

# Measuring gaze overlap on videos between multiple observers

Geoffrey Tien\*

M. Stella Atkins†

School of Computing Science  
Simon Fraser University

Bin Zheng‡

Department of Surgery  
University of Alberta

## Abstract

For gaze-based training in surgery to be meaningful, the similarity between a trainee's gaze and an expert's gaze during performance of surgical tasks must be assessed. As it is difficult to record two people's gaze simultaneously, we produced task videos made by experts, and measured the amount of overlap between the gaze path of the expert surgeon and third-party observers while watching the videos. For this investigation, we developed a new, simple method for displaying and summarizing the proportion of time during which two observers' points of gaze on a common stimulus were separated by no more than a specified visual angle.

In a study of single-observer self-review and multiple-observer initial view of a laparoscopic training task, we predicted that self-review would produce the highest overlap. We found relatively low overlap between watchers and the task performer; even operators with detailed task knowledge produce low overlap when watching their own videos. Conversely, there was a high overlap among all watchers. Results indicate that it may be insufficient to improve trainees' eye-hand coordination by just watching a video. Gaze training will need to be integrated with other teaching methods to be effective.

**CR Categories:** H.5.0 [Information Interfaces and Presentation]: General—;

**Keywords:** eye-tracking, perception

## 1 Introduction

Eye-tracking research initially focused on the eye movement behaviors of a single subject watching a static image [Kundel 1975; Nodine and Kundel 1987]. More recent efforts have been made to analyze eye behaviors while viewing dynamic scenes [Law et al. 2004; Nicolaou et al. 2006]. There are situations where it is of interest to determine the degree of gaze path overlap between multiple observers on a common stimulus such as two people watching a television commercial or a parent and child reading a digital storybook [Guo and Feng 2011], or a single observer over a repeated stimulus such as a person playing a video game and then watching a replay of the recorded game. Recent work by Jarodzka et al. demonstrates a method of scanpath comparison between two viewers on a similar stimulus [Jarodzka et al. 2010], generating similarity measures along several dimensions for saccades between fixations.

\*e-mail: gtien@cs.sfu.ca

†e-mail: stella@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](mailto:permissions@acm.org).

ETRA 2012, Santa Barbara, CA, March 28 – 30, 2012.  
© 2012 ACM 978-1-4503-1225-7/12/0003 \$10.00

In health care, when two surgeons perform video-guided surgery side-by-side, it is crucial for the surgical team to be focused on the same surgical target [Anastakis et al. 2000; Scherer et al. 2003]. This is crucial for effective cooperation among surgical team members. Law et al. have observed differences in expert and novice gaze behavior while performing image-guided surgical tasks [Law et al. 2004]. Similar observations have been confirmed in other studies [Kocak et al. 2005]. This notion is supported by Sailer et al. who reported marked stages of different gaze behavior as surgical skill improved [Sailer et al. 2005]. Together these results suggest the possibility of guiding the development of novices' eye movement behaviors or otherwise accelerating the natural progression towards expert gaze patterns.

We have developed a method enabling us to visualize and measure the amount of gaze overlap between multiple recordings on a video-based stimulus. With this technique available we can quantify similarities or differences between expert and novice gaze behavior, in turn informing courses of action in a gaze training program, as well as opening an opportunity for us to assess inter-operator cooperation and shared mental models [Stout et al. 1999; Zheng et al. 2007]. Furthermore, while fixation-based string-edit techniques have demonstrated utility in earlier scanpath comparison studies [Brandt and Stark 1997; Josephson and Holmes 2006], our method compares along the actual time measure and only requires specification of a single separation parameter.

This paper reports preliminary results using this method to analyze gaze overlap of two gaze recordings on a simulated surgical task. Implications of these results for gaze intervention in surgical skills training are discussed.

## 2 Gaze Separation Study: Gaze Overlap on a Laparoscopic Training Task

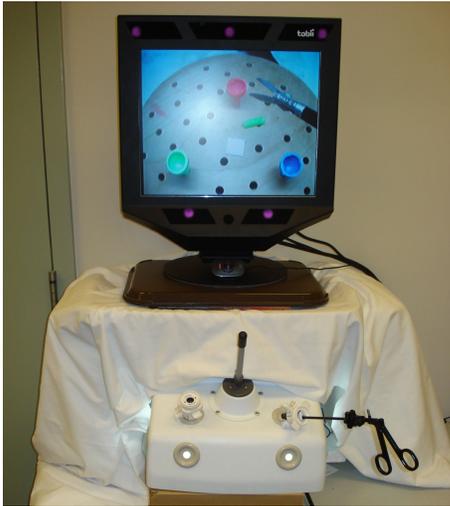
To produce data for testing our gaze overlap software, a small study was conducted to collect eye-tracking data while actively performing a manual task inside a laparoscopic training box, and while passively watching a video recording of the manual task. The recorded videos were watched by both the original operator and by 3rd-party viewers. This created a situation where subjects perceived an identical visual stimulus but had different skill levels and knowledge of the specific task instance.

We developed software which reads the text-based gaze data exported from Tobii Clearview 2.7.X with the Tobii 50-series eye-trackers, which includes the 1750 eye-tracker with an integrated 17" 1280 × 1024 LCD display, and the remote x50 eye-tracker which can be used with any display. From the input gaze data, our software outputs a single value summarizing the amount of overlap between the two gaze recordings.

We hypothesized that recordings from a single subject's task performance and self-review would show a higher overlap than recordings of a different subject watching the task and higher than both subjects watching the task.

## 2.1 Apparatus and Task

The study was comprised of a manual task and a passive watching task. The manual task was performed inside a laparoscopic training box manufactured by 3-D Technical Services. The video from the built-in camera was fed via a NTSC composite connection into a PC running Tobii Clearview 2.7.0 and displayed on the 17" LCD panel of a Tobii 1750 eye-tracker using Clearview's "External video" stimulus. An additional web camera was installed for verification of eye-tracking data loss. Figure 1 shows a reproduction of the apparatus used.



**Figure 1:** *Tobii 1750, web camera, training box and grasper.*

For the manual component, subjects were required to stand in front of the apparatus and use the grasper to transport a small rubber object between three receptacles in a specified order. The receptacles were arranged in a triangular pattern on a peg board with a textured white square in the center.

A single laparoscopic grasper held in the right hand was used to execute the task. The first step in the task was to touch the grasper to the center square. Next, the object had to be picked up from the northern receptacle and placed inside the south-western receptacle, then picked up from south-western receptacle and deposited into the south-eastern receptacle, and finally transported back to the northern receptacle. The grasper was required to return to the center square after each deposition of the object into a new receptacle. If the object was dropped, subjects were instructed to pick up the object and resume the task from the point where the drop occurred. A camera flash was used to mark the beginning and end of each task trial.

## 2.2 Procedure

14 subjects (mean age: 28 years, 9M:5F) from the graduate laboratories of Simon Fraser University participated in this study. Each participant gave signed consent to participate and was given free time to practice the task. When ready, each subject performed the manual task five times and completed a short questionnaire, concluding the manual component of the study.

Out of the 70 total task recordings, 10 were selected to be the ones which were reviewed by others in the study's watching component. These 10 videos were pseudo-randomly assigned to subjects such that each of these videos was watched by 7 other subjects.

At least two weeks following individual participation in the first study component, each subject returned to view their own 5 recordings plus 5 recordings chosen from the bank of 10 described above, in a randomized order. Filenames were obfuscated so subjects would be unable to immediately identify the videos as their own or belonging to others, and participants were not told whether or not any video was their own. These videos were displayed using Clearview's "AVI video" stimulus. Subjects were given a series of questions to answer following review of each of the 10 videos in this watching component.

## 3 Data Characteristics and Implementation

The Tobii GZD is a tab-delimited plain-text file which can be exported from a gaze recording prepared in Tobii Clearview 2.7.X. Our software uses the millisecond timestamp, X and Y gaze points of both eyes, and validity code which denotes the level of confidence that the gaze point was accurately measured.

For our study on dynamic stimuli, we first used an external video stimulus recorded through a NTSC composite video connection. For this arrangement, the exported GZD file contained X gaze point values in the range of 0 to 352 and Y values between 0 and 288. During recording, the input video screen is expanded to fill the display area without preserving the aspect ratio.

Later when using an AVI video stimulus, we non-uniformly resized the video clip to fill the screen at the display's native resolution of  $1280 \times 1024$  pixels. The exported GZD file for this stimulus thus contains gaze point values in this range and the external video GZD must be upsampled before comparison.

Because the point of gaze in GZD files is saved in pixel measurements and we wish to determine gaze separation as degrees of visual angle, we must specify some parameters of the viewing conditions and display characteristics in order to convert a specified visual angle to a pixel value. Using the known monitor diagonal size and resolution, the pixel pitch (in pixels/cm) along a cardinal axis can be approximately calculated using simple trigonometry and unit conversion. Next we must specify an approximate viewing distance in centimeters and the desired visual angle  $\theta$  in degrees. The target separation in pixels is then similarly calculated.

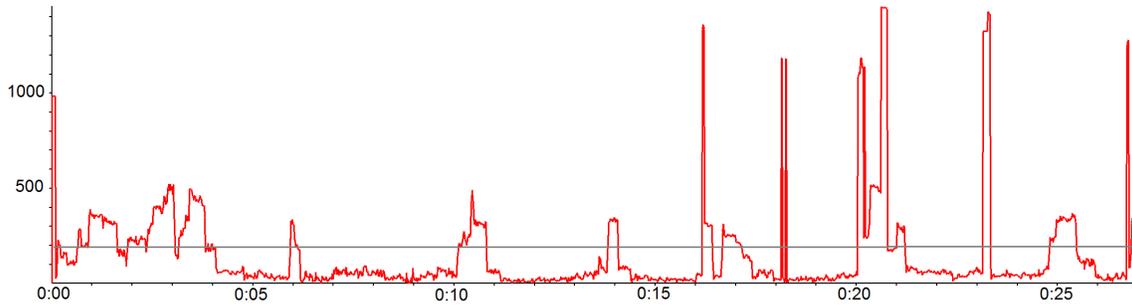
When performing a lengthy or complex task, there may be instances where the operator by necessity looks away from the screen, leading to periods of inactivity during which the eyetracker is unable to detect a point of gaze. Because we cannot make comparisons between recordings where gaze data do not exist, known inactivity periods are manually specified in another text file.

The two input gaze data files do not necessarily contain identical timestamps, so the beginning timestamps are aligned before both gaze data lists can be compared.

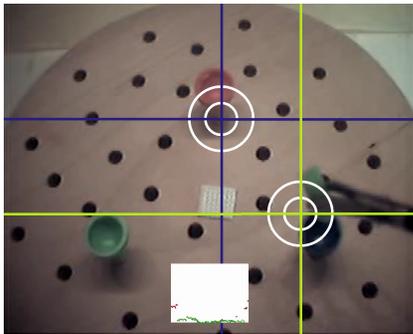
If both recordings have valid gaze points and the timestamp is not within any inactivity period, the Euclidean distance between the points is calculated in pixels and compared to the target separation.

Finally we divide the number of samples where the two gaze paths satisfied our overlap conditions by the total number of comparable samples to attain the average overlap. The overlap can be visualized over time as shown in Figure 2.

A copy of the original stimulus video, with gaze data streams overlaid can also be saved – a sample screen capture of the dual gaze overlaid video is given in Figure 3.



**Figure 2:** Gaze separation visualization, Euclidean screen-space separation on vertical axis, timeline on horizontal axis. The horizontal line at approximately 200 pixels indicates the a gaze separation of 5° visual angle.



**Figure 3:** Dual overlaid screenshot with operator's point of gaze (blue) and 3rd-party watcher's gaze (yellow). 2.5° and 5° circles shown in white.

Subject #	Usable trials	2.5° overlap (% time)	5° overlap (% time)
1	5	65.4	81.8
2	4	57.3	73.2
3	2	72.6	86.9
4	4	76.6	80.9
5	5	82.0	86.8
6	5	42.4	86.0
8	5	75.6	82.5
9	4	74.9	86.6
10	5	55.6	77.9
12	5	69.1	76.9
13	4	84.2	89.2
14	3	64.6	85.8
Mean, std.deviation			
Doing vs. self-rev.		67.9 ± 15.4	82.5 ± 6.4
Doing vs. 3rd-party rev.		70.1 ± 10.6	81.2 ± 6.7
Self-rev. vs. 3rd-party rev.		74.7 ± 9.2	86.9 ± 6.4

**Table 1:** Mean doing vs. self-review overlap, with mean 3rd-party overlap.

## 4 Results

Using the software detailed above, gaze overlap data were compiled for the three scenarios: doing vs. self-review, doing vs. 3rd-party-review, and self-review vs 3rd-party-review. Due to a loss of eye-tracking during a large proportion of the task time in 19 trials including all trials from subjects 7 and 11, these trials where fixations accounted for less than 72% of the task duration were omitted from the doing vs. self-review analysis.

Table 1 lists the percentage of task time where the active and passive gaze for self-review overlapped, averaged over each subject's trials for overlap parameters of 2.5° and 5° visual angle. Note from Figure 3 on our apparatus, the white center square was estimated to be separated from the receptacles by about 9.2° visual angle and the receptacles were separated from one another by roughly 17°. Hence our choice of 2.5° and 5° for analysis are certain to distinguish gazing on separate objects in the scene while still forgiving some gaze jitter.

The overlap amounts are similar for the cases doing versus both self-review and 3rd-party review. However, gaze patterns show higher concordance when both data streams are from passive video review. This indicates that most people will view a task in a similar way, regardless of whether or not there is ownership of the task. Conversely, performing a task first-hand produces different eye movement patterns that cannot be fully reproduced simply from a passive review of the recorded task. The results for the 2.5° are also less stable than for 5°, likely with vulnerability to jitter or sub-optimal calibration being contributing factors.

## 5 Discussion

Our method has advantages over fixation-based string-edit methods due to our tight preservation of temporal data. Our method uses only one user-supplied parameter – the desired visual angle separation. Fixation based measures are affected by the chosen fixation duration and size parameters, and the scanpaths produced in turn can suffer from boundary effects of AOI quantization as well as parameters in scanpath simplification. Furthermore, our method utilizes all available gaze data, including samples which may be omitted from fixations. This allows us to capture gaze overlap which may occur during smooth pursuits, but has the disadvantage that samples collected during the effective blindness of saccades are also included. Saccades are very brief, so relatively few such gaze samples will contribute to the reported overlap.

Rigorous statistical analysis of our results is difficult, because the self-review and 3rd-party review averages were obtained in different ways. More specifically, every subject produced 5 recordings, all of which were available for self-review. In an effort to increase viewership for 3rd-party review without making participation prohibitively lengthy, only 10 videos of the original 70 were used; each one was viewed by 7 other participants. Thus we simply provide a descriptive analysis in this paper. An improved group design in another study will be available for statistical analysis later.

In this study, when a video was watched by multiple reviewers in-

cluding the owner, the gaze points overlap within  $2.5^\circ$  for 75% of the task time, suggesting that a common gaze pattern was employed by all reviewers. In contrast, overlap was lower when comparing an operator's gaze to self-review (68%) and 3rd-party review (70%), which may be explained by a gap in visual reaction or a lack of planning and control while watching passively, regardless of procedural knowledge. This can be supported by our observation that saccades to a target while watching are delayed by 600 ms compared to saccades while doing [Atkins et al. 2012]. With the main task broken down into 9 discrete tool movements, the total watching delay can comprise roughly 5-20% of the task duration, reflected in the reported overlaps.

Wilson et al. have successfully demonstrated that gaze of learners can be used as a feedback resource to improve surgical performance [Wilson et al. 2011]. Our method of analyzing gaze overlap can aid this training approach, but will require more sophisticated scanpath comparison on specified areas of interest as well as sub-task decomposition at key intervals.

The difficulty with experiments modelling tasks such as minimally-invasive surgery is that attempting to measure expert task knowledge is easily confounded by variation in manual expertise. The task chosen for our study was devised to reduce the requirement on both manual dexterity and expertise, so it presented little opportunity to discern expert and novice decision-making by eye metrics.

## 6 Conclusions and Future Work

Eyegaze studies can have many recordings for a given stimulus. It will be useful to batch process a number of gaze recording pairs using the same parameter set and produce an aggregate result.

The graphical timeline is useful for visualizing events where the distance between gaze points is large. Automatic highlighting and indexing of these events can be done as a step towards identifying reasons for gaze pattern discrepancies.

As videos are highly dynamic and variable stimuli, the summary produced by this software is tightly dependent on the actual content viewed. If training value is contained in gaze behaviors of first-hand doing, we can modify our software to compare multiple performances of similar tasks using labeled areas of interest. Being able to give detailed analysis to these areas and periods of interest in the videos over practice can provide insight into the visuo-motor integration process of trainees during skill acquisition of demanding manual tasks. Future studies will be conducted to explore integration of gaze training into surgical education. Additionally, as we have been unable to certainly say that gaze overlap alone is sufficient to measure differences in expertise, we will investigate ways to combine gaze overlap with scanpath analysis during key periods of a new task where expert decision-making will be more evident from eye metrics.

## Acknowledgements

The authors thank the Canadian Natural Sciences and Engineering Research Council (NSERC) for equipment funding.

## References

ANASTAKIS, D. J., HAMSTRA, S. J., AND MATSUMOTO, E. D. 2000. Visual-spatial abilities in surgical training. *American Journal of Surgery* 179, 6, 469 – 471.

ATKINS, M. S., JIANG, X., TIEN, G., AND ZHENG, B. 2012. Saccadic delays on targets while watching videos. In *Proceedings*

*of the 2012 symposium on Eye tracking research & applications, ETRA '12*.

BRANDT, S. A., AND STARK, L. W. 1997. Spontaneous eye movements during visual imagery reflect the content of the visual scene. *J. Cognitive Neuroscience* 9 (January), 27–38.

GUO, J., AND FENG, G. 2011. How eye gaze feedback changes parent-child joint attention in shared storybook reading? an eye-tracking intervention study. In *In Second Workshop on Eye Gaze in Intelligent Human Machine Interaction*.

JARODZKA, H., HOLMQVIST, K., AND NYSTRÖM, M. 2010. A vector-based, multidimensional scanpath similarity measure. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications, ETRA '10*, 211–218.

JOSEPHSON, S., AND HOLMES, M. E. 2006. Clutter or content?: how on-screen enhancements affect how TV viewers scan and what they learn. In *Proceedings of the 2006 symposium on Eye tracking research & applications, ETRA '06*, 155–162.

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 Laparoscopic & Advanced Surgical Techniques, Part A* 15, 6 (December), 575–580.

KUNDEL, H. L. 1975. Peripheral vision, structured noise, and film reader error. *Radiology* 114, 2 (February), 269–273.

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.

NICOLAOU, M., ATALLAH, L., JAMES, A., LEONG, J., DARZI, A., AND YANG, G.-Z. 2006. The effect of depth perception on visual-motor compensation in minimal invasive surgery. In *Medical Imaging and Augmented Reality*, vol. 4091 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 156–163.

NODINE, C. F., AND KUNDEL, H. L. 1987. Using eye movements to study visual search and to improve tumor detection. *Radiographics* 7, 6 (November), 1241–1250.

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. *American Journal of Surgery* 185, 6, 516 – 520.

STOUT, R. J., CANNON-BOWERS, J. A., SALAS, E., AND MILANOVICH, D. M. 1999. Planning, shared mental models, and coordinated performance: An empirical link is established. *Human Factors* 41, 1 (March), 61–71.

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., SWANSTRÖM, L., AND MACKENZIE, C. 2007. A laboratory study on anticipatory movement in laparoscopic surgery: a behavioral indicator for team collaboration. *Surgical Endoscopy* 21, 935–940.