Eye Movement Traits in Differentiating Experts and Laymen

Katarzyna Harezlak[1(✉)], Pawel Kasprowski[1], and Sabina Kasprowska[2]

[1] Institute of Informatics, Silesian University of Technology, Gliwice, Poland
katarzyna.harezlak@polsl.pl
[2] Department of Radiology, District Hospital of Orthopedics and Trauma Surgery, Piekary Slaskie, Poland

Abstract. There is much research indicating that eye tracking methods are a promising approach which can be used in revealing experts' visual patterns and acquiring information regarding their subconscious behaviour while making decisions in professional tasks. The studies presented in this paper extend the aforementioned investigations and were aimed at checking the possibility of differentiating experts and laymen based on their eye movement characteristics. For this purpose, an experiment in the radiology field was chosen. The studies revealed not only significant differences between visual patterns of the analysed groups but also demonstrated that distinguishing experts from novices based on their eye movements is feasible. The classification performance was high and, dependent on the method applied for defining the test set, amounted to 85% or 93% correctly-classified subjects. The investigation concerning the possibility of recognizing who was performing the experiment task—an expert or layman—showed that dependent on the radiology image explored—the performance in the majority of cases was between 79% and 93%.

1 Introduction

Industry, education, healthcare, transport or economics are only a few examples of human life areas in which a group of skilled people called experts may be distinguished. A huge effort is made to create such a group as large as possible - whole educational processes may be classified as such an activity. However, even after graduating, people still have to undertake steps aimed at further skill improvement. Certainly, experience gained during daily work and also cooperation with professionals play an important role in this case. The possibility of observing and being guided by more experienced people may speed up the development progress. Nevertheless, at times transferring the knowledge is a difficult task because skilled behaviour is subconscious, as people act instinctively. For this reason other methods for experts' knowledge acquisition have been intensively searched for.

Among potential solutions, the eye tracking technique seems to be a promising approach - many studies have been devoted to revealing experts' visual

© Springer International Publishing AG 2018
A. Gruca et al. (eds.), *Man-Machine Interactions 5*,
Advances in Intelligent Systems and Computing 659, DOI 10.1007/978-3-319-67792-7_9

patterns - the majority of which concerns medicine. The contribution of eye movement research to the study of expertise in this field was broadly discussed in [15]. Additionally, exhaustive analysis of the research performed in the most explored branch of medicine - radiology - was presented in [5]. The examination of eye movements may also be found *inter alia* in the case of colonoscopy experts in [2], surgeons in [10] and in the anaesthesia field [17].

Obviously, such research studies are also presented in other disciplines. A comprehensive review of eye tracking applications in learning - including patterns of information processing, effects of instructional design, effects of learning strategies or conceptual development - is provided in [11]. Moreover, there are works exploring the usefulness of eye movement analysis in the expertise disclosed in regard to pilots [19], sportsmen [13] or software designers [12].

The research described in this paper was undertaken to extend the aforementioned investigations and was aimed at checking the possibility of differentiating experts and laymen based on their eye movement characteristics. For this purpose, an experiment in the radiology field was chosen. However the presented method may also be utilized in other applications. The description of the performed work is organised as follows: The basis of eye tracking methods are provided in the Sect. 2. The experiment used in the studies is introduced in Sect. 3. The analysis of collected data and obtained results is presented in Sect. 4. Finally, the research is concluded in Sect. 5.

2 Basics of Eye Movement Analysis

The way of eye function is strictly related to its structure, which is responsible for producing continuous images that are instantly transmitted to the brain. To describe briefly: the light entering the eye falls on the retina containing the photoreceptors, among which the most sensitive are placed in a small area - the fovea - ensuring sharp central vision. The fovea is surrounded by a larger peripheral area that delivers information of low resolution. For this reason, the eye, to obtain a sharp image of a whole observed scene, has to be in constant movement, within which two main components may be distinguished. One of them is a fixation, when the eye is almost stable, focused on a small area to acquire information about this part of the scene. Fixations are interlaced with saccades - quick movements to another place of the scene, during which no data is taken [8]. These two components constitute a base for eye movement analysis.

However, the first step in eye movements processing is their registration with usage of a specialized device called an eye tracker. Eye trackers commonly used at present are equipped with small cameras and infrared light sources, which directed to the eye, facilitate recording its images (Fig. 1).

Such images are subsequently transformed to obtain the pupil's center and calculate changes in its position with time. Based on this data, gaze directions are ascertained. Between these two actions - calculating a pupil's center position and estimating a point of gaze - a calibration step is required. It is necessary to build a mapping function (usually a regression model), which associates coordinates

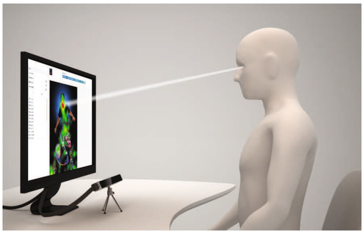


Fig. 1. Example of an eye tracker - The EyeTribe system [18] - the device used during the experiment

of pupil's position provided by the eye tracker in its internal system (x_e, y_e) to values adequate to the scene's coordinate system (x_g, y_g) [3,7]. Equation 1 presents an example of a polynomial regression:

$$x_g = A_x x_e^2 + B_x y_e^2 + C_x x_e y_e + D_x x_e + E_x y_e + F_x \tag{1}$$

$$y_g = A_y x_e^2 + B_y y_e^2 + C_y x_e y_e + D_y x_e + E_y y_e + F_y$$

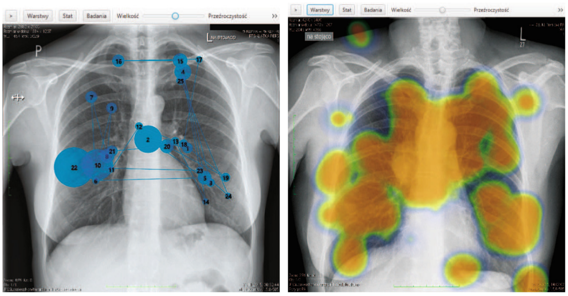
Before the application of such a model for the whole eye tracking session, it should be verified in terms of the provided results accuracy. The method, which may be used for this purpose is Root Mean Squared Error (RMSE—Eq. 2) measuring the difference between the target location and a gaze point calculated from the model [9].

$$RMSE = \sqrt{\frac{1}{n}\sum_{i}(y_i - \widehat{y_i})^2} \tag{2}$$

where y_i is an observed value and $\widehat{y_i}$ a value calculated by the model.

Each value obtained by means of the determined calibration model represents one recording and in order to form a fixation or saccade - sets of recordings - appropriate methods have to be used. Among them the most popular are dispersion- and velocity-based fixation identification algorithms. The first one (I-DT) identifies fixations as groups of consecutive points within a dispersion defined by a chosen threshold. The second algorithm (I-VT) classifies each point as a fixation or saccade based on a velocity threshold: if the point-to-point velocity is below the defined threshold, it becomes a fixation point, otherwise it is classified as a saccade [6,16].

Parts of eye movement signal selected in this way may be subsequently used in a quantitative and qualitative analysis and visualised by means of various methods such as scan paths or heatmaps (Fig. 2).



(a) (b)

Fig. 2. Examples of eye movement visualisation - a scan path on the left and a heat map on the right

3 Experiment

As mentioned above, the research studies presented in this paper were undertaken to investigate the possibilities of differentiating skilled and unskilled people when dealing with a task specific for a particular domain. For this purpose two groups of people were invited to take part in our radiological experiment. They were:

– Laymen [L] - seven people not working as radiologists (most of them were sales representatives of companies selling medical equipment),
– Experts [S] - seven skilled specialists in the field of radiology with over 10 years of experience.

All participants signed a consent form allowing publication of their eye tracking data and results obtained using them, anonymously. During the experiment participants were presented 12 chest X-Rays with and without various diseases. Every image was shown twice: for 6 and 12 s respectively. After each image presentation, a participant was expected to make a diagnosis based on four proposed answers. After the second image display, participants had the opportunity to change the decision previously taken. Participants' eye movements were recorded by means of the Eye Tribe eye tracker with 60 Hz sampling rate. Images were presented on the 24 in. screen. The users were sitting centrally at a distance of 60 cm (see Fig. 1).

Each recorded session started with the calibration process and only participants with a calibration error of less than 1° were allowed to continue the experiment. Fixations were selected with the usage of the I-DT algorithm. Additionally, in order for a sequence of gaze points to be qualified as a fixation it had to last more than 150 ms. Subsequent fixations separated by less than a 40 ms gap and located within one degree of visual angle were merged together.

Sets of fixations obtained for all participants individually were further used in the visual patterns analysis utilizing several eye movement metrics: total observation time (*elaspsed*), total number of fixations (*fixNum*), averaged fixation duration (*fixAvgDur*), total length of saccades (*sacLen*) and averaged length of saccades (*sacAvgLen*). They were calculated for the first and second image observation independently as well as for both decision making parts.

4 Data Analysis

The analysis of experts' and laymen's visual patterns was started from the comparison of values averaged for each metric, each group and each image presentation separately. The obtained results are presented in Table 1. As it may be noticed, in many cases the differences between the studied groups are evident. Experts eye movements are characterized by a greater number of fixations, shorter fixation duration and longer saccades. The significance of all these differences were confirmed statistically by the use of the t-test. The null hypothesis stating that the mean values for experts and laymen are equal was rejected for each of the aforementioned metrics (Table 1 - upper part, column titled 'p-values (t-value)'). Similarities in eye movement features were only revealed in the case of the diagnosis made after the second image presentation (Table 1 - lower part) and in these cases the null hypothesis was not rejected. The same relates to the averaged fixation duration at the time of an image assessment after its first display (Table 1 - lower part).

The similarities in eye movement patterns disclosed after the second observation may result from the fact that participants belonging to both groups had previously seen this part of the presentation and they knew its content - proposed answers and layout - thus for all of them it was repetition of the same, previously completed work.

Obtaining the confirmation that experts and novices are characterised by different eye movement characteristics was the motivating factor to undertake a subsequent step of the eye movement analysis. During it we checked the possibility of predicting qualifications of a person based on his/her way of image observation. The *random forest* classifier - its implementation in 'randomForest' package available in R with default parameters and the number of trees set to 500 - was used for this purpose. Samples for train and test sets were vectors formed from values calculated for all previously-described metrics. One vector was related to one person and one image. Thus, for 14 participants and 12 images there were 168 samples obtained.

During the classification process we applied *the leave-one-out cross-validation* method constituting a train set from 167 samples and leaving one for the test set. Hence, each of 168 samples was classified as being an expert or layman based on the remaining data using the following formula:

$$c(i) = \begin{cases} 1 & \textit{predicted class was equal to the true sample label} \\ 0 & \textit{otherwise} \end{cases} \tag{3}$$

These results were subsequently summed up and normalized in terms of samples belonging: (1) to the same participant (14 sets) and (2) recorded for the

Table 1. Averaged values of eye movement measures for both image presentations

Image observation parts						
Feature	First image presentation			Second image presentation		
	Laymen	Experts	p-value (t-value)	Laymen	Experts	p-value (t-value)
Elapsed (ms)	4568.28 (1739.82)	5651.07 (1011.62)	$p < 0.001$ (−4.86)	6034.14 (4004.04)	8899.14 (3552.51)	$p < 0.001$ (−4.84)
fixNum	5.54 (3.23)	9.18 (2.44)	$p < 0.001$ (−8.08)	7.59 (5.93)	13.42 (5.94)	$p < 0.001$ (−6.29)
fixAvgDur (ms)	718.61 (491.07)	466.41 (127.45)	$p < 0.001$ (4.477)	695.87 (440.19)	495.86 (140.76)	$p < 0.001$ (3.89)
Saclen (deg)	29.57 (26.01)	66.41 (24.42)	$p < 0.001$ (−9.35)	45.54 (46.89)	101.63 (54.94)	$p < 0.001$ (−7.04)
sacAvgLen (deg)	4.42 (2.54)	7.05 (1.64)	$p < 0.001$ (−7.85)	4.62 (2.87)	7.03 (2.11)	$p < 0.001$ (−6.12)
Decision making parts						
Feature	First image presentation			Second image presentation		
	Laymen	Experts	p-value (t-value)	Laymen	Experts	p-value (t-value)
Elapsed (ms)	2774.16 (1739.37)	3819.61 (1791.59)	$p < 0.001$ (−3.79)	1845.12 (1298.32)	1756.13 (1109.77)	$p = 0.638$ (0.47)
fixNum	2.91 (2.05)	4.19 (2.43)	$p < 0.001$ (−3.65)	2.17 (1.55)	2.241 (1.81)	$p = 0.795$ (−0.26)
fixAvgDur (ms)	875.19 (609.28)	819.74 (511.06)	$p = 0.529$ (0.63)	667.83 (530.93)	708.29 (415.06)	$p = 0.588$ (−0.54)
sacLen (deg)	8.04 (9.51)	13.14 (10.97)	$p = 0.002$ (−3.19)	5.89 (7.61)	5.86 (8.61)	$p = 0.986$ (0.02)
sacAvgLen (deg)	1.97 (1.69)	2.61 (1.48)	$p = 0.010$ (−2.61)	1.86 (1.87)	1.61 (1.78)	$p = 0.387$ (0.87)

same image (12 sets). The accuracy of the classification for the grouped sets was calculated in accordance with Eq. 4:

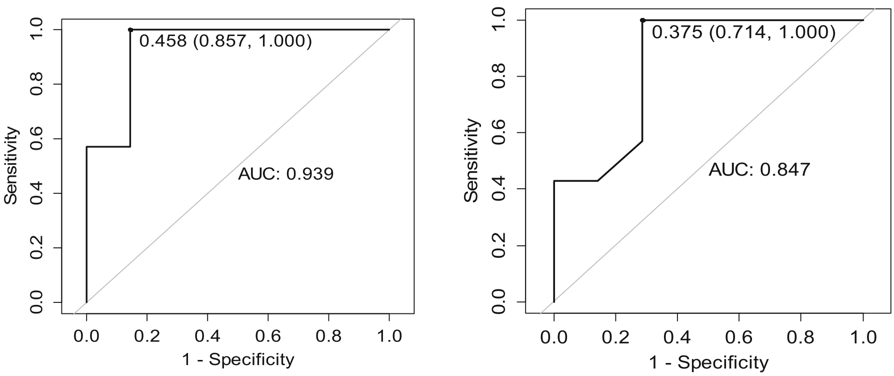
$$accuracy = \frac{\sum_{i=1}^{S} c(i)}{S} \tag{4}$$

where S denotes the number of samples in a set—12 for participant-related collections and 14 for image-related ones.

Additionally, for the participant-related sets, the acceptance rate a_i for different thresholds th ranging from 0 to 1 was calculated (5).

$$a_i(th) = \begin{cases} 1 & p_i \geq th \\ 0 & otherwise \end{cases} \tag{5}$$

where i is the number of sets and p_i is the accuracy calculated according to Eq. 4. Values evaluated in this way were utilized to ascertain the performance of the classifier. For this purpose *sensitivity* (True Positive Rate (TPR)) and *specificity* (True Negative Rate (TNR)) were determined [14]. The trade-off between these factors was explored in the ROC analysis and was presented in the form of the ROC curve in Fig. 3(a), together with the cut-off value and area under the curve (AUC) - an indicator summarizing the performance of a classifier into a single metric [4]. It is visible that the curve tends to bend to the upper left, thus it may be reasoned that the classification model is able to differentiate the positive and negative data well. It is confirmed by the determined AUC value, which being equal to 0.939 represents an excellent performance [1].



(a) leave-one-out cross-validation (b) leave-one-subject-out cross-validation

Fig. 3. The ROC curve presenting the trade-off between sensitivity and specificity for various threshold values

Taking the estimated cut-off point and related threshold (0.458) into account, the final accuracy of the classification amounted to 93% of correctly classified subjects.

The results for sets, each of which consisted of recordings related to one image, are presented in the chart visible in Fig. 4. It may be noticed that for 7 out of 12 images observers were classified with the efficiency at least on 85% level and the next two reached almost 80%. In the case of the remaining three images it was between 57% and 71%.

These outcomes, apart from confirming the good performance of the classification model, also revealed that some type of images are more useful in distinguishing observers than others. The X-Rays 5 and 6 characterized by the lowest classification accuracy consisted of abnormalities (namely *pneumothorax* and *emphysema*) which were - contrary to the abnormalities in other images - darker than the background ("clearer" in radiologic language). It may be reasoned that radiology specialists concentrated on searching for whiter anomalies

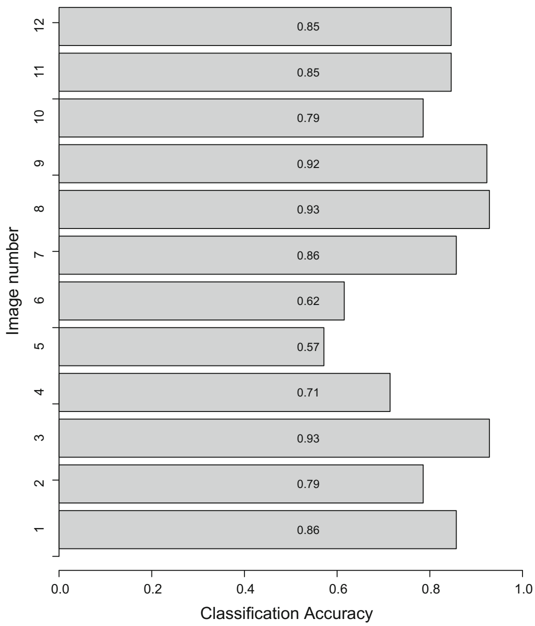


Fig. 4. Percentage of correctly classified observers for various images

("shadows") as more frequent and the darker ones were searched during the second pass. It might explain why the specialists' scan-paths are more similar to the novices ones. However, further studies are necessary to confirm this finding.

Undertaking the task of differentiating observers it was assumed that an explored image influences subject's eye movement characteristics. However, it cannot be excluded and even may be expected that people keep their visual patterns regardless of changing tasks. For this reason, the efficiency of the classification was also checked with the usage of *the leave-one-subject-out cross-validation* method. It means that recordings belonging to each participant were omitted in a corresponding training set and used only for the model testing purpose. The results obtained in this case are presented in Fig. 3(b) also in the form of the ROC curve. It may be seen that the AUC value is lower than in the former case, yet still on the high level. For the estimated cut-off point and corresponding threshold (0.375), the final classification accuracy reached 85%.

5 Conclusions

The research presented in this paper was aimed at investigating the possibilities of differentiating two groups of people: experts and novices in the realm of a specific domain. Our studies were focused on the radiology field, however application of the research method used is not limited to this area. This method utilisation demonstrated both the significant differences between visual patterns of the analysed groups and that distinguishing experts from novices based on their eye movements is feasible. The performance of the classifier assessed based on the area under the ROC curve turned out to be very high: 85% correctly classified participants when the *leave-one-subject-out* cross-validation was applied and 93% for *leave-one-out* one. Another interesting subject to study was ascertaining if it is possible to recognise who was assessing an image - an expert or layman. Also this time, the classification efficiency - meant as a percentage of the correctly determined observers' classes - was good: for 75% of images it was approximately 79% and higher, up to 93%.

All these outcomes are promising and lead to the conclusion that in the case of a well-adjusted experiment it is possible to identify a class of its participants based on eye movement characteristics if a distinction is made between experts and laymen. Obviously, there are other levels of expertise between these two groups. The performance of the method should also be checked against them, which is planed as future studies. Probably, an extended set of eye movement features will be required to achieve this goal.

Acknowledgements. The research presented in this paper was partially supported by the Silesian University of Technology Rector's Pro-Quality Grant 02/020/RGJ17/0103 and by the Silesian University of Technology grant BK/263/RAu2/2016.

References

1. Bekkar, M., Djemaa, H.K., Alitouche, T.A.: Evaluation measures for models assessment over imbalanced data sets. J. Inf. Eng. Appl. **3**(10), 27–38 (2013)
2. Bernal, J., Sánchez, F.J., Vilariño, F., Arnold, M., Ghosh, A., Lacey, G.: Experts vs. novices: applying eye-tracking methodologies in colonoscopy video screening for polyp search. In: ETRA 2014, Safety Harbor, USA, pp. 223–226 (2014)
3. Cerrolaza, J.J., Villanueva, A., Cabeza, R.: Study of polynomial mapping functions in video-oculography eye trackers. ACM Trans. Comput. Hum. Interact. **19**(2), 10:1–10:25 (2012)
4. Fielding, A.H., Bell, J.F.: A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ. Conserv. **24**(01), 38–49 (1997)
5. 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. Adv. Health Sci. Educ. **22**, 1–23 (2016)
6. Harezlak, K., Kasprowski, P.: Evaluating quality of dispersion based fixation detection algorithm. In: ISCIS 2014, Krakow, Poland, pp. 97–104 (2014)
7. Harezlak, K., Kasprowski, P., Stasch, M.: Towards accurate eye tracker calibration-methods and procedures. Procedia Comput. Sci. **35**, 1073–1081 (2014)
8. Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., Van de Weijer, J.: Eye Tracking: A Comprehensive Guide to Methods and Measures. Oxford University Press, Oxford (2011)
9. Kasprowski, P., Harezlak, K., Stasch, M.: Guidelines for the eye tracker calibration using points of regard. In: Piętka, E., Kawa, J., Wieclawek, W. (eds.) Information Technologies in Biomedicine, Volume 4, AISC, vol. 284, pp. 225–236. Springer, Cham (2014)

10. Khan, R.S.A., Tien, G., Atkins, M.S., Zheng, B., Panton, O.N.M., Meneghetti, A.T.: Analysis of eye gaze: do novice surgeons look at the same location as expert surgeons during a laparoscopic operation? Surg. Endosc. **26**(12), 3536–3540 (2012)
11. Lai, M.L., Tsai, M.J., Yang, F.Y., Hsu, C.Y., Liu, T.C., Lee, S.W.Y., Lee, M.H., Chiou, G.L., Liang, J.C., Tsai, C.C.: A review of using eye-tracking technology in exploring learning from 2000 to 2012. Educ. Res. Rev. **10**, 90–115 (2013)
12. Nivala, M., Hauser, F., Mottok, J., Gruber, H.: Developing visual expertise in software engineering: an eye tracking study. In: EDUCON 2016, pp. 613–620. Abu Dhabi, United Arab Emirates (2016)
13. Panchuk, D., Vine, S., Vickers, J.N.: Eye tracking methods in sport expertise. In: J. Baker, D. Farrow (eds.) Routledge Handbook of Sport Expertise, Routledge, pp. 176–187 (2015)
14. Provost, F.J., Fawcett, T.: Analysis and visualization of classifier performance: comparison under imprecise class and cost distributions. In: KDD 1997, vol. 97, Newport Beach, pp. 43–48 (1997)
15. Reingold, E.M., Sheridan, H.: Eye movements and visual expertise in chess and medicine. In: Oxford Handbook on Eye Movements, pp. 528–550 (2011)
16. Salvucci, D.D., Goldberg, J.H.: Identifying fixations and saccades in eye-tracking protocols. In: ETRA 2000, Palm Beach Gardens, USA, pp. 71–78 (2000)
17. Schulz, C., Schneider, E., Fritz, L., Vockeroth, J., Hapfelmeier, A., Brandt, T., Kochs, E., Schneider, G.: Visual attention of anaesthetists during simulated critical incidents. Br. J. Anaesth. **106**(6), 807–813 (2011)
18. The eye tribe: the eye tribe system (2016). theeyetribe.com. Accessed Mar 2017
19. Yang, J.H., Kennedy, Q., Sullivan, J., Ronald, D., Fricker, J.: Scan patterns on overland navigation in varying route difficulty: is total-flight-hours (TFH) a good measure of expertise? In: HFES 2012, vol. 56, no. 1, pp. 1406–1410 (2012)