Using Eye Tracking to Evaluate and Develop Innovative Teaching Strategies for Fostering Image Reading Skills of Novices in Medical Training

Nora Castner, Shahram Eivazi, and Enkelejda Kasneci

Perception Engineering
Eberhard Karls Universität Tübingen
Tübingen 72076, Germany
castnern@informatik.uni-tuebingen.de
{kasneci,eivazi}@uni-tuebingen.de
http://www.ti.uni-tuebingen.de/Perception.perception-engineering.0.html

Abstract. In the medical domain, developing skills, such as performing surgery, involves copious training and time. Eye movement behavior of experts during domain specific tasks shows measurable differences to those from a student. With this study, we analyze eye movement data from expert and novice microsurgeons while viewing surgical videos. It was found that level of expertise had no effect on fixations and saccades during a surgery video task. However, level of expertise had an effect on the relationship of saccade length and fixation duration. A comparative analysis of areas of interest with respects to level of expertise was also performed and differences were found. Understanding the visual search strategies of experts in specific domains of the medical field can influence the training of students, such as models for gaze-based adaptive learning employed during training of medical students.

Keywords: Eye Tracking, Visual Search, Expertise, Microsurgery, Medical training, Gaze-based training

1 Introduction

Expertise remains an appealing topic; one of the most interesting research areas has been devoted to measure the effectiveness of training programs on progression from novice to expert. Expertise, for any specific skill set, not only develops over time, but also requires intense practice. Experts are less error prone when compared to novices in their field and their accuracy has been shown to evolve over time through practice and exposure [14, 6]. In the medical field, expertise has been shown to be related to faster and more accurate identification of anomalies, such as in radiograph interpretation [2, 12]. It has also been found that expertise relates to the efficiency of performance, where expert laparoscopic surgeons perform surgery more quickly than their novice counterparts [17] and were also more likely to have more fluid hand motions and more “elegant” surgical maneuvering [16].

Within the medical field, novices learning a specific task may take up to two times longer to complete said task: for instance, detecting anomalies [7] or completing a surgery [17]. Gegnenfurtner and colleagues [6] attribute the performance abilities of novices to their inability to reduce extraneous processing demands. Novices may not have solid foundation of information in their memory, thus rapid pinpointing of what is pertinent to the task may be rather unclear. Also, they found that novices are more likely to be less confident in their skills and tend to allocate more time and attention to check their performance rather than trust on their own abilities [6].

Expertise of any domain is heavily reliant on visual processing and regarding the eye movement behavior of both experts and novices, clear differences are apparent. Experts tend to hone their attention more towards task relevant information and reduce any noise from non-important information [14, 4]. For example, expert surgeons look at their tools significantly less than their less experienced counterparts [17]. It has also been found that experts, when performing a task related visual search, employ shorter fixations and larger saccades compared to novices on the same task [6, 14, 7]. Gegenfurtner and colleagues [6] assumed that these eye movement behaviors are indicative of rapid information processing (fixation duration), selective attention allocation (fixation on relevant areas), and thorough global image analysis (saccade length). Differences between experts and novices could also be due to more systematic approaches employed by experts and because of novices’ lack of knowledge and practice [6, 13, 7].

Typically, eye movement behavior patterns, such as the aforementioned, are similar across in experts across domains [6, 1], although task appropriate eye movements are also visible [7, 3, 5, 9]. For instance, expert laparoscopic surgeons focus their attention more to target locations while performing surgeries rather than to the tools employed whereas novice surgeons shift their gaze more often between tools and target locations [17]. On the other hand, Kundel and colleagues [11] found that visual search patterns of medical students were closer to a localized central search method, when compared to staff radiologists, where the search methods were closer to a wider circular pattern. Consequently, the task related eye movements during surgery are likely to be more procedural [15] in comparison to the analyzing and interpreting of radiographs, where the task related eye movements are more reflective of employing visual search strategies [7, 14].

In this study, we also explore the differences in the eye movement behavior of experts and novices in the medical fields. Specifically, we looked at the eye movement data of microneurosurgeons with different levels of expertise while watching video segments of surgery. These eye movement patterns can offer insight into to what we can expect for future study; for instance, one concentrating on visual search and radiography. Since we know from the literature that eye movement behavior is indicative of expertise, we intend to support these assertions and further supplement them with our data from microneurosurgeons.

Furthermore, we will look at expert areas of interest in contrast to non expert areas of interest. We hope to gain insight from these area differences in order to set the foundation for future research regarding gaze-based learning paradigms for medical student.

2 Method

The experiment was conducted at the Neurosurgery Department of Helsinki University Central Hospital. Forty neurosurgeons ($M_{age} = 41.46$ years, $SD_{age} = 10$, 2 female) with varying degrees of expertise from 16 countries participated in the experiment. All participants had at least 2 years of experience in neurosurgery and performed at least 10 surgeries. The average surgeons' years of experience was 13.17 ($SD = 1.59$) and average number of hours performing surgeries was 2213.5 hours ($SD = 2458.76$). Subjects were divided into three groups: Novice (students and residents with less than 1000 hours of surgery experience), intermediate (between 1000 and 2000 hours of surgery experience), and expert (with 2000 and over hours of surgery experience). Data for one subject was excluded for the last video segment due to calibration errors.

2.1 Experiment design and apparatus

We recorded microneurosurgeons' eye movements while they watched video segments of a brain aneurysm clipping. The brain aneurysm clipping video was divided into 6 segments with the length of 20 seconds on average. After each segment, the video was paused and then during the frozen video image, participants had to answer aloud to three questions related to state of the neurosurgery. For recording the eye movement data, a Tobii T120 eye tracker with a 17 inches screen was used with a 60-65 cm viewing distance, 1024 by 768 pixel resolution, and 60Hz sampling rate.

Prior to the experiment, consent form and demographic information was taken by the instructor. First, participants received a brief introduction to the experiment and then were asked to perform 9 points calibration routine for the eye-tracker. After calibration, participants viewed an instruction slide and then proceeded to the video segments.

The study generated a large volume of data and a preliminary analysis indicated that there was great variety in the strategies exhibited. To focus our analysis, in this paper we employed only the data from the first and last video segments: each with a length of 20 seconds. We did not analyze the other segments because they exhibited a slight effect of blurring. Analysis of the gaze data from the chosen video segments was conducted with the software *EyeTrace* [10]. For simplicity and presentation of Area of Interest (AOI) analysis, we chose to look at one image from each segment rather than the whole video since the video clips were not highly dynamic.

2.2 Data analysis

Aligning with much of the previous literature, we measured eye movement differences in experts, intermediates, and novices. This measurement was based on two principle measures of visual attention: fixations and saccades. Specifically, we looked at the differences in number of fixations, mean fixation duration, mean saccades length (duration), and mean saccade amplitude for both video segments. To evaluate whether the time watching the video had an effect on eye movements, we compared these variables between the first and last video segment. Then, for each video segment, we compared the eye movement behavior between the three groups. Lastly, we measured the areas of interest (AOI) for each group over the time course of each video clip.

Using the *EyeTrace* software, fixations were calculated as having a minimum duration of 100ms and saccades were calculated from the time between two fixations. For the AOIs, we calculated gradient-based heat maps from the fixation data for each group. Heatmaps were created for 5 second intervals for each video segment. Since the videos were not highly dynamic regarding dramatic movements, tool or otherwise, we chose a random frame from each segment for the AOI overlay visualization. The heatmap calculation based off fixations is a two dimensional Gaussian distribution. For the calculations, we chose $\sigma = 10$ as determined empirically. We also defined fixation inclusion (a pre-threshold value in *Eyetrace*) to 5%, meaning at least 5% of the total fixations could be considered for an AOI candidate. Thus, for each of the 5 second intervals, this translated to an average minimum of 2 fixations.

3 Results

Eye movement data was compared between the first and the last video segment. A paired samples t-test revealed significantly more fixations in the first segment (M = 30.74, SD = 9.25) than in the second (M = 26.10, SD = 7.79); $t(38) = 2.5$, $p = .017$. There were also significant differences between the mean duration of fixations for the first (M = 678.60, SD = 274.96) and last segment (M = 576.07, SD = 179.90); $t(38) = 2.09$, $p = .044$.

Concerning saccades, both length and duration were shorter in the first segment than in the second: M = 53.85, SD = 15.50 and M = 60.83, SD = 22.06 for saccade length in both segments respectively and M = 61.89, SD = 16.50 and M = 66.26, SD = 17.82 for saccade duration in the respective segments. However, there were no significant differences regarding both saccade measures from the first to last segments.

3.1 Expertise differences

Table 3.1 details the eye movement features for each expertise group. Regarding expertise level, a one-way ANOVA revealed no significant differences regarding a main effect of expertise on eye movements behavior in either video segment (see Table 3.1).

Table 1: Mean and standard deviations of subgroups

Variable	Segment 1 Novice	Segment 1 Intermediate *M(SD)*	Segment 1 Expert	Segment 2 Novice	Segment 2 Intermediate *M(SD)*	Segment 2 Expert
n	16	7	17	15 [a]	7	17
Num. Fixations	31.44 (8.45)	27.86 (7.56)	31.88 (10.77)	22.67 (7.77)	29.14 (6.82)	27.88 (7.45)
Mean Dur. Fixations	638.17 (253.80)	729.44 (215.69)	682.00 (319.94)	607.53 (178.58)	521.83 (180.03)	570.64 (185.90)
Mean Len. Saccade	54.96 (18.26)	55.52 (11.12)	52.49 (14.38)	65.72 (30.20)	66.60 (15.00)	54.13 (13.73)
Mean Dur. Saccade	60.97 (9.17)	71.36 (25.54)	56.90 (16.09)	67.01 (14.68)	68.39 (25.09)	64.71 (18.01)

[a] Calibration issue for one participant resulted in *incomplete* data.

In our data, saccades and fixations during either video segment were then not indicative of level of expertise. Consequently, we looked at the correlation between fixations and saccades to see whether level of expertise had an effect on fixations and saccades.

Table 2: One way ANOVA effects of expertise on eye movements

	Segment 1 $F(2,39)$	**Segment 2** $F(2,38)$
Num. Fixation	$F = .483$ $p = .621$	$F = 2.465$ $p = .085$
Mean Dur. Fixation	$F = .277$ $p = .760$	$F = .542$ $p = .586$
Mean Len. Saccade	$F = .142$ $p = .868$	$F = 1.424$ $p = .254$
Mean Dur. Saccade	$F = 2.138$ $p = .132$	$F = .122$ $p = .886$

A partial correlation analysis revealed that in the first video segment, hours of surgery experience had and effect on the correlation of saccade length and fixation duration (r = -.394, p = .014) as well as saccade length and number of fixations (r = .470, p = .003). However, the effect was not significant in the last video segment. Years of experience also revealed an effect on the same correlations as well: r = -.400, p = .013 and r = .476, p = .003 respectively. Figure 1 depicts the relationship between saccade length and fixation duration as well as number of fixations. It should be worth noting that without the effect of level of expertise: Both Saccade length and fixation duration for the first video segment was significant (r = -.405, p = .01) and saccade length and fixation number for the first segment had significant correlations independent on level of expertise (r = .479, p = .002).

From the correlation analysis, it is clear that longer saccade lengths correlate to both shorter fixation durations and higher numbers of fixations. These correlations were in turn effected by level of expertise. Though it is unclear as to how much since correlations existed independent of expertise.

3.2 pattern of eye movements over time in AOI

To measure the effect of level of expertise on areas of interest over the timecourse of the video segments, we calculated gradient- based heat maps based on the

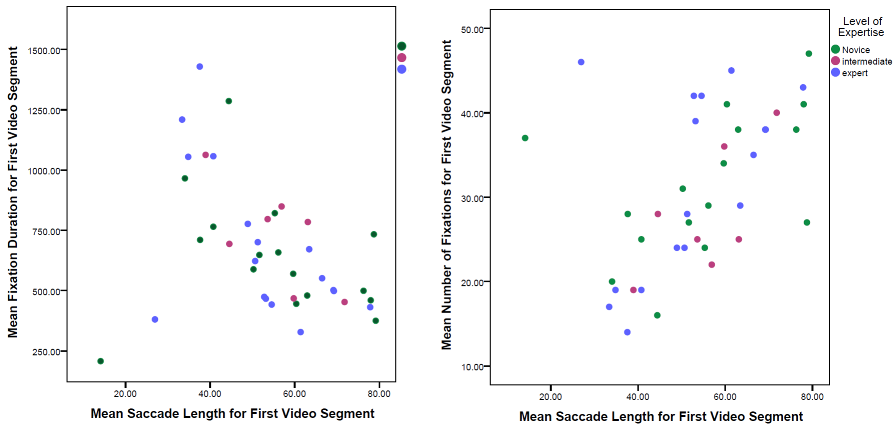

Fig. 1: Correlation of eye movements for the first video segment: Mean fixation duration and mean saccade length (left) and mean number of fixation numbers and mean saccade length (right). Level of expertise is marked on the graph and shows its effect on both correlations.

fixation information for 5 second intervals in each segment. Even though, the video was not highly dynamic, differences in the areas of interest with respect to level of expertise over time is apparent. Figure 2 shows the differences in AOI size as well as clustering as a result of level of expertise. Here, we notice that the overall positioning of the AOIs has a high overlap, which is due to the surgical procedure remaining in a condensed area of the images (i.e. relatively static). However, we can see that the overall number of AOI clusters for the novices is higher than that of the other groups, with 39 clusters for novices, 23 for the intermediates, and 29 for the experts. To an extent, the size of the "main" or largest clusters are also larger for the novices compared to both intermediates and experts. Of all the groups, the intermediates had both the fewest and the smallest number of clusters. These trends in the size of the clusters were also apparent in the first segment, though there was an exception regarding number of clusters: with 39, 23, and 29 for novices, intermediates, and experts respectively.

4 Discussion

In this study, we used eye tracking technology to quantify neurosurgeons eye movements when watching video segments of a microscopic brain surgery. Series of eye fixations (visual processing) and saccades (visual search) characteristics were used to analyze and precisely evaluate neurosurgeons visual attention patterns. These eye movement behaviors gave us a unique opportunity to capture the nuanced differences among different level of experienced surgeons when attending to even a short segment of aneurysm treatment.

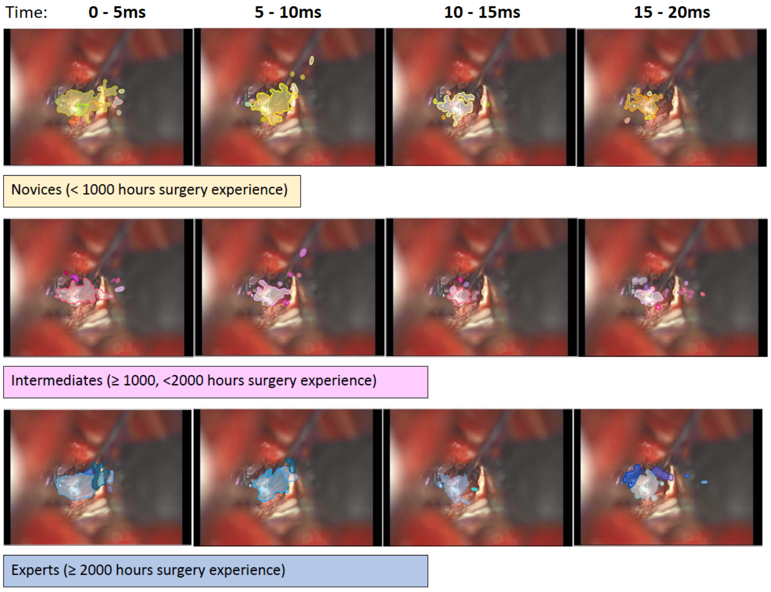

Fig. 2: Comparison of fixation patterns between novice, intermediate, and advanced microsurgeons over time course of a microsurgery video.

Here, we show that the gaze measures were not affected by the expertise of the surgeons, though there may be an effect of expertise on the relationship between fixations and saccades, such that experts employ both shorter and fewer fixations in association with longer saccades. Thus, our results do not support the general findings, mainly that experts employ shorter fixations and longer saccades [6][13] [8] [15], however further testing may indicate whether expertise effects the relationship between fixations and saccades.

There are important dependencies contingent on the progress of the treatment and this is reflected in the subject viewing behavior in the first stage of surgery compared to last stage. We exposed these intricate relationships by analyzing eye-movements in time. We found that initially when watching the surgery segments all participants had significantly higher and longer fixations for the first segment compared to last segment. Though fixation amount and duration was not dependent on group, we can see a highest difference (mean difference of 8.77) in the mean number of fixations for the novice group which could indicate they employed less fixations over the time course as a result of *acquisition* to the task. In this sense, *acquisition* to the task in the video was also apparent for the experts, though to a lesser extent (mean difference of 4).

Number of fixations can be indicative of the information-reduction hypothesis, meaning experts may reduce the number of fixations on areas of redundant information and centralize their fixations on only the relevant areas [6]. This model could also be relevant for viewing behavior over time as we noticed by the trends in our data, such that selective attention becomes more precise over time, though possibly faster for experts. The number of fixations as well as the fixation duration can support the information-reduction hypothesis. In our data, the mean fixation duration data supported this hypothesis as well though not in support of expertise effects; the fixation duration was significantly shorter for all groups in the last video segment, though no significant differences between groups. Rather these results could suggest changes in long-term working memory in general [6].

It was interesting that expertise affected the eye-movements mainly regarding fixation areas. The behavior of the participants did not differ in general on eye movement level; however, the size and the clustering of the AOIs differed between groups and over time as well. Although, the video images were highly localized and exhibited minimal tool use, there were moments in time where novices' fixations regions revealed attention to the tool and sometimes more clusters outside the *central focal area* when compared to intermediates and experts in the same time interval. Overall, this behavior was found for both segments. Concerning the experts and the novices, their areas of attention can be indicative of their *processing* of the task at hand, where experts tend to attend to areas as *chunks* close to each other and novices determine a large main area to selectively focus then scatter attention to smaller areas near the outer regions of the *central focal area*.

Further, a more in-depth analysis can offer better insight into areas of interest as determined by level of expertise; Such an analysis could be comparing the area of AOI clusters between groups or even center extraction of clusters and then looking at the distance from the nearest expert cluster From our AOI data, we see that the intermediate group employed fewer and more concentrated Fixations. This behavior can be attributed to two *things*, one being that the surgeons grouped as intermediate though less experienced in terms of total hours of surgery, *still have plenty of experience. They may have also taken the experiment more seriously compared to the experts, especially those that were more advance. This Source, citation,* also noticed that experts were more likely to not participate in studies seriously when compared to their less experienced counterparts. The second *thing* for the intermediate eye movement behavior being that there were only seven participants compared to the 16 and 17 in the novice and expert groups. Recruiting more participants meeting the criteria for an intermediate microneurosurgeon sample could possibly provide a more comprehensive understanding of AOIs for this level of expertise.

The eye movement behavior of all groups could have been influenced by the notion that subjects were watching a video of microneurosurgeries and not actually performing the surgeries themselves. Actually performing the surgeries creates an environment that facilitates context based behaviors, such as hand or eye movements *(Find source)*. Simply watching a video may not be enough to evoke the natural task-based behaviors. *find source supporting this.* From our data we had initially suspected that we would not see precise surgery based eye movements but rather general eye movements indicative of expertise but more importantly we expected to see more eye movement behavior related to interpretation of the task. Generally, a microneurosurgeon is debriefed in advance concerning what type of surgery he needs to perform. For our data, we asked the subjects to watch the video and interpret the surgery. In this context, the visual behaviors may not be indicative of procedural tasks but rather *scanning and evaluating* tasks though can also necessary for surgical performance as well.

Visual attention skills, both procedural and analytical, are essential for developing and mastering surgical techniques [**?**,**?**,17, **?**]. In micro-neurosurgery, moreover, the learning curve for eye-hand coordination is long and may depend on the subjective psychomotor abilities. Surgeons have to learn to scan across the operating field while concentrating on tiny objects in the operating field. As such, they have to maintain attention to follow the procedure and avoid obstacles as safely moving the instrument through delicate structures. Our work employed gaze-based measures that are more likely to be indicative of visual search strategies for interpreting and analyzing images. Visual search is also a concept in the medical domain that has been of interest since expertise can also be trained; for instance, interpreting radiographs.

4.1 future work

Gegenfurtner and colleagues [6] promote that for any skill learning, training aids often direct novices concerning what to focus on and in which order they should focus on specific information. When students are trained to employ expert visual search strategies, improvements in their perceptual performance were measurable [7]. For instance, the **the scanpaths of experts can guide learners' visual attention to relevant areas, otherwise known as Eye Movement Modeling Examples [8].** Thus, gaze-based adaptive training techniques are of interest because this information can facilitate learning.

In this study, we also explore the differences in the eye movement behavior of experts and novices in the medical fields that deal with image interpretation. We aim to develop a technique to investigates and model the change in viewing behavior during expertise development so it can be used for assessment of medical practitioners and furthermore can be employed as a learning tool for students.

The prospective study will look at radiograph interpretation among dentistry students, novice and advanced, and certified practitioners and will attempt to model the development. Radiographs and other medical image interpretation is a crucial component to diagnostics and general healthcare. Both medical and dental students are required to undergo rigorous training in order to develop the skills for fast and accurate interpretation of these images. Reading radiographs, especially for dentists, is non-trivial due to the routine use of them as well as their likelihood for error [13].

The proposed study intends to evaluate massed practice, domain knowledge and task experience. Then, develop a low-cost intervention based on eye movement modeling. This intervention will be compared against other gaze-based interventions (e.g. EMME).

how can we improve the training of radiograph analysis and interpretation.

These results will then be indicative of eye movement behaviors we can expect to find regarding expert and novice dentists when employing visual search strategies on radiograph interpretation.

5 References

References

1. Charness, N., Reingold, E.M., Pomplun, M., Stampe, D.M.: The perceptual aspect of skilled performance in chess: Evidence from eye movements. Memory & cognition 29(8), 1146–1152 (2001)
2. Donald, J.J., Barnard, S.A.: Common patterns in 558 diagnostic radiology errors. Journal of medical imaging and radiation oncology 56(2), 173–178 (2012)
3. Eivazi, S., Bednarik, R., Tukiainen, M., von und zu Fraunberg, M., Leinonen, V., Jääskeläinen, J.E.: 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. pp. 377–380. ACM (2012)
4. Eivazi, S., Hafez, A., Fuhl, W., Afkari, H., Kasneci, E., Lehecka, M., Bednarik, R.: Optimal eye movement strategies: a comparison of neurosurgeons gaze patterns when using a surgical microscope. Acta Neurochirurgica 159(6), 959–966 (2017)
5. Eivazi, S., Slupina, M., Fuhl, W., Afkari, H., Hafez, A., Kasneci, E.: Towards automatic skill evaluation in microsurgery. In: Proceedings of the 22nd International Conference on Intelligent User Interfaces Companion. pp. 73–76. ACM (2017)
6. Gegenfurtner, A., Lehtinen, E., Säljö, R.: Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. Educational Psychology Review 23(4), 523–552 (2011)
7. van der Gijp, A., Ravesloot, C., Jarodzka, H., van der Schaaf, M., van der Schaaf, I., van Schaik, J., ten Cate, T.J.: How visual search relates to visual diagnostic performance: a narrative systematic review of eye-tracking research in radiology. Advances in Health Sciences Education pp. 1–23 (2016)
8. Jarodzka, H., Balslev, T., Holmqvist, K., Nyström, M., Scheiter, K., Gerjets, P., Eika, B.: Conveying clinical reasoning based on visual observation via eye-movement modelling examples. Instructional Science 40(5), 813–827 (2012)
9. Kübler, T., Eivazi, S., Kasneci, E.: Automated visual scanpath analysis reveals the expertise level of micro-neurosurgeons. In: MICCAI Workshop on Interventional Microscopy (2015)
10. Kübler, T.C., Sippel, K., Fuhl, W., Schievelbein, G., Aufreiter, J., Rosenberg, R., Rosenstiel, W., Kasneci, E.: Analysis of eye movements with eyetrace. In: International Joint Conference on Biomedical Engineering Systems and Technologies. pp. 458–471. Springer (2015)
11. Kundel, H.L., La Follette Jr, P.S.: Visual search patterns and experience with radiological images 1. Radiology 103(3), 523–528 (1972)
12. Kundel, H.L., Nodine, C.F., Conant, E.F., Weinstein, S.P.: Holistic component of image perception in mammogram interpretation: gaze-tracking study 1. Radiology 242(2), 396–402 (2007)
13. Nodine, C.F., Kundel, H.L., Mello-Thoms, C., Weinstein, S.P., Orel, S.G., Sullivan, D.C., Conant, E.F.: How experience and training influence mammography expertise. Academic radiology 6(10), 575–585 (1999)

14. Nodine, C.F., Mello-Thoms, C.: The nature of expertise in radiology. Handbook of Medical Imaging. SPIE (2000)
15. Tien, T., Pucher, P.H., Sodergren, M.H., Sriskandarajah, K., Yang, G.Z., Darzi, A.: Differences in gaze behaviour of expert and junior surgeons performing open inguinal hernia repair. Surgical endoscopy 29(2), 405–413 (2015)
16. Uemura, M., Tomikawa, M., Kumashiro, R., Miao, T., Souzaki, R., Ieiri, S., Ohuchida, K., Lefor, A.T., Hashizume, M.: Analysis of hand motion differentiates expert and novice surgeons. Journal of Surgical Research 188(1), 8–13 (2014)
17. Wilson, M., McGrath, J., Vine, S., Brewer, J., Defriend, D., Masters, R.: Psychomotor control in a virtual laparoscopic surgery training environment: gaze control parameters differentiate novices from experts. Surgical endoscopy 24(10), 2458–2464 (2010)