3. Pasadena, pp 1186–1191
9. Peters JH, Fried GM, Swanson LL, Soper NJ, Sillins LF, Schirmer B, Hoffman K (2004) Development and validation of a comprehensive program of education and assessment of the basic fundamentals of laparoscopic surgery. *Surgery* 135:21–27
10. Rayner K, Pollatsek A (1989) The psychology of reading. Prentice Hall, Englewood Cliffs, p 529
11. Richstone I, Schwartz M, Seideman C, Caddeu J, Marshall S, Kavoussi LR (2010) Eye metrics as an objective assessment of surgical skill. *Ann Surg* 252:177–182
12. Sadasivan S (2004) Use of eye movements as feedforward training for a synthetic aircraft inspection task. Clemson University, Clemson
13. Sadasivan S, Greenstein J, Gramopadhye A, Duchowski T (2005) Use of eye movements as feedforward training for a synthetic aircraft inspection task. In: proceedings of the SIGCHI conference on human factors in computing systems. ACM, New York, pp 141–149
14. Smith CD, Farrell TM, McNatt SS, Metreveli RE (2001) Assessing laparoscopic manipulative skills. *Am J Surg* 181:547–550
15. Smith JD, Graham TCN (2006) Use of eye movements for video game control. In: proceedings of the 2006 ACM SIGCHI international conference on advances in computer entertainment technology. ACM, Hollywood
16. Stoyanov D, Mylonas GP, Yang GZ (2008) Gaze-contingent 3D control for focused energy ablation in robotic-assisted surgery. In: Miccai proceedings of medical image computing and computer-assisted intervention, vol 5242. New York 347–355
17. The DaVinci Si HD surgical system, Intuitive, CA
18. Wilson M, McGrath J, Vine S, Brewer J, Defriend D, Masters R (2010) Psychomotor control in a virtual laparoscopic surgery training environment: gaze control parameters differentiate novices from experts. *Surg Endosc* 24(10):2458–2464