Towards automatic skill evaluation in microsurgery

Shahram Eivazi
University of Tübingen
Perception Engineering, Sand 14, 72076 Tübingen, Germany
shahram.eivazi@mnf.uni-tuebingen.de

Michael Slupina
University of Tübingen
Neural Interfaces and Brain Signal Decoding
Sand 14, 72076 Tübingen, Germany
slupina@informatik.uni-tuebingen.de

Wolfgang Fuhl
University of Tübingen
Perception Engineering, Sand 14, 72076 Tübingen, Germany
wolfgang.fuhl@uni-tuebingen.de

Hoorieh Afkari
University of Eastern Finland
School of computing, Joensuu, Finland
hoorieh.afkari@uef.fi

Ahmad hafez
Helsinki university central hospital
Department of Neurosurgery, Töölö Hospital, Helsinki, Finland
ext-ahmad.hafez@hus.fi

Enkelejda Kasneci
University of Tübingen
Perception Engineering, Sand 14, 72076 Tübingen, Germany
enkelejda.kasneci@uni-tuebingen.de

IUI'17 Companion, March 13-16, 2017, Limassol, Cyprus
ACM 978-1-4503-4893-5/17/03.
http://dx.doi.org/10.1145/3030024.3040985

Abstract

In the past decade, eye tracking has emerged as a promising answer to the increasing needs of understanding surgical expertise. The implicit desire is to design an intelligent user interface (IUI) to monitor and assess the competency of surgical trainees. In this paper, for the first time in microsurgery, we explore the potential for a surgical automatic skill assessment through a combination of machine learning techniques, computational modeling, and eye tracking. We present primary findings from a random forest classification method where we achieved about 70% recognition rate for the detection of expert and novice group. This leads us to a conclusion that prediction of the micro-surgeon performance is possible, can be automated, and that the eye movement data carry important information about the skills of micro-surgeons.

Author Keywords

Eye tracking; machine learning; skill assessment; Micro-neurosurgery

ACM Classification Keywords

H.1.2 [User/Machine Systems]: Human factors

Introduction

In modern-day medical interventions, the surgical skill assessment of trainees is central. At the end of a medical

Figure 1: In microsurgery the surgeon uses a magnified view with miniaturized instruments to operate a caliber vessel in the range of 1-2 mm in diameter under a surgical microscope.

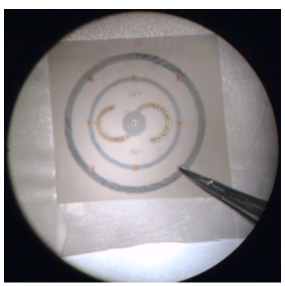

Figure 2: The actual experiment design. Participants were asked to cut and suture precisely along two inner green lines in the left and right side of central circle.

training program, the trainee will typically undergo a subjective assessment. A mentor evaluates the trainee's performance to determine whether a certain objective was met. As such, studies reported in the medical literature on health care and training systems, focus on measurable qualitative outcomes, such as respect for tissue, instrument handling, and quality of the results [9]. Subjective assessment systems have been validated in various clinical settings and found to be the most effective techniques to date [9]. However, the main disadvantage of such an assessment method is that it requires an experienced senior to observe trainees directly, which makes it labor intensive. Moreover, questions regarding how precisely a trainee's knowledge and skills are developed remain unanswered. To improve the assessment techniques for the development of better surgical training systems, other researchers, namely physiologist and cognitive researchers focus on objective analysis methods. Thus, they investigate the correlation between medical evaluation scores and individual surgical skills.

Eye tracking is considered as a tool that provides a behavioral index of ongoing visual and cognitive processes. In medicine, eye tracking has been used as early as the 1970s to study visual search [8] and later to understand the nature of expertise [2]. To date, researchers have used big data analysis methods namely clustering, pattern recognition, and machine learning to propose eye tracking as a suitable objective metric for the automatic assessment of ones level of expertise in medical domains. For example, Richstone et al. [10] and Ahmidi et al. [1] introduced an automated method for assessing surgical skill using eye movement measures. Richstone et al. applied classification methods, linear discriminant analysis (LDA) and neural network analysis (NNA), to group 21 surgeons into expert and novice classes. In a simulated surgery, the system was able to accurately distinguish novice and expert surgeons with 91.9% and 92.9% accuracy, respectively, using eye movement metrics namely, blink rate, fixation rate, pupil size, and vergence. The performance of the prediction system was 81.0% and 90.7% in the live surgery.

While there are numerous studies of visual attention patterns in the context of image-guided surgery, namely laparoscopic surgery [5, 12], gaze patterns have not been studied in any microsurgical OR. In microsurgery, the surgeon uses a microscope to magnify the view and operate on micro-anatomical structure with a size of 1-2 mm (See Figure 1). As such microsurgery requires special psychomotor and visuomotor skills. We believe gaze analysis may provide the details of these skills in such a complex task. Until now, it was not possible to record eye movements in ocular-based setups; thus, there was a lack of evidence of empirically demonstrated visual attention patterns in this domain (e.g., Eivazi et al. [4] or Kübler et al. [6, 7] who studied gaze in microsurgery without the presence of the microscope).

In this paper, we used the recent microscope eye tracker developed by Eivazi [3] to record micro-neurosurgeon eye movements when conducing a cutting and suturing task under surgical microscope. We then applied a machine learning, random forest classification, technique to predict surgeons' performance group. The goal is to evaluate, whether the eye tracking data can be used to detect expertise level for a purpose of proactive intelligent systems.

Method

To answer the question whether gaze data can be used to classify and predict surgeons' expertise level we used novel eye tracker developed in Eivazi's doctoral dissertation [3]. Moreover, we test on two datasets collected from the P(IV) experiment by Eivazi [3]. The author had instructed a group

Measures median	Age	Years of Experience
Participants	33	4
Novices (4)	31	6
Experts (5)	36	2.75

Figure 3: The detailed description of the participants.

		Predicted Labels	
		Novice	Expert
True Labels	Novice	440	232
	Expert	222	496

Figure 4: Results from 1390 predicted data points using scikit-learn random forest classifier in the cutting task.We achieved accuracy of 67% in average.

		Predicted Labels	
		Novice	Expert
True Labels	Novice	7702	4253
	Expert	3369	9444

Figure 5: Results from 24768 predicted data points using scikit-learn random forest classifier in the cutting task.We achieved accuracy of 69% in average.

of surgeons to cut (first dataset) and suture (second dateset) precisely along the lines drawn on the top of a latex glove sheet (see Figure 2). Ten (eight male and two female) neurosurgeons with varying years of training participated in this experiment. We removed data from one participant for the purpose of this study due to low quality gaze estimation. Figure 3 shows participants' expertise level. We divided participants in two groups based on median years of experience (four years).

We analyzed gaze data using Matlab and Paython. Using raw gaze data, exported from our eye tracker software, we employed a velocity-based algorithm for fixation and saccade identification [11]. The velocity threshold was set to 20 deg/sec, minimum fixation duration threshold was set to 100 ms, and maximum distance between two gaze points was set to 20 pixels. Drawing on eye tracking literature, in this study we computed position, duration, amplitude, average euclidean distance, and area covered by fixation and saccade as feature vectors to be used for scikit-learn random forest classifier.

Result

In this paper we analyzed two datasets. Figure 4 shows the recognition accuracy of 67% for the classifier in the cutting task. In total we had 718 data vectors for experts and 672 for novices. The baseline performance was established as a classification accuracy of a majority classifier. Given the fact that more data belonged to the expert group, the majority classifier would perform with accuracy of 52%.

Second, we investigate how well any given data vector informs about the expertise group in the suturing task. Figure 5 shows the recognition accuracy of 69% for the classifier in the cutting task. In total we had 12813 data vector for expert and 11955 for novices. The baseline performance was established as a classification accuracy of a majority classifier. Given the fact that more data belonged to the expert group, the majority classifier would perform with accuracy of 52%.

In summery both tables presents a similar recognition accuracy (about 70%) of predicting into which of the two groups an arbitrary vector of data belongs. This accuracy might be considered relatively low, however the baseline classifier would achieve only 52%. The reason can be found in the small sample size and low number of features that we used in this study.

Discussion

For several decades now, researchers have demonstrated the potential of eye tracking to provide reliable tools for automatic performance assessment. In this study with the help of state-of-the-art surgical microscope eye tracker [3], we extend this knowledge to microsurgery by apply gaze for the assessment of surgeons in micro-neurosurgery. The novel result presented here, in the domain of microsurgery, shows that although the differences in individual and related eye-movement data are subtle and multidimensional, they can be automatically recognized using a radnom forest classification, with about 70% accuracy. This leads us to a conclusion that prediction of the micro-surgeon performance is possible, can be automated, and that the eye movement data carry important information about the skills of micro-surgeons. While the accuracy of classifications was not extremely great, we plan to take data type consideration into account, by both increasing the sample size and extending the features.

The future steps of this research include a development of a real-time system that dynamically captures and classifies surgeons performance based on eye-movement data. In the domain of microsurgery this will enable us to build an intelligent surgical microscope that closely follows the user action and can proactively provide guidance, for example for the purposes of learning.

Acknowledgements

We thank all the surgeons who participated in our study.

REFERENCES

1. N. Ahmidi, G. D. Hager, L. Ishii, G. Fichtinger, G. L. Gallia, and M. Ishii. 2010. Surgical task and skill classification from eye tracking and tool motion in minimally invasive surgery. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 295–302.
2. R. Bertram, L. Helle, J.K. Kaakinen, and E. Svedström. 2013. The effect of expertise on eye movement behaviour in medical image perception. *PloS one* 8, 6 (2013), e66169.
3. S. Eivazi. 2016. *Eye gaze patterns in micro-neurosurgery: from remote to ocular-based eye tracker*. Ph.D. Dissertation. University of Eastern Finland, Dissertations in Forestry and Natural Sciences.
4. S. Eivazi, R. Bednarik, M. Tukiainen, M. von und zu Fraunberg, V. Leinonen, and J. E. Jääskeläinen. 2012. Gaze behaviour of expert and novice microneurosurgeons differs during observations of tumor removal recordings. In *Proceedings of the Symposium on Eye Tracking Research and Applications*. ACM, 377–380.
5. F. Hermens, R. Flin, and I. Ahmed. 2013. Eye movements in surgery: A literature review. *Journal of Eye Movement Research* 6, 4 (2013).
6. T. Kübler, S. Eivazi, and E. Kasneci. 2015. Automated Visual Scanpath Analysis Reveals the Expertise Level of Micro-neurosurgeons. In *MICCAI Workshop on Interventional Microscopy*.
7. T. C. Kübler, C. Rothe, U. Schiefer, W. Rosenstiel, and E. Kasneci. 2016. SubsMatch 2.0: Scanpath comparison and classification based on subsequence frequencies. *Behavior Research Methods* (2016), 1–17.
8. H. L. Kundel and P. S. La Follette Jr. 1972. Visual search patterns and experience with radiological images 1. *Radiology* 103, 3 (1972), 523–528.
9. S. Ramachandran, A. M. Ghanem, and S. R. Myers. 2013. Assessment of microsurgery competency-where are we now? *Microsurgery* 33, 5 (2013), 406–415.
10. L. Richstone, M. J. Schwartz, C. Seideman, J. Cadeddu, S. Marshall, and L. R. Kavoussi. 2010. Eye metrics as an objective assessment of surgical skill. *Annals of surgery* 252, 1 (2010), 177–182.
11. Dario D Salvucci and Joseph H Goldberg. 2000. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*. ACM, 71–78.
12. T. Tien, P. H. Pucher, M. H. Sodergren, K. Sriskandarajah, G. Yang, and A. Darzi. 2014. Eye tracking for skills assessment and training: a systematic review. *journal of surgical research* 191, 1 (2014), 169–178.