Gaze Characteristics of Video Watching in a Surgical Setting

Bin Zheng* University of Alberta

Xianta Jiang† Simon Fraser University

Roman Bednarik‡ University of Eastern Finland

M. Stella Atkins† Simon Fraser University

ABSTRACT

Observing surgical videos has been integrated intensively into surgical education. We test whether video watching can elicit the same eye motion pattern of an observer as the operator in performing the surgical procedure. While observing the task performed using tools, subjects started to move their eyes off the previous target 0.7 s after the tool but reached the next target 1.5 s before the tool. These two events in observing occurred approximately 0.5 s after times in operating. Fixations performed while observing the video were often shorter than while performing, especially before the tool touched the target. Participants in observing were primarily checking outcomes of tool movement; differing to actively collecting information on the target for guiding tool movement during task operating. This result has indication for developing new strategies of improving education outcomes from video watching.

Keywords: Eye-hand coordination, video watching, proactive eye movement, image-guided surgery, visuomotor integration.

Index Terms: H.5.2 [Information Interfaces and Presentation]: User Interfaces—Evaluation/methodology

1 INTRODUCTION

Learning the skills in image-guided tele-manipulation, such as a laparoscopic procedure, is more complicated than learning skills in the direct performance [1, 2]. Tele-manipulation requires sophisticated eye-hand coordination built among eyes, hands and tools of the performing surgeon [3]. Currently tele-manipulation skills are often taught through video watching [4, 5]. Cases performed by experts are recorded and watched by trainees in the training settings. Several researchers have studied the effects of the video watching on skills acquisition and many have reported positive outcomes [4-6]. The underlying assumption of the video watching for training skills is that once trainees had the opportunity to watch an expert's performance, they would be able to implement similar eye motor programs that connect to motor representation of the manual actions recorded in the videos [7].

The assumption of implementing equivalent eye motion programs triggered by watching others' actions is derived from pioneering works in the study of eye-hand coordination in goal-directed movements done by Flanagan and Johansson [7]. In their study, subjects were required to observe others in performing a grasping and stacking task. The observers displayed similar spatiotemporal features in their eye-hand coordination as if the task was performed by themselves. Similar to the actors, observers could perform proactive gaze movements, where the gaze reached the target before the hands did. The preservation of proactive gaze in observing was an important piece of evidence for the *equivalent motor program* between eye motion program in observing and hand motion program in operating [7, 8]. More supporting evidence has been collected from subsequent behavioral and neurological studies [9-11].

In this study, we planned to compare eye-hand coordination between operating and observing the actions in the laparoscopic setting. Laparoscopic surgery is performed by placing a digital camera and several long-shafted instruments into the abdominal cavity through keyholes on the abdominal wall [2]. Skill learning for laparoscopic surgery is difficult and requires long training hours before a surgeon gaining confidence and proficiency [1, 2].

In this study, eye-hand coordination was described by spatiotemporal characteristics of the eye and the tool trajectories. We hypothesized that the same spatiotemporal characteristics would be revealed from both proactive eye leaving (eye gaze shifting away from a target before the tool left the target) and gaze-guidance (eye gaze arriving at a target before the tool reached it) with a similar time gap between performing and observing a laparoscopic procedure, which would further suggest that a common motor program would be elicited by watching tool movements in surgical videos.

2 METHODS

Data collection was conducted at the Medical Imaging Research Laboratory of the Simon Fraser University, where ethics approval was obtained from the University *Research Ethics Board*. Fourteen university students (9 males and 5 females, age: 20 ~ 36, mean = 28) with zero surgical experience participated in the study, as we intended to eliminate the influence of surgical expertise on the performance. Sample sizes were estimated based on Sailer, Flanagan & Johansson 2005 study where 10 subjects were recruited to a learning visuomotor task with a similar repeated measure design [12]. All subjects were right-handed with normal or corrected to normal vision. Written consent was obtained from each participant prior to entering the study.

2.1 Apparatus

The experimental apparatus is shown in Figure 1. The remote eye-tracker (Tobii 1750, Tobii Technology, Danderyd, Sweden) was placed on top of the training box 60 ~ 70 cm away from the standing subject. A USB web camera (C525 HD Webcam, Logitech, Fremont, CA) was placed below the Tobii monitor to record the facial movement of the operator. Video taken by this web camera was used for checking the validity of gaze data, and for checking potential periods of lost eye tracking data, such as during blinks and large head movements.

2.2 Task

Subjects used a laparoscopic grasper (Ethicon Endo-Surgery, Cincinnati, Ohio) to move the object (4.5 × 7 mm green cylinder) over three dishes (13 mm in diameter) in a pre-determined order (Figure 1 B). A complete trial included 9 steps comprising three types of tasks (Figure 1B); reaching and grasping (R), transporting and releasing (T), and homing (H). One complete trial took 60 ~ 90 s.

* bin.zheng@ualberta.ca
† xiantaj, stella@sfu.ca
‡ roman.bednarik@uef.fi

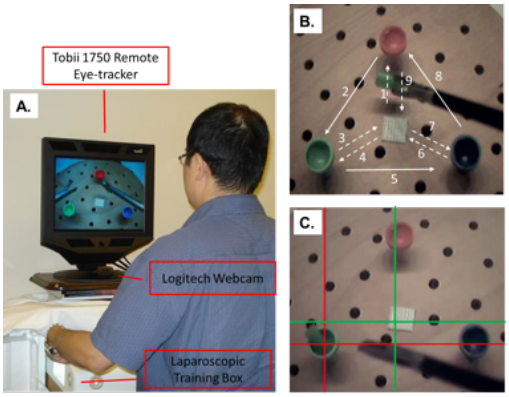

Figure 1: Experimental apparatus. The screen shows surgical site inside the training box, consisting of the peg board with 3 colored dishes of 13 mm diameter, a white homing square (12 × 12 mm), and a green cylinder object of 4.5 mm diameter. The distance between the home plate and each disk is 37 mm, and disks are set 65 mm from each other.

2.3 Procedure

Subjects practiced the task for five minutes before recording to familiarize themselves with the tools and movements. After calibrating their eye gazes, each subject performed 5 trials of the task while their eye movements were recorded, with a short break between each trial. Two weeks after performing the task, subjects returned and watched videos of the trials they had performed on the same display monitor while their eye movements were recorded. Subjects were instructed to watch the video as if they were performing the task.

2.4 Video synchronization

The task scene was captured with a television tuner card (Hauppauge HVR2250, Hauppauge, New York) using a NTSC composite video connection and displayed on the Tobii 17" monitor using Clearview 2.7.0 version of external video stimulus.

The web camera recorded at 30 frames/s, whereas the Tobii 1750 recorded eye-tracking data at 50 frames/s. To establish temporal correspondence between the three recording systems (i.e. the surgical video, eye-tracking signals, and web-cam video) we introduced camera flashes at the start and end of the trial. The videos were synchronized using the start and end flashes, in order to obtain accurate temporal correspondence.

The surgical scene video was recorded at a considerably lower resolution (352 × 288 pixels) than the display monitor of the Tobii 1750 (1240 × 1024 pixels). Methods used for aligning videos with different resolutions have been reported elsewhere [13].

2.5 Data analysis

After the surgical scene and eye-tracking signals were synchronized in time and spatial coordinates, the two eye scanning paths (operating and observing) were overlaid on the scene video (Figure 1C). A custom-designed algorithm was developed using C++ (Microsoft Visual Studio, Microsoft, Redmond, WA) and OpenCV Library to identify the location of the tooltip during these videos [13]. The onset of each step was defined by the moment when the tooltip departed from the home plate or a dish. Explicitly for each step, the following moments were recorded: 1) gaze in operating leaving and arriving at a target dish or the home plate, 2) tool leaving and arriving at a target dish or the home plate, 3) gaze in observing leaving and arriving at a target dish or the home plate.

2.6 Measures

The gaze might start to shift away from the home plate or cup before or after the tool. A positive gaze leaving (*GL*) was recorded when the gaze started to leave for the next target before the tool (Figure 2). This was similar to the proactive gaze movement in Flanagan's study. Conversely, a negative *GL* meant that the gaze left after the tool left. The gaze might also arrive at the target before or after the tool. When the gaze arrived at the target before the tool, we recorded a positive gaze guidance (*GG*) (Figure 2). Conversely, when the gaze arrived on the target after the tool, a negative GG was recorded. The time gaps of GL and GG were reported in seconds (s).

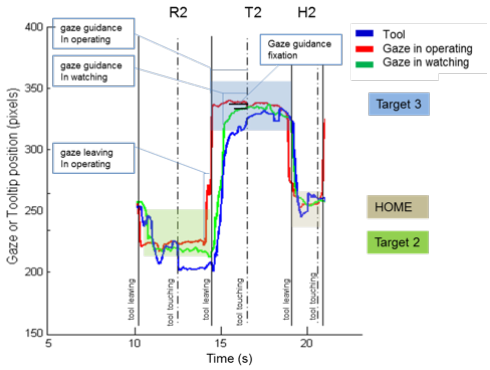

Figure 2: Gaze and tool motion trajectories. The y-axis is the horizontal position of gaze or tool measured in pixels.

We further examined the duration of the gaze fixation prior to touching. Fixation was detected using Salvucci's I-DT dispersion threshold algorithm (Salvucci et al. 2000) with a minimum duration of 100 ms and a maximum dispersion of 40 pixels relative to the captured video scene frame. We named this measure the gaze guidance fixation which is close to the "quiet eye" measure defined by Vickers [14] and is also related to the "target locking" measure used by Wilson and Vine et al in analyzing sequential tasks like surgery [15]. We did not name this gaze guidance fixation as the "quiet eye", because our tasks were sequential and different to the nonsequential ballistic task (i.e. free shooting a basketball) used in Vickers' studies.

2.7 Statistics

Completion time for each subtask was recorded and compared by ANOVA. The proactive gaze variables (*GL, GG*) were subjected to a 2 action mode (operating vs. observing) × 3 task type (R vs. T vs. H) within subject ANOVA. Partial eta squared (η_p^2) were used to calculate effect sizes for means comparisons in ANOVA. A p value less than 0.05 was considered significant. The results are reported in this paper as mean ± standard deviation unless stated otherwise.

3 Results

Fourteen subjects performed a total of 70 trials (5 trials for each subject). However, eye-tracking data were not appropriately recorded 19 trials. Of the valid 51 trials performed by 12 subjects, there were a total of 459 steps (51 trials with 9 steps each). There were 4 invalid steps where the gaze signal was missing during crucial times either in operating or observing. Eye-hand coordination variables could not be obtained on these 4 steps because the calculation required both operating and observing signals to be simultaneously valid. Therefore, we had 455 valid steps for analysis.

3.1 Eye-hand coordination

On average, subjects complete a trial in 44 ± 16 s. The completion times were different between the 3 subtasks (R: 5.4 ± 2.2 s, T: 6.5 ± 2.4 s, H: 2.7 ± 1.5 s, $F_{2,33} = 12.2$, $p < 0.001$, $\eta_p^2 = 0.43$). Post-Hoc (Bonferroni) test revealed the significant differences were present only between Homing and the other two subtasks (R vs. H: $p = 0.05$; T vs. H: $p < 0.001$), but not between R and T ($p = 0.358$).

3.1.1 Gaze leaving (GL)

GL is defined as the time gap between the gaze and the tool starting a rapid movement to the next target, breaking contact with either the home plate or a dish. The gaze generally shifted away from a target after the tool moved off, therefore, most *GL* were recorded with negative values. In operating, positive *GL* occurred in 20.6% of all steps, mostly in reaching and grasping (Figure 4, red diamonds, with a positive value); this rate dropped significantly to only 3.8% in watching the task video ($F_{1,11} = 127.4$, $p < 0.001$, $\eta_p^2 = 0.84$). The low percentage of positive *GL* reflects that the operators and observers were more often *following* the tool when aiming to the next target.

ANOVA analysis on the *GL Duration* revealed significance of action mode ($F_{1,11} = 203.2$, $p < 0.001$, $\eta_p^2 = 0.95$), task type ($F_{2,22} = 42.0$, $p < 0.001$, $\eta_p^2 = 0.79$) but no interaction effect ($F_{2,22} = 1.6$, $p = 0.230$, $\eta_p^2 = 0.13$). While operating, the mean *GL* was - 0.3 ± 0.2 s; whereas in observing, the mean *GL* was -0.7 ± 0.3 s. The mean difference of *GL* gap between operating and observing was 0.47 s. *GL* varied between task type (R: - 0.2 ± 0.2 s, T: - 0.6 ± 0.3 s, H: - 0.6 ± 0.3 s). Post-Hoc (Bonferroni) test revealed significant difference presented between R and T ($p < 0.001$), R and H ($p < 0.001$), but not between T and H ($p = 0.876$). Performing subtasks with low precision requirements, such as Homing, subjects' gazes leave the target much later than the tool when compared to that while performing subtasks requiring more precision (i.e., reaching and grasping).

3.1.2 Gaze Guidance (GG)

GG describes the time gap between the gaze and the tool reaching the target. In our study, subjects' gazes were locked onto the target before the tool in the majority of all the steps both while operating and observing; there was an insignificantly higher occurrence of positive *GG* while performing the task than in observing (100% vs. 96.4%; $F_{1,11} = 0.022$, $p = 0.886$, $\eta_p^2 = 0.36$). The percentage of positive *GG* was similar over the three subtasks (R: 97.4%, T: 98.7%, H: 98.5%; $F_{2,22} = 1.396$, $p = 0.226$, $\eta_p^2 = 0.27$). No significant interaction was found on the rate of positive *GG*.

For *GG Duration*, ANOVA reported significance on action mode ($F_{1,11} = 133.1$, $p < 0.001$, $\eta_p^2 = 0.92$), task type ($F_{2,22} = 23.4$, $p < 0.001$, $\eta_p^2 = 0.68$), but not for interaction ($F_{2,22} = 1.2$, $p = 0.323$, $\eta_p^2 = 0.10$). Subjects' eyes fixated on the target 2.0 ± 0.5 s before tools in operating and 1.5 ± 0.5 s in observing. On average there was a 0.5 s time gap on the *GG* between operating and observing the tasks.

The length of *GG* varied between three types of tasks (R: 2.0 ± 0.5 s, T: 1.9 ± 0.6 s, H: 1.4 ± 0.4 s). Post-Hoc (Bonferroni) test revealed significant difference presented between R and H ($p < 0.001$), T and H ($p = 0.002$), but not between R and T ($p = 0.101$). While performing subtasks with higher precision requirements, such as reaching and grasping, subjects' gazes entered the target much earlier than the tool when compared to that while performing subtasks requiring less precision (i.e. Homing).

3.1.3 Gaze Guidance Fixation

Analysis of the duration of gaze fixations in gaze guidance revealed significant differences with the main effects from the action mode ($F_{1,11} = 4.8$, $p = 0.050$, $\eta_p^2 = 0.31$) and task type ($F_{2,22} = 11.7$, $p = 0.001$, $\eta_p^2 = 0.52$) but not for interaction ($F_{2,22} = 0.4$, $p = 0.699$, $\eta_p^2 = 0.03$).

Subjects performed longer fixations on the target before the tool reached the target while operating (1.1 ± 0.4 s) compared to observing the video (0.91 ± 0.4 s). The *GG* fixation was longest in the R (1.2 ± 0.5 s) and shorter in T (0.9 ± 0.3 s), and H (0.7 ± 0.3 s). Post-Hoc (Bonferroni) test revealed significant difference only presented between R and H ($p = 0.005$), not between R and T ($p = 0.056$) and T and H ($p = 0.081$). Results suggested that performing subtasks with high precision requirements, subjects' eyes fixated on the target for a longer time prior to the tool arriving at the target. Post-hoc tests revealed differences between R and T ($p = 0.035$), and R and H ($p < 0.001$) but not between T and H ($p = 1.111$).

3.1.4 Overview on spatiotemporal features

To give a global overview on the spatiotemporal features between operating and observing, all trials were normalized by the median of each subtask time as reported by Flanagan [7]. The eye-tracking data collected from both operating and observing and the tooltip position data were normalized using a customized Matlab script to create spatiotemporal plots (Figure 3). Clear gaps were observed between operating and observing in many steps, which suggest different motor programming was issued between operating and watching a surgical procedure.

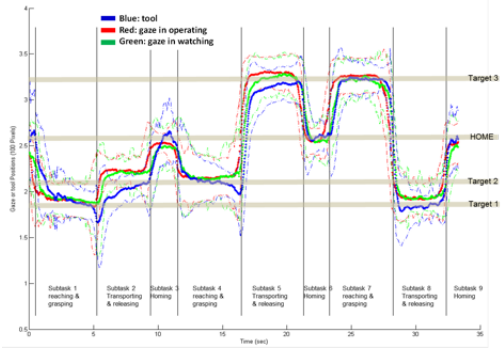

Figure 3: Overview of eye-hand coordination between observing and operating a laparoscopic task. The y-axis is the horizontal position of gaze or tool measured in pixels.

4 Discussion

A better way to describe motor behaviors should include more spatiotemporal information of eye movement as we believe that eye-hand coordination of an operator is influenced by central mechanisms [15, 16]. In our study, we plot the entire eye-hand coordination curve to display a global view, reflecting different involvement of gazes between operating and watching the performance and how each measure is adjusted by task requirement.

4.1 Gaze leaving

In observing actions recorded in the video, gaze leaving before tool was rarely seen (3%); this rate increased significantly in operating (27%). The low rate of positive *GL* may be caused by novice subjects we included in the study. Keeping in mind that the subjects are university students, and therefore are unlikely to have used laparoscopic instruments before, they often drop the object (cylinder) during the grasping (loading) or releasing (loading) phases of movement. The unskillful subjects were reluctant to shift their vision away from their current action before they could ensure that the task was completed with accuracy. This explanation can be supported by the results of *GL* duration over

three different types of subtasks. Before heading for the home plate, *GL* was 0.6 s after the tool. Subjects needed more time to ensure the cylinder was loaded well into the cup in the previous action. In contrast, *RG* task recorded the shortest *GL* (0.2 s) because subjects were not so concerned about tool position at the home plate before they performed the reaching and grasping subtask.

As an average over all subtasks, the GL in operating (0.2 s) was shorter than observing (0.6 s); subjects in operating required a much shorter time staring at the previous target than in watching the video. They would rather spend more time to check the accuracy of others rather than quickly move their vision away for the next task. Another explanation may come from the fact that haptic feedback in the video observation is lost; subjects can only rely on visual feedback [3]. So that observers needed more time to collect information on target movement.

4.2 Gaze guidance

Although gaze often left the current target after the tool, the gaze reached the target earlier than the tool did. In operating, the rate of positive *GG* was 100%, whereas in observing the *GG* rate was still as high as 93%. This indicates clearly that watching a tool's motion can trigger a proactive type of gaze movement.

In observing, proactive gaze movement on the target can serve different purposes except for bonding with a hand movement. Therefore, subjects did not have urgent needs to fixate on the target. The GG duration dropped to 1 second in the video watching. Subjects may use this time to inspect how tasks were performed.

After the gaze reached the target, it may scan over a number of locations related to the target. Only the last fixture before the tool reaches the target may be directly related to guiding the action follow-up. The last fixation is close to the quiet eye phase defined by Wilson [17] and Vickers [14], it lasts 1.1 s in operating, and 09 s in watching. Over three subtasks, it decreases as the task challenge was reduced. This evidence suggests that task requirements can regulate vision for taking information from the target. Subjects can always perform proactive type gaze when required, regardless whether they are performing or observing the task.

The tasks performed in this study are simple and straightforward. Subjects practice for five minutes to ensure they know the task procedure. Observers can predict the upcoming actions and move their eyes to the next target, which explained why the gaze guidance occurred in as high as 100% in operating and 96.4% in observing. However, when complex surgical videos are watched by junior surgeons who are not capable of anticipating upcoming actions, the gaze guidance may drop significantly [16].

4.3 Implication and future direction

The existence of different eye-hand coordination between video watching and operating help us re-consider how to improve outcome for surgical skill learning. When we encourage junior surgeons to watch surgical videos, we teach them how to perform a surgical procedure step-by-step. At the same time, watching the video can help them to learn how to handle tissues with the right instruments at the right time. However, watching videos has limited capacity to teach surgeons to develop a smart strategy in searching for critical visual cues for guiding movement and making good decisions for safe surgery. A possible solution for enhancing the teaching value of video watching would be to embed the expert's eye gaze into training videos. Once the junior surgeons had a chance to observe the eye scanning pattern of experts, they would have a chance to learn from experts how to take proactive visual inputs for guiding their performance in hands [18]; they could also improve their vigilance in the operating room for safe surgery [19]. In this sense, further studies on gaze behaviors of surgeons may help to improve laparoscopic performance and patient safety.

In the future, we would like to continue to gather information and evidence regarding eye-hand coordination during video-watching, as our long-term goal is to study surgeons' behavior during image-guided surgery and to develop new technology to improve skills training. This goal can be reached by a series of studies: examining eye-hand coordination in complex surgical procedure recorded directly from surgeons in the operating room, and examining the impact of watching gaze of expert surgeons on the skill learning of novice surgeons.

5 Conclusion

The eyes of subjects were primarily checking outcomes of tool movement while observing task videos; different to actively collecting information on the target for guiding tool movement during task operating. Landmark behaviors in gaze were adjusted differently by task requirements which questions whether one motor program can regulate eye-hand coordination in both operating and observing.

Acknowledgments

We thank Mr. Bo Fu, Dr. Eric Fung for editing the manuscript and Dr. Lee Swanström for providing critical comments on the manuscript. We thank the Canadian Natural Sciences and Engineering Research Council (NSERC), and the Royal College of Physicians and Surgeons in Canada (RCPSC) for funding the surgical education project.

References

[1] R. Aggarwal, *et al*., "Laparoscopic skills training and assessment," *Br J Surg*, vol. 91, pp. 1549-58, Dec 2004.

[2] A. Cuschieri, "Cost efficacy of laparoscopic vs open surgery. Hospitals vs community," *Surg Endosc*, vol. 12, pp. 1197-8, Oct 1998.

[3] F. Tendick, *et al*., "Technological aspects of minimal access surgery," *Proc Inst Mech Eng [H]*, vol. 211, pp. 129-44, 1997.

[4] M. N. Akl, *et al*., "The efficacy of viewing an educational video as a method for the acquisition of basic laparoscopic suturing skills," *J Minim Invasive Gynecol*, vol. 15, pp. 410-3, Jul-Aug 2008.

[5] L. A. Scherer, *et al*., "Videotape review leads to rapid and sustained learning," *Am J Surg*, vol. 185, pp. 516-20, Jun 2003.

[6] P. Yeung, *et al*., "Comparison of text versus video for teaching laparoscopic knot tying in the novice surgeon: A randomized, controlled trial," *J Minim Invasive Gynecol*, vol. 16, pp. 411-5, 2009.

[7] J. R. Flanagan and R. S. Johansson, "Action plans used in action observation," *Nature*, vol. 424, pp. 769-71, Aug 14 2003.

[8] B. Gesierich, *et al*., "Human gaze behaviour during action execution and observation," *Acta Psychol (Amst)*, vol. 128, pp. 324-30, Jun 2008.

[9] D. Elliott, *et al*., "Goal-directed aiming: two components but multiple processes," *Psychol Bull*, vol. 136, pp. 1023-44, Nov 2010.

[10] V. Caggiano, *et al*., "Mirror neurons encode the subjective value of an observed action," *Proc Natl Acad Sci U S A*, vol. 109, pp. 11848-53, Jul 17 2012.

[11] C. Sinigaglia and G. Rizzolatti, "Through the looking glass: self and others," *Conscious Cogn*, vol. 20, pp. 64-74, Mar 2011.

[12] U. Sailer, *et al*., "Eye-Hand Coordination during Learning of a Novel Visuomotor Task," *J Neurosci*, vol. 25, pp. 8833-8842, 2005.

[13] X. Jiang, *et al*., "Video processing to locate the tooltip position in surgical eye-hand coordination tasks," *Surg Innov*, p. Manuscript submitted in Sept 2013, 2013.

[14] J. N. Vickers, Perception, Cognition, and Decision Training:The Quiet Eye in Action. Champaign, IL, US: Human Kinetics, 2007.

[15] M. C. Bowman, *et al*., "Eye-hand coordination in a sequential target contact task," *Exp Brain Res,* vol. 195, pp. 273-83, May 2009.

[16] M. M. Hayhoe, *et al*., "Visual memory and motor planning in a natural task," *Journal of Vision,* vol. 3, pp. 49-63, 2003.

[17] M. Wilson, *et al*., "Psychomotor control in a virtual laparoscopic surgery training environment: gaze control parameters differentiate novices from experts," *Surg Endosc,* vol. 24, pp. 2458-64, Oct 2010.

[18] M. R. Wilson, *et al*., "Gaze training enhances laparoscopic technical skill acquisition and multi-tasking performance: a randomized, controlled study," *Surg Endosc,* vol. 25, pp. 3731-9, Dec 2011.

[19] B. Zheng, *et al*., "Surgeon's vigilance in the operating room," *Am J Surg,* vol. 201, pp. 673-7, May 2011.