

Saccadic delays on targets while watching videos

M. Stella Atkins*

Xianta Jiang†

Geoffrey Tien‡

School of Computing Science
Simon Fraser University

Bin Zheng§

Department of Surgery
University of Alberta

Abstract

To observe whether there is a difference in eye gaze between doing a task, and watching a video of the task, we recorded the gaze of 17 subjects performing a simple surgical eye-hand coordination task. We also recorded eye gaze of the same subjects later while they were watching videos of their performance.

We divided the task into 9 or more sub-tasks, each of which involved a large hand movement to a new target location. We analyzed the videos manually and located the video frame for each sub-task where the operator's saccadic movement began, and the frame where the watcher's eye movement began. We found a consistent delay of about 600 ms between initial eye movement when doing the task, and initial eye movement when watching the task, observed in 96.3% of the sub-tasks.

For the first time, we have quantified the differences between doing and watching a manual task. This will help develop gaze-based training strategies for manual tasks.

CR Categories: H.1.2 [Information Systems]: Models and Principles—User/Machine Systems; Human Factors; I.4.8 [Computing Methodologies]: Image Processing and Computer Vision—Scene Analysis; Tracking;

Keywords: Eye tracking, saccadic movement, eye hand coordination, watching vs. doing, laparoscopic surgery

1 Introduction

Laparoscopic (keyhole) surgical procedures involve complex eye-hand coordination, where the surgeon operates using long tools inserted through small holes in the body, while viewing the surgical scene on a display monitor. For decades, laparoscopic skills have been taught by having surgeons-in-training watch videos of surgical procedures [Rosser et al. 2000; Scherer et al. 2003; Birch et al. 2007], by teaching learners to imitate expert surgeons' actions. This takes hours of practice of manual skills in the laboratory or clinical environment. The learning process is long and stressful, and outcomes are questionable. This teaching model is proven to be inefficient and sometimes even unsafe when surgeons are overloaded [Wilson et al. 2011].

*e-mail: stella@cs.sfu.ca

†xiantaj@cs.sfu.ca

‡gtien@cs.sfu.ca

§e-mail: bin.zheng@ualberta.ca

Copyright © 2012 by the Association for Computing Machinery, Inc. Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions Dept, ACM Inc., fax +1 (212) 869-0481 or e-mail permissions@acm.org.

ETRA 2012, Santa Barbara, CA, March 28 – 30, 2012.

© 2012 ACM 978-1-4503-1225-7/12/0003 \$10.00

A new training model has emerged recently by shifting surgeons' focus from the endpoints of tool movement, to where the expert surgeon is looking. The theory supporting this new learning strategy is called implicit learning [Masters et al. 2008]. According to this theory, a surgeon can integrate key visual information to the motor program once he understands where to look, and further develop coordination implicitly in his motion system [Masters et al. 2008; Wilson et al. 2011].

Studies of laparoscopic surgeons' eye motions reveal differences between novices and expert surgeons [Law et al. 2004; Wilson et al. 2010] – novices tend to follow the tool tip much more than experts, who generally focus on the target. Other studies [Kocak et al. 2005; Sodergren et al. 2010; Tien et al. 2010; Wilson et al. 2010; Zheng et al. 2011] have also shown that novices and experts have different eye gaze patterns during simulated surgery tasks.

Several authors have demonstrated that by teaching junior surgeons to copy expert surgeons' gaze strategy, trainees can learn faster [Masters et al. 2008; Sodergren et al. 2010; Wilson et al. 2011]. Wilson et al. have shown that knowing where to look improves overall surgical performance of trainees more than by receiving verbal feedback from an instructor or just watching a video [Wilson et al. 2011].

To endorse this new learning strategy, it is important for us to display an expert surgeon's eye-tracking onto the video stream of a performed surgical procedure. A practical question we encountered was whether or not we can record an expert surgeon's eye-tracking *outside* the operating room while he watches a surgical procedure, for easily creating an eye-tracking superimposed surgical video for teaching. In a previous study, Law et al. have documented experts performed a saccade to the target to collect information about target location and target size, before the hand starts to move [Law et al. 2004]. Hence the early onset of saccade is essential for goal-directed movement.

The goal of this project is not to compare gaze differences between experts and novices during laparoscopic procedures; instead, we decided to investigate whether saccades would be performed in the same way between the expert surgeons while doing and later while watching a surgical procedure. Since we aim to provide feedback to trainee surgeons based on the experts' eye gaze during surgical tasks, it is important to understand fully the differences in eye gaze pattern between doing a surgical task, and later watching a video of the task. We hypothesized that the eye scanning strategy will be different when someone is merely watching a video, compared with actually doing the task. Here we quantify the differences in saccadic movements performed by university students between doing a simple surgical task, and watching the video of the same task.

2 Method

2.1 Task and Apparatus

A one-handed task was designed, to move a peg inside a box from one dish into another using a long grasper. We used a standard la-

paroscopy training box (3D Technical Services) which consists of a closed box containing the task peg board, illuminated with a camera, with the inside image displayed on a monitor. A right-handed grasper is inserted into a hold entering the box. One trial of the task is to use the grasper to move the peg from the red dish at the top to the green dish on the left, to the blue dish on the right, and finally back to the red dish, touching the center white square with the grasper tip at the start and after each move. This emulates a simple surgical training task, and aims to teach eye-hand coordination using laparoscopic tools. A single trial takes about 60-90 seconds to complete. We recorded 17 right-handed subjects (14 male, 3 female, aged 24-45, mean 29) who completed a pre-test questionnaire and practiced the task for a few minutes, until they were ready to begin. They then performed 5 trials of the task, with a short break between trials. At least two weeks later the subjects returned and watched videos of 15 trials on the same display monitor while being eye tracked. The 15 trials included 10 trials from other subjects, randomly interspersed with their own 5 trials. We had performed a pilot study and shown there was no significant difference in eye gaze behavior between viewing a video sitting down or standing up, so for comfort we allowed the subjects to view the videos while seated.

The eye-tracker, a Tobii 1750, is incorporated into the monitor, and records eye gaze on the 17" LCD display at 50 Hz using the cameras built into the base of the monitor, accurate to within 1°. The user must be 60-70 cm away, and must not make excessive head movement. The task scene was captured with a Hauppauge HVR2250 using a NTSC composite video connection at 352 × 288 pixels and displayed on the Tobii 1750 using their Clearview software's (version 2.7.0) "external video" stimulus. Additionally, a USB web camera was used for verification of valid eye gaze data and potential periods of lost eye-tracking data.

Note that the white center square was estimated to be separated from the dishes by a visual angle of about 9°, and the dishes were separated from one another by about 17°.

2.2 Data Analysis

2.2.1 Establishing Temporal and Spatial Correspondence

The web camera was recording at 30 frames/sec, whereas the Tobii 1750 records eye-tracking data at 50 frames/sec. Hence we had to synchronize the videos using camera flashes at the trial start and end.

The scene video recorded at a considerably lower resolution (352 × 288 pixels) than the display monitor of the Tobii 1750 (1280 × 1024 pixels). From prior tests we observed that Tobii's Clearview software automatically scales the external video stimulus to the display monitor's native resolution without aspect ratio preservation. However, when using Clearview's "AVI video" stimulus for the watching phase, the selected video is displayed in the center of the screen in its original dimensions, with black borders added on all sides. Hence to present the same visual stimulus as what was seen while doing the task, the AVI video needed to be non-uniformly scaled up for watching, by a factor of 1280/352 horizontally and by 1024/288 vertically. For our analysis, to overlay the "doing" eye gaze properly on the upscaled "watching" video, we also had to upscale the recorded "doing" eye gaze coordinates to the range 1280 × 1024.

2.2.2 Comparing Gaze between Doing and Watching

The two eye scanning paths (doing and watching) were overlaid on the scene video, and manually analyzed by viewing the overlaid videos frame-by-frame. A screenshot of a typical frame is shown

in Figure 1, where the yellow cross indicates the eye gaze while doing, and the green cross indicates the eye gaze while watching the video.

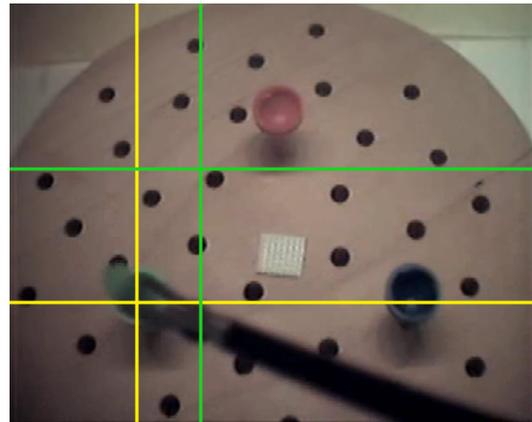


Figure 1: Enhanced screenshot from subject 5 trial 1, showing the dual eye gaze overlaid as a yellow cross, now over the grasper (while doing) and a green cross (while watching).

We used a trial data sheet for each trial of each subject, indicating each sub-task, as shown in Figure 2. The trial data sheet shows the start and end times of the trial (in secs), and a schematic picture of the peg board task, with labeled arrows corresponding to each sub-task. The table at the side allows us to record the time when the corresponding eye or hand movement started. This particular trial data sheet is filled with sample data from Subject 5 trial 1.

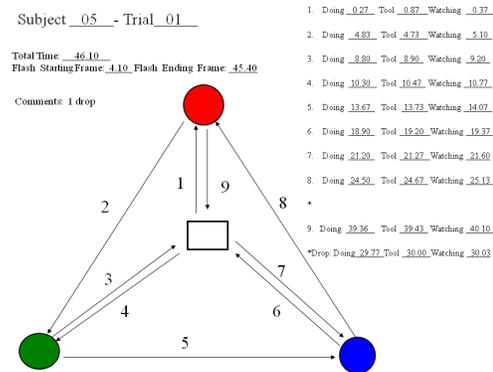


Figure 2: The trial data sheet for subject 5 trial 1. The numbers 1-9 on the labeled arrows indicate the successive sub-tasks of the trial. The start time (secs) of each sub-task is identified in the columns on the right.

The screenshot in Figure 1 shows data from subject 5 trial 1, summarized in Figure 2, at about 4.83 secs, with the doing eye gaze at the left target at the end of sub-task 2, and the watching eye gaze still far from this target.

Extra sub-tasks such as dropping the peg occurred several times, and were recorded as intermediate data within the sub-task sheet, as seen in Figure 2 where the peg was dropped between sub-task 8 and the last sub-task (sheet slightly modified for clarity in this publication). Data were later transferred to a spreadsheet for analysis.

3 Results

There were 85 “doing” trials (5 for each of 17 subjects). 34 trials were discarded where fixations were recorded for less than 70% of the elapsed time due to head movement. All the watching trials were valid, as the subjects were seated.

Of the 51 valid doing trials, there was a total of 459 sub-tasks available for analysis (51 trials with 9 sub-tasks each). There were just 4 invalid sub-tasks (< 1% sub-tasks) where the gaze signal was missing during crucial parts of the trial; dual eye gaze tracking analysis requires both signals to be valid simultaneously. There were 28 drops which we counted as additional valid sub-tasks because we required subjects to pick up the dropped peg in order to resume and complete each sub-task. Therefore we had 483 valid sub-tasks for analysis (459 – 4 + 28).

Typical data for the start of the eye saccade for each sub-task while “doing” and “watching” (from the data sheet shown in Figure 2), are shown in Table 1, with the saccade gap in ms between “doing” and “watching” in the last column. Of the main sub-tasks excluding drops, the average gap between eye saccade movement during “doing” and “watching” was 574 ± 330 ms. Sub-task 2 had the longest average gap of 737 ± 363 ms, and all the other sub-tasks had an average saccade delay between 514 ms and 591 ms. This is a very consistent result over 455 sub-tasks.

The saccade delays over a complete trial are visualized over time in Figure 3, where the top row shows the X-pixel value of the gaze and the second row shows the Y-pixel coordinate of the gaze, for the same subject 5 trial 1 shown in Figure 2 and Table 1. In Figure 3 the black line shows the gaze points while doing and the purple line shows the gaze points while watching. The times at which the eye saccades start for each sub-task have been overlaid as dashed vertical lines, using data taken from Table 1. Note that in this example trial, sub-task 2 showed a low saccade gap of only 270 ms. However, the visualization in Figure 3 illustrates that this saccade was paused half-way, possibly changing to a smooth pursuit.

4 Discussion and Conclusions

For the first time, we have quantified the differences of the onset of saccadic eye movement on a target, between “doing” and “watch-

| Sub-task number | Doing start time (s) | Watching start time (s) | Saccade gap (ms) |
|-----------------|----------------------|-------------------------|------------------|
| 1 | 0.27 | 0.37 | 100 |
| 2 | 4.83 | 5.10 | 270 |
| 3 | 8.80 | 9.20 | 400 |
| 4 | 10.30 | 10.77 | 470 |
| 5 | 13.67 | 14.07 | 400 |
| 6 | 18.90 | 19.37 | 470 |
| 7 | 21.20 | 21.60 | 400 |
| 8 | 24.50 | 25.13 | 630 |
| 9 | 39.36 | 40.10 | 740 |
| Drop | 29.77 | 30.03 | 340 |

Table 1: Saccade start times for each sub-task (sub. 5 trial 1).

ing” a manual task. The consistent delay of saccadic movement in “watching” indicates that a vital piece of gaze strategy may be absent during watching. Early onset of saccadic eye movement to a target provides information for guiding upcoming hand movement, which is the essence of eye-hand coordination junior surgeons must learn. Since this component of gaze strategy is missing in video watching, we suggest that gaze-augmented teaching video should only be created by recording the operator’s gaze while doing the procedure.

It may appear from Figure 3 that the gaze locations are very similar between “doing” and “watching”, but we have determined that the gaze overlap is on average only 68%-82%, depending on the visual field considered as overlapping [Tien et al. 2012].

It is interesting to examine what causes the 600 ms delay between “doing” and “watching”. In doing a task, active visual search to the target initiates before moving the hand, whereas in watching the task, active visual search is not necessary. Previous studies on goal-directed movement showed that there can be a 100-400 ms delay in time between saccadic eye movement to the target and hand movement [Elliott et al. 2010; Deconinck et al. 2011]. After noticing the tool movement in video watching, subjects tended to follow the movement of the tool in a smooth pursuit with a reaction time of 200-300 ms [Sailer et al. 2005; Deconinck et al. 2011]. Adding these times together, we are not surprised to see there is a delay of

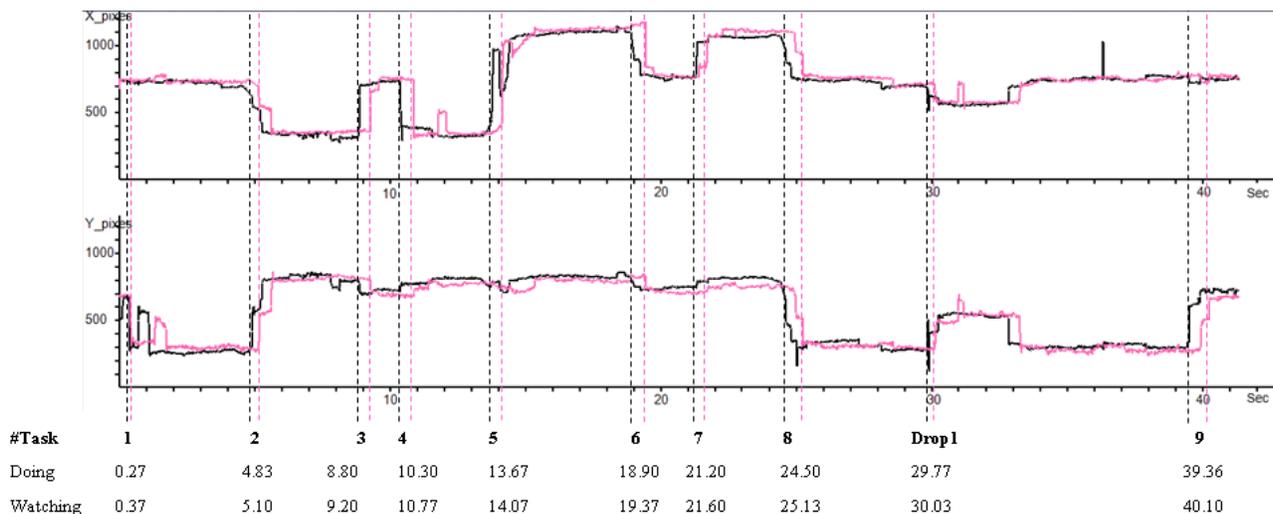


Figure 3: Gaze points for doing (black) and watching (purple) over time, for subject 5 trial 1. The top row shows the X-pixel location, and the second row shows the Y-pixel value. The dashed lines show the start of the eye saccade for doing (black) and watching (purple).

600 ms between the onset of saccadic eye movement in “doing” and “watching”.

The gap between doing and watching can be explained by lack of planning and control while watching passively, regardless of knowledge of the task instance. The doers perform target fixations whereas the watchers likely do not fixate on the target, instead probably performing tool tracking, although faster sampling rates would be needed to establish this point.

Results of this study shed light in the development of gaze-based training strategies for surgical tasks involving eye-hand coordination. Given the delay of eye movement in video watching, displaying the gaze strategy obtained from expert video watching is unlikely to facilitate eye-hand coordination for guiding trainees’ actions. Instead, displaying the gaze of a surgeon recorded live in the operating room is needed to accelerate the development of eye-hand coordination in goal-directed movement. In future work, we will further study eye-hand coordination using dual eye-tracking technology, and will observe tool movements in the same context. The time delay between the onset of saccadic movement and the onset of tool movement will enrich our knowledge on the visual feedback loop in performing goal-directed movements.

Acknowledgements

We thank Yifan Hao for manually annotating the videos, and the Canadian Natural Sciences and Engineering Research Council (NSERC), and the Royal College of Physicians and Surgeons in Canada (RCPC) for funding this project.

References

- BIRCH, D. W., SAMPLE, C., AND GUPTA, R. 2007. The impact of a comprehensive course in advanced minimal access surgery on surgeon practice. *Canadian Journal of Surgery* 50, 1 (February), 9–12.
- DECONINCK, F., VAN POLANEN, V., SAVELSBERGH, G., AND BENNETT, S. 2011. The relative timing between eye and hand in rapid sequential pointing is affected by time pressure, but not by advance knowledge. *Experimental Brain Research* 213, 99–109.
- ELLIOTT, D., HANSEN, S., GRIERSON, L. E. M., LYONS, J., BENNETT, S. J., AND HAYES, S. J. 2010. Goal-directed aiming: Two components but multiple processes. *Psychological Bulletin* 136, 6 (November), 1023–1044.
- KOÇAK, E., OBER, J., BERME, N., AND MELVIN, S. 2005. Eye motion parameters correlate with level of experience in video-assisted surgery: Objective testing of three tasks. *Journal of Laparoendoscopic & Advanced Surgical Techniques, Part A* 15, 6 (December), 575–580.
- LAW, B., ATKINS, M. S., KIRKPATRICK, A. E., AND LOMAX, A. J. 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*, ETRA ’04, 41–48.
- MASTERS, R., LO, C., MAXWELL, J., AND PATIL, N. 2008. Implicit motor learning in surgery: Implications for multi-tasking. *Surgery* 143, 1, 140–145.
- ROSSER, J. C., HERMAN, B., RISUCCI, D. A., MURAYAMA, M., ROSSER, L. E., AND MERRELL, R. C. 2000. Effectiveness of a cd-rom multimedia tutorial in transferring cognitive knowledge essential for laparoscopic skill training. *The American Journal of Surgery* 179, 4, 320–324.
- SAILER, U., FLANAGAN, J. R., AND JOHANSSON, R. S. 2005. Eye-hand coordination during learning of a novel visuomotor task. *Journal of Neuroscience* 25, 39 (September), 8833–8842.
- SCHERER, L. A., CHANG, M. C., MEREDITH, J., AND BATTISTELLA, F. D. 2003. Videotape review leads to rapid and sustained learning. *The American Journal of Surgery* 185, 6, 516–520.
- SODERGREN, M. H., ORIHUELA-ESPINA, F., CLARK, J., TEARE, J., YANG, G.-Z., AND DARZI, A. 2010. Evaluation of orientation strategies in laparoscopic cholecystectomy. *Annals of Surgery* 252, 6 (December), 1027–1036.
- TIEN, G., ATKINS, M. S., ZHENG, B., AND SWINDELLS, C. 2010. Measuring situation awareness of surgeons in laparoscopic training. In *Proceedings of the 2010 Symposium on Eye Tracking Research & Applications*, ETRA ’10, 149–152.
- TIEN, G., ATKINS, M. S., AND ZHENG, B. 2012. Measuring gaze overlap on videos between multiple observers. In *Proceedings of the 2012 symposium on Eye tracking research & applications*, ETRA ’12.
- WILSON, M., MCGRATH, J., VINE, S., BREWER, J., DEFRIEND, D., AND MASTERS, R. 2010. Psychomotor control in a virtual laparoscopic surgery training environment: gaze control parameters differentiate novices from experts. *Surgical Endoscopy* 24, 2458–2464.
- WILSON, M., VINE, S., BRIGHT, E., MASTERS, R., DEFRIEND, D., AND MCGRATH, J. 2011. Gaze training enhances laparoscopic technical skill acquisition and multi-tasking performance: a randomized, controlled study. *Surgical Endoscopy*, 1–9.
- ZHENG, B., TIEN, G., ATKINS, S. M., SWINDELLS, C., TANIN, H., MENEGHETTI, A., QAYUMI, K. A., AND PANTON, O. N. M. 2011. Surgeon’s vigilance in the operating room. *The American Journal of Surgery* 201, 5, 673–677.