Surgical Endoscopy
https://doi.org/10.1007/s00464-019-07141-x

Virtual pointer for gaze guidance in laparoscopic surgery

Yuanyuan Feng[1] · Hannah McGowan[1] · Azin Semsar[1] · Hamid R. Zahiri[2] · Ivan M. George[3] · Adrian Park[2] · Andrea Kleinsmith[1] · Helena Mentis[1]

Received: 20 March 2019 / Accepted: 20 September 2019

Abstract

Background A challenge of laparoscopic surgery is learning how to interpret the indirect view of the operative field. Acquiring professional vision—understanding what to see and which information to attend to, is thereby an essential part of laparoscopic training and one in which trainers exert great effort to convey. We designed a virtual pointer (VP) that enables experts to point or draw free-hand sketches over an intraoperative laparoscopic video for a novice to see. This study aimed to investigate the efficacy of the virtual pointer in guiding novices' gaze patterns.

Methods We conducted a counter-balanced, within-subject trial to compare the novices' gaze behaviors in laparoscopic training with the virtual pointer compared to a standard training condition, i.e., verbal instruction with un-mediated gestures. In the study, seven trainees performed four simulated laparoscopic tasks guided by an experienced surgeon as the trainer. A Tobii Pro X3-120 eye-tracker was used to capture the trainees' eye movements. The measures include fixation rate, i.e., the frequency of trainees' fixations, saccade amplitude, and fixation concentration, i.e., the closeness of trainees' fixations.

Results No significant difference in fixation rate or saccade amplitude was found between the virtual pointer condition and the standard condition. In the virtual pointer condition, trainees' fixations were more concentrated ($p = 0.039$) and longer fixations were more clustered, compared to the Standard condition ($p = 0.008$).

Conclusions The virtual pointer effectively improved surgical trainees' in-the-moment gaze focus during the laparoscopic training by reducing their gaze dispersion and concentrating their attention on the anatomical target. These results suggest that technologies which support gaze training should be expert-driven and intraoperative to efficiently modify novices' gaze behaviors.

Keywords Laparoscopic training · Gaze guidance · Intraoperative video annotation · Eye tracking · Fixation concentration

Minimally invasive surgery (MIS), which minimizes post-operative complications, blood loss, and recovery time, has prompted a significant paradigm shift in the advancement of surgery [1]. However, learning how to perform laparoscopic surgeries is extremely challenging, as it relies not only on the novices' attainment of fundamental knowledge and proficiency in technical skills but also on the novices' ability to perceive and use the in situ knowledge that is conveyed by the trainers during an operation [2, 3].

A major component of this in situ knowledge is professional vision—knowing where to see, how to see it, and how to incorporate it into actions [2, 4, 5]. It has been shown that guiding novices' gaze to a desired target has become the main focus in laparoscopic training [7]. Yet, such guidance requires great effort for the trainers, requiring them to gesture or point over the monitor while providing detailed verbal explanations to reveal the subtle changes in the video [2, 6]. The surgical instruments are also used to point at a target to elucidate the structures embedded in the tissues. However, these gestures are often implicit, which requires novices to parse, envision, and make sense of experts' behaviors to perceive the locations and directions [2, 6]. For instance, in a laparoscopic cholecystectomy case, the expert surgeon drew an "imaginary

✉ Yuanyuan Feng
fengy1@umbc.edu

1 Department of Information Systems, University of Maryland, Baltimore County, Baltimore, MD, USA

2 Department of Surgery, Anna Arundel Medical Center, Annapolis, MD, USA

3 Department of Surgery, Johns Hopkins Medicine, Baltimore, MD, USA

Published online: 04 October 2019

line" with a grasper. This gesture led to a prolonged conversation, which engendered a series of checking, clarifying, and aligning of the location of the line [2].

To facilitate the conveyance of professional vision, we have designed a low-cost and easy-to-deploy virtual pointer (VP) that enables experts to point or draw free-hand sketches over an intraoperative laparoscopic video for a novice to see [8]. In this, the VP transforms the implicit guidance into explicit and salient visual cues. In our previous study, we have shown that there was a negligible impact of the VP on the task completion time, yet a significant reduction in the novices' number of movements [8]. Yet, it is unclear how the VP affects novices' gaze behaviors. This understanding is important to assess the efficacy of the VP in supporting laparoscopic training, and more broadly, in identifying design directions for future technologies in surgical education.

Previous research has shown that novices and experts have different gaze patterns [9–11]. Experts usually have longer fixations and concentrate their gaze on surgical objects, while novices more frequently visually trace the movements of their instruments [9, 10]. Surgical performance is often affected by how seeing is elicited and the objects on which gaze is fixated [11, 12]. Expert gaze patterns have been demonstrated to be associated with faster task completion, fewer movements, and shorter tool paths [12]. To identify the benefits of gaze training, Wilson et al. asked novices to complete an eye-hand coordination task in a surgical simulator. One-third of them watched experts' gaze behaviors on offline videos of the task before practice. They found that the gaze training led to faster task completion times [11]. In a similar experiment, Vine et al. demonstrated that gaze training significantly improved novices' skill retention and transfer to more complex tasks [13]. A real-time gaze pointer was introduced by Chestwood et al. for laparoscopic training. They superimposed experts' gaze onto the novices' monitor during an eye-hand coordination task and found that the real-time gaze pointer improved task performance [14]. These gaze training studies explain the benefits of gaze training on task performance. Yet, it is unknown how novices use these professional vision cues to guide their focus during a surgical task.

In this study, we aim to investigate the efficacy of the virtual pointer in guiding novices' gaze patterns. The virtual pointer is essentially different from the gaze pointer [14]. The VP is controlled by the experts, meaning that the experts are actively involved in determining what information is important to share at which moment and how to present this information. In this way, the VP leverages on the experts' knowledge and experience in laparoscopic training, as well as the pairs' common ground during the task. In comparison, the gaze pointer automatically presents the experts' gaze movements to the novices without keeping the experts in the loop of the guidance.

We hypothesize that with the use of VP, novices' gaze behaviors will more approach that to the experts, i.e., they will demonstrate fewer fixations and concentrate their fixations more on surgical objects.

Methods

Virtual pointer

The virtual pointer is a video referencing system, using touchless natural user interfaces based on the Microsoft Kinect sensor (Microsoft Corporation, USA) [8]. Real-time image manipulation, such as pointing and free-hand drawing, can be performed by a surgeon using a combination of audio key words and hand movements to facilitate the conveyance of expert gaze to the trainees (Fig. 1). Specifically, the verbal command of "Kinect ready" awakens the Kinect and allows it to start detecting movement of the trainer. The trainer can then use the verbal commands "Kinect point" or "Kinect draw" to engage either the pointing or drawing tool, respectively. To use the pointing mode, the trainer controls the Virtual Pointer (displayed as a green dot) using an open hand. To engage the drawing mode, the trainer closes their hand to draw and opens it when finished drawing. To clear any drawings, the verbal command "Kinect clear" must be spoken and "Kinect end" is spoken when finished to close the application.

The system, developed in C Sharp for the Windows Operating System, allows for the application to be overlaid to a laparoscopic video captured directly from the laparoscopic camera. In addition, the visual feedback of voice commands and body movements timely communicates the system

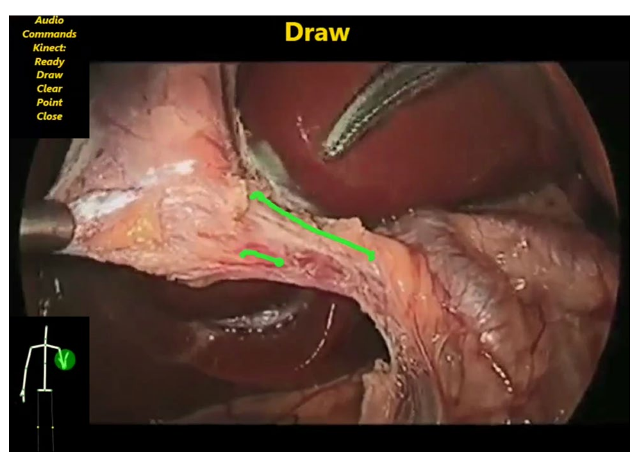


Fig. 1 Virtual Pointer user interface with list of verbal command (top left), current mode (top center), gesture recognition feedback window (bottom left), drawing (green lines on anatomy), and pointer (a green dot; not shown) (Color figure online)

response to the interaction to minimize the learning curve in using the system.

Experimental design and procedure

The experimental design is a 2×4 (training conditions and tasks) counter-balanced, within-subject design. We performed a controlled experiment with two training conditions—a standard training condition as the control and a virtual pointer-supplemented (VP) training condition as the intervention. In the Standard condition, trainer instruction was conducted as it would be normally, through verbal instructions with un-mediated gestures. In the VP condition, the virtual pointer was used by the trainers as an addition to standard guidance to facilitate instruction. The trainers determined when and what VP instructions should be given based on their ongoing assessment of the trainees' knowledge of the task and understanding of the presented instructions. The trainees worked on four simulated laparoscopic tasks under the trainers' guidance. Two tasks were in the VP condition and two were in the Standard condition. The tasks were selected based on a hierarchical task analysis of the laparoscopic cholecystectomy procedure and confirmed by an attending surgeon that they were of similar difficulty levels and required both skills of anatomical structure identification and instrument manipulation. The tasks were performed on validated laparoscopic training physical models [15], including (1) mobilizing the cystic duct and the cystic artery, (2) clipping the cystic duct, (3) clipping the cystic artery, and (4) cutting the cystic artery and the cystic duct (Fig. 2). The orders of the training conditions and the tasks are counter-balanced by constructing a Latin square [16]. The study was approved by the University of Maryland, Baltimore County institutional review board. Informed consent was obtained from all participants before their participation.

Participants

The participants were recruited from the Department of General Surgery in the Anne Arundel Medical Center, Annapolis, MD. A total of 7 surgical trainees, including 1 surgical fellow, 1 research fellow, and 5 surgical residents (3 PGY-1 and 2 PGY-2) were recruited. One attending surgeon and one surgical fellow were recruited as the trainers. The attending surgeon guided the surgical fellow in performing the tasks and the surgical fellow guided the rest of the trainees.

Study setting

The experiment took place on a Park Trainer (Stryker Corporation, USA), which housed the physical anatomical models for the simulated laparoscopic tasks (Fig. 3). The Microsoft Kinect sensor was set to the left of the Park Trainer. This was determined to be the best location for this study as the tasks used called for the trainer to stand to the left of the trainee using their right hand to manipulate the laparoscopic camera. The Windows laptop computer running the Virtual Pointer application was then placed on a table to the right of the Park Trainer.

A Tobii Pro X3-120 eye-tracker was used to capture the trainees' eye movements on the laparoscopic video. The eye-tracker was mounted beneath the trainees' laparoscopic monitor (Fig. 3). Eye calibration was performed at the beginning of each task for each trainee to ensure the eye-tracker would correctly capture trainees' eye movements. The laparoscopic

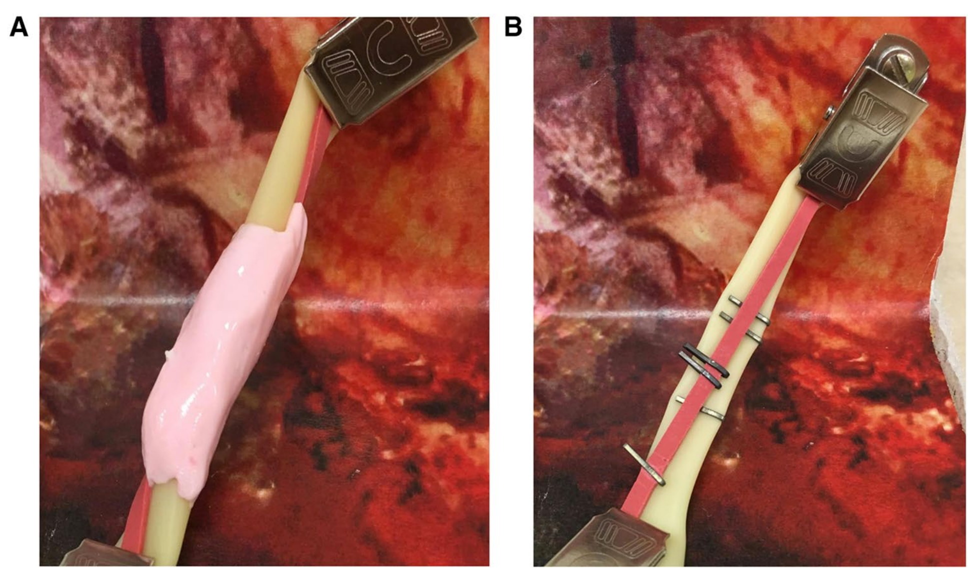


Fig. 2 Task models used in the study: **A** was used for task 1—mobilizing the cystic duct and artery; **B** without the staples was used for task 2—clipping the cystic duct, and task 3—clipping the cystic artery; **B** with staples was used for task 4—cutting the cystic artery and duct

Fig. 3 Study setting with the virtual pointer system, Tobii Pro X3-120 eye-tracker and park trainer

video accompanied with the trainees' eye movements was captured in Tobii Pro Studio (Tobii Technology, Sweden).

Gaze behavior measures

According to previous studies on eye metrics [10, 17, 18, 20], three categories of gaze measures were selected for the comparison, including fixation rate, saccade amplitude, and fixation concentration. Fixation rate was measured to reflect the efficiency of gaze strategy. It was defined as "the number of fixations made per second per trial" [10]. Experts have been found to have fewer and longer fixations than novices [19]. Saccade amplitude, i.e., the velocity of eye movements [20], was measured to reflect the visual search strategy. Experts are reported to have a short amplitude in reading the medical images [19]. Fixation concentration was measured to reflect the focus of the novice [14]. Fixation concentration was first visually compared with gaze heat maps, which were generated in the Tobii Pro Studio (Tobii Technology, Sweden), and was then quantitatively evaluated by the standard deviation of the Euclidean distance between the fixation position and the origin point [18]. It has been shown that experts tend to concentrate their gaze at the target, while novices' gaze behavior is more dispersed [9, 10]. The Tobii Pro Studio used the Velocity-Threshold Identification (I-VT) fixation classification algorithm to identify the fixation [21]. I-VT classifies the eye movements into either fixations or saccades based on the velocity of directional shifts of the eye [22]. The fixation is defined as the eye-movement data point that is below the velocity threshold. In this study, we used the validated velocity threshold of 30°/s [21].

Table 1 Linear mixed models for fixation rate, fixation duration, saccade amplitude, and fixation concentration with training conditions as the fixed effect and the tasks, task orders, and participants as the random effects

Mixed model	Training conditions				
	Est. coefficient	Std. error	*df*	*t* value	*p* value
Fixation rate	0.502	0.385	17.385	−0.630	0.269
Saccade amplitude	0.4189	0.210	17.209	2.00	0.062
Fixation concentration (all fixations)	−15.250	6.902	20.000	−2.209	0.039
Fixation concentration (ten longest fixations)	−23.358	7.757	17.726	−3.011	0.008

Data analysis

The data analysis focused on the changes of trainees' eye movements between the two training conditions. Since we had a within-subject, counted balanced experimental design with four tasks, we used linear mixed models to control the confounding factors—repeated measure, tasks, and task orders. We first developed the linear mixed models with training conditions and task orders as fixed effects, and participants and tasks as random effects. The results showed that the task and the task orders contributed little to the variations of the dependent variables. Thus, we regarded the task orders as a random effect in our final linear mixed models. The Linear mixed models were developed in R version 3.2.2 (R Foundation for Statistical Computing, Vienna, Austria) for the comparison of the three measure categories between the Standard and Virtual Pointer conditions. For all tests, a p value of less than 0.05 was considered statistically significant.

Results

A total number of 28 runs were conducted among the 7 participants, including 14 runs for the Virtual Pointer condition and 14 runs for the Standard condition. In the Virtual Pointer condition, the system was used for an average of 5 times per run, with a standard deviation of 1.75.

There were no significant differences in the fixation rate ($p=0.269$), and saccade amplitude ($p=0.062$) between the standard and virtual pointer conditions (Table 1).

The fixation concentration was first visually evaluated using heatmaps that plotted fixation points over the surgical task field (Fig. 4). In the heatmaps, the color of the highest value was set to be bright red and the lowest dark blue, with corresponding transitional colors between the two extremes. In the fixation count heatmaps, the value equals to the fixation frequency at one spot in one task. In absolute fixation duration, the value equals to the total fixation duration at one spot in one task. The values were scaled for each participant in each task. Compared to the standard condition, in the VP condition, fixation points were more concentrated on the task area (Fig. 4, fixation count), and the longer fixations were more prone to cluster together (Fig. 4, absolute fixation duration).

This trend was confirmed by fitting the fixation concentrations among all fixations in the linear mixed model with the training conditions ($p = 0.039$) as the fixed effect, controlled by the tasks, the task orders, and participants as the random effects (Table 1). Based on the number of bright red clusters on the absolute fixation duration heatmaps, we further evaluated the fixation concentration among the ten longest fixations and found a significant impact of the training conditions ($p = 0.008$) (Table 1). Figure 5 further shows that the Virtual Pointer concentrated the trainees' fixations with a significant reduction of the standard deviation among the fixations.

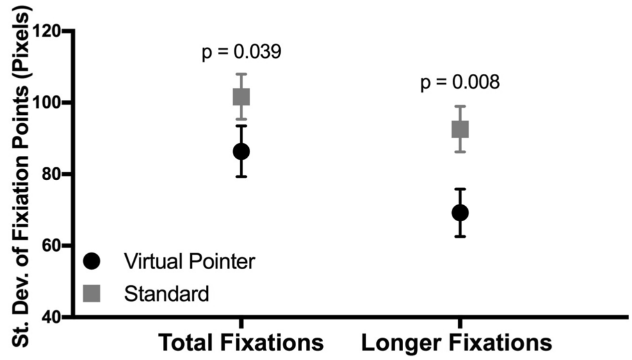
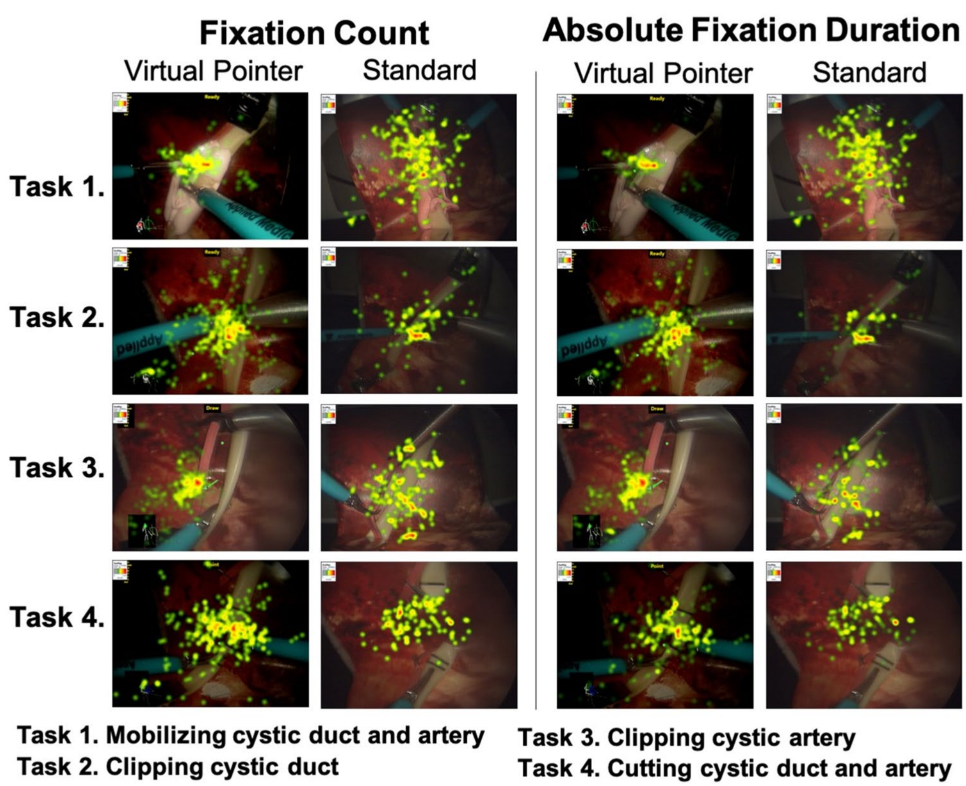


Fig. 5 Fixation concentration for total fixations and the ten longest fixations in the virtual pointer condition and Standard condition

Discussion

Gaze guidance has become a main focus in laparoscopic training because of its necessity in interpreting the operative field [2, 7]. In this study, we present the use of a novel instructional technology, the virtual pointer, to facilitate intraoperative gaze guidance in a laparoscopic training simulation. It is the first approach that is aimed at directly changing novices' gaze behaviors during training. This approach transforms implicit guidance into explicit and salient visual cues using expert-generated annotations on a live

Fig. 4 Examples of heatmaps for fixation count and absolute fixation duration in each task



laparoscopic video. As shown in the results, this technology could substantially improve surgical novices' in-the-moment gaze focus—reducing the novices' gaze dispersion and concentrating their attention on the target.

Previous approaches in improving novices' gaze behaviors have been attempted through offline videos—the novices were asked to watch and memorize the expert gaze behaviors on an offline laparoscopic video before a task [11, 13]. These studies have shown that the gaze training supported technical skill acquisition and annotations on laparoscopic videos are an effective approach in technical skills acquisition [10]. Yet, the use of offline videos in this approach limits its application to tasks that are replicable for both experts and novices. More importantly, it depends on novices' identification and memorization of expert gaze patterns and lacks the support for in situ learning, where the knowledge is tailored specifically to modify individual novices' gaze patterns [2, 23].

As explained by previous research [2–5], professional vision is far more than just knowing where to look at. It is the in situ knowledge that incorporates when, where, how, and why a gaze should be focused. The design of the VP allows experts to point at a specific target and to draw free-hand sketches. The experts are able to explicitly illustrate their gaze strategies by highlighting the target, as well as reinforce these strategies when necessary. The success of the VP in improving novices' gaze patterns confirms that the use of the VP facilitates the conveyance of professional vision. More importantly, compared to the automatic gaze visualization [14], the VP essentially serves as an instructional source, which the experts can leverage in designing their guidance. The success of using the VP to concentrate the novices' fixations indicates that technologies supporting laparoscopic gaze training should be *expert-driven* and *intraoperative* to efficiently modify novices' gaze behaviors.

Compared to directly visualizing experts' gaze patterns, the use of the VP is at the control of experts—by monitoring the novices' progress of task, experts evaluate when and what cues should be given. The information that the VP provides is tailored to the novices' needs in the problem-solving process, as opposed to requirements or dictations that drive novices away from their current work. Thus, it concentrates novices on solving the problem, as manifested by more fixations, especially longer fixations, on surgical objects.

Moreover, this study highlights the need for supporting intraoperative gaze guidance. As shown in the results, novices' longest fixations were more concentered when exposed to the VP. According to the association between fixation duration and visual scene processing [17], the greater concentration reduces the amount of information that the novices need to process, potentially reducing the errors from extra processing, and improving the efficiency of the task performance. Thus, there may be outcome benefits when improved gaze guidance takes place during an operation.

It is interesting to see the insignificant difference in fixation rate between the VP condition and the Standard condition. This result may be due to two reasons. First, the fixation rate reflects the difficulty of a participant in perceiving the task environment. Due to the similar tasks in our study, participants' fixation rates may not vary significantly. Further, the similar fixation rates between the two conditions indicate that there were no learning effects between tasks. Secondly, we used a general fixation filter, which is based on the velocity of the eye movements in simple searching tasks [24]. Surgical tasks are more complex than simple searching task. Thus, there might be variation in determining fixation in surgery. When we were considering the distances between two fixations for fixation concentration, the effect of the fixation filter threshold was minimized.

Our study found no significant difference in the saccade amplitude between the two training conditions. The saccade amplitude reflects the complexity of the scene—with a more complex scene, the saccade amplitude increases [25]. Yet, in this study, we used the same structures—cystic artery and duct—across the four tasks. Thus, limited changes were found in the saccade amplitude. This result is also due to the design of Virtual Pointer, i.e., an expert-driven approach to selectively highlight the target within the instructions, instead of an exploration of the scene.

It is noteworthy that most of our participants were junior residents, with limited laparoscopic experience. Further evaluation should be taken when applying the results to broader trainees. In addition, the tasks are limited to laparoscopic cholecystectomy, which is a relatively basic procedure. The results may vary when novices are working on a complex procedure that requires advanced laparoscopic skills. Our next step is to further evaluate the efficacy of the Virtual Pointer among surgeons with different levels of expertise in different procedures.

Conclusion

To facilitate intraoperative gaze guidance, we designed a video annotation system, the Virtual Pointer, that enables experts to point or draw free-hand sketches over an intraoperative laparoscopic video for a novice to see. We conducted a controlled study to evaluate the efficacy of the system in guiding novices' gaze patterns. We found that the Virtual Pointer reduces novices' gaze dispersion and concentrates their attention on the target. This shows that an easily deployed virtual pointer could substantially improve surgical trainees' in-the-moment gaze focus. In this, we discuss the design implications for future technologies to support laparoscopic gaze training.

Acknowledgements The authors would like to acknowledge Dr. Timothy Turner for his support in preparing for data collection and the SAIL Center at Anne Arundel Medical Center for providing the equipment and space. This work was sponsored by NSF Grant IIS #1422671 and #1552837.

Funding This study is sponsored by National Science Foundation Grant IIS #1422671 and #1552837.

Compliance with ethical standards

Disclosures Ms. Feng, Ms. McGowan, Ms. Semsar, Dr. Zahiri, Mr. George, Dr. Park, Dr. Kleinsmith, and Dr. Mentis have no conflicts of interest or financial ties to disclose.

References

1. Lengyel J, Morrison C, Sagar PM (2010) Trends towards increased use of the laparoscopic approach in colorectal surgery. Colorectal Dis 12(10):1007–1012
2. Mentis HM, Chellali A, Schwaitzberg S (2014) Learning to see the body: supporting instructional practices in laparoscopic surgical procedures. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM pp 2113–2122
3. Cope AC, Mavroveli S, Bezemer J, Hanna GB, Kneebone R (2015) Making meaning from sensory cues: a qualitative investigation of postgraduate learning in the operating room. Acad Med 90(8):1125–1131
4. Koschmann Timothy, LeBaron Curtis, Goodwin Charles, Feltovich Paul (2011) “Can you see the cystic artery yet?” A simple matter of trust. J Pragmat 43(2):521–541
5. Goodwin C (1994) Professional vision. Am Anthropol 96(3):606–633
6. Feng Y, Mentis HM (2017) Improving common ground development in surgical training through talk and action. In: AMIA annual symposium proceedings. American Medical Informatics Association, vol. 2017, p 696
7. Feng Y, Wong C, Park A, Mentis H (2016) Taxonomy of instructions given to residents in laparoscopic cholecystectomy. Surg Endosc 30(3):1073–1077
8. Feng Y, McGowan H, Semsar A, Zahiri HR, George IM, Turner T, Mentis HM et al (2018) A virtual pointer to support the adoption of professional vision in laparoscopic training. Int J Comput Assist Radiol Surg 13(9):1463–1472
9. 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. pp 41–48
10. 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
11. Wilson MR, Vine SJ, Bright E, Masters RS, Defriend D, McGrath JS (2011) Gaze training enhances laparoscopic technical skill acquisition and multi-tasking performance: a randomized, controlled study. Surg Endosc 25(12):3731–3739
12. Richstone L, Schwartz MJ, Seideman C, Cadeddu J, Marshall S, Kavoussi LR (2010) Eye metrics as an objective assessment of surgical skill. Ann Surg 252(1):177–182
13. Vine SJ, Chaytor RJ, McGrath JS, Masters RS, Wilson MR (2013) Gaze training improves the retention and transfer of laparoscopic technical skills in novices. Surg Endosc 27(9):3205–3213
14. Chetwood AS, Kwok KW, Sun LW, Mylonas GP, Clark J, Darzi A, Yang GZ (2012) Collaborative eye tracking: a potential training tool in laparoscopic surgery. Surg Endosc 26(7):2003–2009
15. Seagull FJ, George I, Ghaderi I, Vaillancourt M, Park A (2009) Surgical abdominal wall (SAW): a novel simulator for training in ventral hernia repair. Surg Innov 16(4):330–336
16. MacFie HJ, Bratchell N, Greenhoff K, Vallis LV (1989) Designs to balance the effect of order of presentation and first-order carry-over effects in hall tests. J Sensory Stud 4(2):129–148
17. Henderson JM (2003) Human gaze control during real-world scene perception. Trends Cognit Sci 7(11):498–504
18. Victor T, Johansson E (2005) Gaze concentration in visual and cognitive tasks: using eye movements to measure driving information loss
19. Wood BP (1999) Visual expertise. Radiology 211(1):1–3
20. Nicolaou M, James A, Darzi A, Yang GZ (2004) A study of saccade transition for attention segregation and task strategy in laparoscopic surgery. In: International conference on medical image computing and computer-assisted intervention. Springer, Berlin pp 97–104
21. Olsen A (2012) The Tobii I-VT fixation filter. Tobii Technology, Danderyd Municipality
22. Komogortsev OV, Gobert DV, Jayarathna S, Koh DH, Gowda SM (2010) Standardization of automated analyses of oculomotor fixation and saccadic behaviors. IEEE Trans Biomed Eng 57(11):2635–2645
23. Lave J, Wenger E (1991) Situated learning: legitimate peripheral participation. Cambridge University Press, Cambridge
24. Sen T, Megaw T (1984) The effects of task variables and prolonged performance on saccadic eye movement parameters. Adv Psychol 22:103–111
25. Pomplun M (2006) Saccadic selectivity in complex visual search displays. Vision Res 46(12):1886–1900