Guidelines for eye tracker calibration using points of regard

Conference Paper · June 2014

DOI: 10.1007/978-3-319-06596-0_21

Pawel Kasprowski
Silesian University of Technology

Katarzyna Harezlak
Silesian University of Technology

Mateusz Stasch
Silesian University of Technology

This is a pre-print. The original version was published in Springer, Advances in Intelligent Systems and Computing Volume 284, 2014, and is available via
http://dx.doi.org/10.1007/978-3-319-06596-0_21

Guidelines for the eye tracker calibration using points of regard

Pawel Kasprowski, Katarzyna Harężlak, and Mateusz Stasch

Silesian University of Technology, Institute of Informatics,
kasprowski@polsl.pl, katarzyna.harezlak@polsl.pl, mateusz.stasch@polsl.pl

Abstract. Eye movement data may be used for many various purposes. In most cases it is utilized to estimate a gaze point - that is a place where a person is looking at. Most devices registering eye movements, called eye trackers, return information about relative position of an eye, without information about a gaze point. To obtain this information, it is necessary to build a function that maps output from an eye tracker to horizontal and vertical coordinates of a gaze point. Usually eye movement is recorded when a user tracks a group of stimuli being a set of points displayed on a screen. The paper analyzes possible scenarios of such stimulus presentation and discuses an influence of usage of five different regression functions and two different head mounted eye trackers on the results.

Keywords: eye movements, eye tracking, calibration

1 Introduction

Eye movement data may be used for many various purposes. In most cases it is used to estimate a gaze point - that is a place where a person is looking at. Most devices registering eye movements, called eye trackers, return information about relative position of an eye, without information about the gaze point. To obtain this information it is necessary to build a function that maps an output from the eye tracker to horizontal and vertical coordinates of a gaze point. It is typically done using information about eyes position when an examined person is looking at a set of points (called Points of Regard or PoRs). There are several problems that must be addressed when preparing such a function:

- How many points to use
- How to locate the points - i.e. point's layout
- How long to present the stimulus in each point
- What type of mapping function to use
- Which measurements to use as mapping function input
- How to check the validity of the function

The paper discusses some of issues enlisted above using two different eye trackers and a considerable amount of data registered during a couple of sessions.

This is a pre-print. The original version was published in Springer, Advances in Intelligent Systems and Computing Volume 284, 2014, and is available via http://dx.doi.org/10.1007/978-3-319-06596-0_21

The main contribution of the paper is defining guidelines which may be used when preparing own calibration procedures.

While it seems obvious that more calibration points and a longer presentation of points gives more data to build mapping model, it should also be taken into account that too complicated calibration procedure is inconvenient for participants. Participants may be tired or annoyed with a long preparation phase and tend to loss their concentration during the calibration process itself and - what may be even worse - during the subsequent experiment. Therefore, the main objective of the calibration step is to gather and analyze sufficient amount of data during a procedure that is as short and simple as possible.

The pattern which is most widely (eg. [13][18]) used to present stimuli to participant is the square grid, which typically consist of 9 to 25 points. Ramanauskas et al. [13] confirm what was mentioned above, that higher number of points in this configuration usually results with better performance. Ohno et al. [19] propose an eye tracking system which uses only 2 calibration points, however this setup requires to maintain very strict relations between a camera, an IR illuminator and an eye.

2 Experiment setup

There were two different eye trackers used during the experiment.

The first one was a head mounted Jazz-Novo eye tracker (product of Ober-consulting) that records eye positions with 1000Hz. It uses Direct Infra Red Oculography (IROG) and utilizes pairs of IR emitters and sensors. The optoelectronic transducers are located between the eyes, thus hiding the sensor assembly behind the "shadow" of the nose.

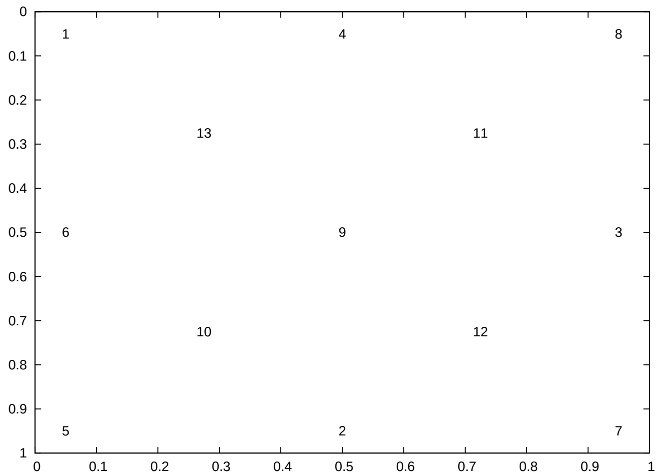
The second eye tracker was the VOG head-mounted eye tracker developed with a single CMOS camera with USB 2.0 interface (Logitech QuickCam Express) possessing 352 x 288 sensor and lens with IR-Pass filter. The camera was mounted on the arm attached to head and was pointing at the right eye. The eye was illuminated with single IR LED placed off the axis of the eye that caused "dark pupil" effect, which was useful during a pupil detection. The system generated 20 - 25 measurements of a center of a pupil per second.

Both eye trackers were used in a usual points of regard calibration experiment. The participants of the experiment were looking at a stimulus presented on a screen. The stimulus was a circle pulsating on the screen to attract participant's attention. There were 29 different stimulus locations (Fig 1). The stimulus was displayed for about 3 seconds in each location. The order of stimuli presentation was the same for each session.

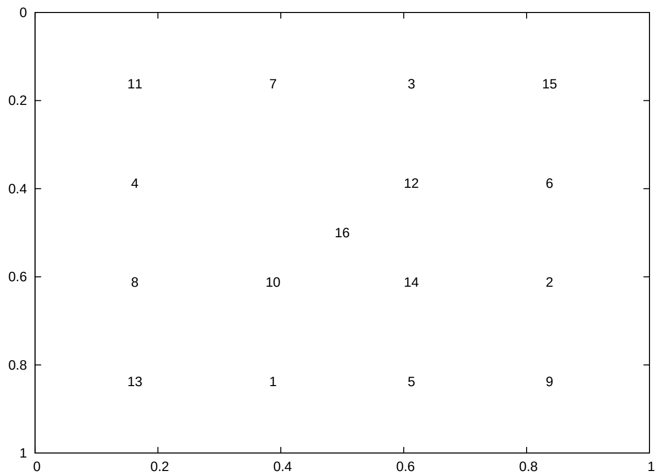
In both cases the experiment was done on a 1280 x 1024 (370mm x 295mm) flat screen. The eye-screen distance was 500mm, vertical gaze angle was 40 deg and horizontal gaze angle was 32 deg.

There were overall 88 sessions with 39 sessions for Jazz-Novo eye tracker and 49 sessions for VOG eye tracker.

This is a pre-print. The original version was published in Springer, Advances in Intelligent Systems and Computing Volume 284, 2014, and is available via
http://dx.doi.org/10.1007/978-3-319-06596-0_21



(a) Training points



(b) Testing points

Fig. 1. Points used during sessions

3 Calibration algorithms

The objective of an eye tracker is to determine a gaze point - that is a place where the user is looking at - with the best possible accuracy. There are three main factors that influence the accuracy of an VOG eye tracker:

– Quality of the registered image
– Quality of the algorithm used for extracting image features

This is a pre-print. The original version was published in Springer, Advances in Intelligent Systems and Computing Volume 284, 2014, and is available via http://dx.doi.org/10.1007/978-3-319-06596-0_21

– Quality of the algorithm that is used to map the extracted features to point of regard (gaze point)

It is worth emphasizing that the aim of described study is to analyze how to improve the third of the aforementioned elements, that is mapping image features to point of regard (PoR).

The data was divided into training and testing part. The first 13 stimulus locations (points) were used for training (Fig 1(a)). All tests and errors of the estimation were always calculated on the last 16 points (Fig 1(b)). Various combinations of 13 points were used to create 61 sets differing in the number of points used and their locations. The sets can be divided into five groups differing in the number of points. There were 19 sets of 5 points, 20 sets of 7 points, 9 sets of 9 points and 12 sets of 11 points prepared. Additionally, there was a one "full" set of all 13 points used. Due to limited space, the detailed description of the sets is not presented here.

Each of defined sets was then used to build a model mapping an eye tracker output to gaze coordinates on the screen. Such a model consists of two functions:

$$x_s = f(x_e, y_e) \tag{1}$$
$$y_s = f(x_e, y_e)$$

where x_e and y_e represent data obtained from an eye tracker and x_s and y_s are estimated gaze coordinates on a screen.

There are multiple possible regression functions to be used. In this study three types of such functions were used: the polynomial functions, the artificial neural network (ANN) and the support vector regression (SVR).

3.1 Polynomial functions

The most common choice for a calibration function is usage of polynomial functions, which can differ in the degree and number of terms. Two comprehensive studies analyzed possible solutions [3][1]. There were three classic functions used in this work: a linear function, a quadratic function and a cubic function with all possible terms.

Linear equation

$$x_s = A_x x_e + Bx y_e + C_x \tag{2}$$
$$y_s = A_y x_e + B_y y_e + C_y$$

Quadratic equation

$$x_s = A_x x_e^2 + B_x y_e^2 + C_x x_e + D_x y_e + E_x \tag{3}$$
$$y_s = A_y x_e^2 + B_y y_e^2 + C_y x_e + D_y y_e + E_y$$

Cubic equation

$$\begin{aligned} x_s = A_x x_e^3 + B_x y_e^3 + C_x x_e^2 y_e + D_x x_e y_e^2 + E_x x_e y_e \\ + F_x x_e^2 + G_x y_e^2 + H_x y_e + I_x y_e + J_x \end{aligned} \tag{4}$$

This is a pre-print. The original version was published in Springer, Advances in Intelligent Systems and Computing Volume 284, 2014, and is available via http://dx.doi.org/10.1007/978-3-319-06596-0_21

$$y_s = A_y x_e^3 + B_y y_e^3 + C_y x_e^2 y_e + D_y x_e y_e^2 + E_y x_e y_e + F_y x_e^2 + G_y y_e^2 + H_y y_e + I_y y_e + J_y$$

For each function the coefficients were calculated based on training points values using Levenberg-Marquardt algorithm.

3.2 ANN

The second type of function was an artificial neural network (ANN). An activation network with sigmoid function as an activation function was used. The network was trained until the total train error was lower than 0.1, using the Back Propagation algorithm, with normalized samples recorded during a session. Configuration of the network consisted of two neurons in the input layer, 10 neurons in one hidden layer and two neurons as the output. ANN has been already used in several eye tracing applications [6][17].

3.3 SVR

The Support Vector Regression (SVR) [14] was of the last of analyzed types. The RBF kernel with parameters $C = 10$ and $\gamma = 8$ was used. The similar function has been utilized for an eye tracker calibration in [12] and [9] but in completely different setups.

4 Results

Experiments were conducted for all sets of calibration points and for all sessions what gave 11.925 (39 sessions x 61 sets x 5 functions) models for Jazz-novo and 14.935 (49 sessions x 61 sets x 5 functions) models for VOG eye tracker. Every model represented combination of a function, a session and a set of training points. All models were checked using 16 testing points from the same session (Fig 1(b)). The error represented in degrees (E_{deg}) was calculated based on them. Additionally, it was determinant coefficient (R^2) calculated independently for both axes.

$$E_{deg} = \frac{1}{n} \sum_i \sqrt{(x_i - \widehat{x_i})^2 + (y_i - \widehat{y_i})^2} \tag{5}$$

$$R^2 = 1 - \frac{\sum_i (y_i - \widehat{y_i})^2}{\sum_i (y_i - \bar{y})^2} \tag{6}$$

where y_i, x_i represent an observed value, $\widehat{y_i}, \widehat{x_i}$ represent a value calculated by model and $\bar{y}, \bar{x}$ is the mean of observed values.

It must be emphasized, that it takes some time an eye to react to stimulus position change to fixate on another position. Such occurrence is called *saccadic latency* and lasts approximately 100-300 msec. In many cases this first fixation is not accurate and is corrected. During earlier experiments (not published yet) it was calculated that the safest range of measurements to include for further

This is a pre-print. The original version was published in Springer, Advances in Intelligent Systems and Computing Volume 284, 2014, and is available via http://dx.doi.org/10.1007/978-3-319-06596-0_21

studies is from 700 msec. to 1800 msec. after the stimulus position changed. Therefore, only these measurements were taken into account in both training and validation phases. It gave about 20-30 samples per point for VOG eye tracker and more than 1000 samples for the Jazz-Novo eye tracker. As so big number of features was computationally difficult for the ANN and the SVR methods, Jazz-Novo data was downsampled to 50Hz by calculating a median for every 20 subsequent measurements. Such downsampling process didn't affect results of the calibration, which was checked using the polynomial functions.

4.1 Model validity

While checking the correctness and the accuracy of defined models, it occurred that some models gave completely incorrect results for testing samples. It resulted in R^2 coefficient values to become lower than zero, which means that the modeled curve had errors higher than the average of testing samples. It was decided to reject such models as completely not feasible. Additionally, the number of rejections was calculated for each function and group (i.e. sets with the same number of points). The model was rejected if any of R_x^2 or R_y^2 coefficients values for that model was lower than zero. Such strict condition resulted in rejection of 14.7% models for VOG eye tracker. The rejection percentage for different groups and functions is presented in Table 1. It can be observed that the rejection per-

Group	x^1	x^2	x^3	ANN	SVR
5	16.2%	32.4%	87.2%	10.4%	15.9%
7	5.7%	6.1%	35.1%	2.2%	3.8%
9	3.9%	2.7%	14.5%	1.4%	0.9%
11	2.6%	2.0%	5.1%	0.5%	0.3%
13	2.0%	2.0%	2.0%	0.0%	0.0%

Table 1. Rejections percentage for VOG eye tracker

centage was lower when more calibration points were taken into account. For 5 points almost 90% of models were not feasible for the cubic (x^3) function. It must be remembered that a cubic function requires calculation of 18 parameters so it needs more data than e.g. a linear function. The ANN and the SVR functions were stable for each class and the number of rejections was always lower than for the polynomial functions. Similar results - but with higher rejection rates - were achieved for the Jazz-novo eye-tracker. The only important difference was that the ANN and the SVR functions gave much higher rejection rates for this device.

4.2 Multidimensional analysis of the recorded data

Filtered data was analyzed taking few aspects into account. It was interesting to check how the average error changed when various number of points constituted

This is a pre-print. The original version was published in Springer, Advances in Intelligent Systems and Computing Volume 284, 2014, and is available via
http://dx.doi.org/10.1007/978-3-319-06596-0_21

training set and what is an influence of a function used to build a regression model. And finally a question on the impact of a calibration points layout on defined model accuracy was asked. All issues were considered separately for both eye trackers used in the experiments.

Functions comparisons. The first step of the analysis mentioned above was to compare accuracy of different functions using E_{deg} value - that is average angular error. As can be seen in table 2 the best average results for the VOG system

VOG			Jazz		
Func	E_{deg}	**SD**	**Func**	E_{deg}	**SD**
x^2	2.717	1.846	x^2	4.012	1.895
SVR	2.786	1.680	x^1	4.070	2.0188
x^1	3.135	1.965	x^3	4.671	1.762
ANN	3.268	1.650	ANN	5.388	1.672
x^3	3.485	1.770	SVR	7.019	1.705

Table 2. Errors by function

gave the quadratic function x^2. However the difference between x^2 and SVR is not significant ($p = 0.07$). Significant differences can be noticed in relation to other functions.

For Jazz the best function was x^2 although the difference between x^1 and x^2 was not significant. What was interesting that the SVR and the ANN functions were significantly worse for the Jazz-novo eye-tracker.

When errors for various number of points were analyzed, it occurred that for 5-points groups and VOG results the linear function (x^1) became better than SVR and the significant difference can only be noticed for the ANN and the cubic (x^3) function being the worst methods (table 3). For Jazz x^1 function is significantly the best function for 5-points calibration. On the contrary for 13-

VOG			Jazz-novo		
Func	E_{deg}	**SD**	**Func**	E_{deg}	**SD**
x^1	4.427	1.945	x^2	3.415	2.029
x^2	4.786	1.783	x^1	3.561	2.130
ANN	6.037	1.623	SVR	3.588	1.867
x^3	6.082	1.618	ANN	3.986	1.797
SVR	8.100	1.1486	x^3	5.421	1.638

Table 3. Errors by function for sets with 5 points

points group the SVR method became the best method and the linear function x^1 became significantly the worst one. For the Jazz-novo device all polynomial functions were the best ones but with no significant differences (table 5).

This is a pre-print. The original version was published in Springer, Advances in Intelligent Systems and Computing Volume 284, 2014, and is available via
http://dx.doi.org/10.1007/978-3-319-06596-0_21

VOG			Jazz-novo		
Func	E_{deg}	**SD**	**Func**	E_{deg}	**SD**
SVR	1.995	1.401	x^3	3.349	1.585
x^3	2.138	1.510	x^2	3.629	2.017
x^2	2.323	1.572	x^1	4.079	2.321
ANN	2.328	1.385	ANN	4.302	1.554
x^1	2.930	1.894	SVR	5.301	1.799

Table 4. Errors by function for 13 points set

Number of points comparison. When average errors were calculated for set with various numbers of points it occurred that, as it could be expected, 13-points set gave lowest errors (table 5). However, the differences between 13 and 11 points groups were not significant for both eye trackers. Additionally, when similar comparison was done only for models that used x^2 function (table 6), it turned out that the differences between groups 13,11,9 and 7 were not significant. It shows that higher number of calibration points not necessarily causes lower error rates.

VOG			Jazz-novo		
Points	E_{deg}	**SD**	**Points**	E_{deg}	**SD**
13	2.341	1.594	13	4.455	2.737
11	2.522	1.629	11	4.710	2.679
9	2.806	1.656	9	5.130	2.769
7	3.073	1.704	7	5.546	2.817
5	3.722	1.983	5	6.338	2.697

Table 5. Errors by number of points in set

	VOG		Jazz-novo	
Points	E_{deg}	**SD**	E_{deg}	**SD**
13	2.323	1.572	3.629	2.017
11	2.396	1.628	3.662	1.948
9	2.461	1.699	3.801	1.933
7	2.580	1.799	3.891	1.770
5	3.415	2.029	4.786	1.783

Table 6. Errors by number of points in set for x^2 function

Finding the best models. Searching for the best models, the comparison of the E_{deg} errors was performed. All results were sorted with ascending order and

This is a pre-print. The original version was published in Springer, Advances in Intelligent Systems and Computing Volume 284, 2014, and is available via http://dx.doi.org/10.1007/978-3-319-06596-0_21

ten sets with the lowest error values were analyzed. The analysis was done both for each function independently and for all functions together, however data from different eye trackers were treated as autonomous sets.

In table 7 there were presented the best results in form of ten sets with the lowest E_{deg} average error values. Once again, due to limitation of the paper size, only results subset concerning polynomial functions were showed. It can

Jazz-novo				VOG			
Function	**Points**	**No**	E_{deg}	**Function**	**Points**	**No**	E_{deg}
x^3	13	1	3.349	x^3	11	9	2.058
x^2	7	16	3.357	x^3	13	1	2.138
x^1	7	13	3.377	x^2	11	9	2.256
x^2	11	12	3.450	x^2	9	2	2.261
x^2	9	7	3.476	x^2	7	1	2.265
x^2	7	13	3.480	x^3	11	1	2.267
x^2	9	2	3.518	x^2	7	7	2.270
x^2	11	1	3.545	x^3	11	12	2.287
x^2	11	7	3.547	x^2	7	10	2.289
x^2	11	10	3.573	x^2	11	8	2.291

Table 7. Best ten sets with lowest E_{deg}.

be observed that for both devices the set 13_1 - that is the reference set using all training points - together with the most complicated polynomial function x^3 provided one of the best results (the lowest E_{deg} error). But it turned out that the differences between this set and other sets with lower number of points are not significant. Especially for Jazz-novo there are two models (x2_7_16 and x1_7_13) for which the results are almost the same. It shows that the number of calibration points not necessarily results in lower error rates.

Interesting situation can be observed when analyzing the SVR and the ANN functions. It occurs that for the VOG system SVR outperformed all other functions (with 9 of 10 best models) and ANN gave results comparable to the polynomial functions (with the best result 2.166 for ANN_11_8). Quite differently in case of Jazz-novo eye-tracker the SVR methods provided very bad results (with 5.13 E_{deg} for set 13_1) and ANN was only a fraction better (with 4.30 for set 13_1). Such bad results require further studies.

In the next step sets differing in number and layouts of points were analyzed in terms of their usefulness for building a calibration model. This analysis, based on ten sets for which the lowest E_{deg} error values were obtained, was done for each of the functions used. It was expected that groups of the chosen sets would be dominated by sets with higher number of calibration points. The conducted studies confirmed this assumptions for cubic, ANN and SVR methods where sets with 11 points constituted majority of a group. However, in case of linear and quadratic functions, greater diversity of sets types was observed. There were sets with 5, 7 and 9 points especially for the first of mentioned methods.

This is a pre-print. The original version was published in Springer, Advances in Intelligent Systems and Computing Volume 284, 2014, and is available via http://dx.doi.org/10.1007/978-3-319-06596-0_21

These findings regard both types of eye-trackers although some differences were noticed. It concerned the quality of results achieved for the particular functions. The average error rates obtained in case of the Jazz-novo eye-tracker with regard to linear and quadratic functions were lower than for cubic, ANN and SVR ones, even though they were calculated using lower number of points. The opposite situation was determined in case of the second eye-tracker. The cubic, ANN and SVR methods provided the lower average error values than other polynomial functions.

Point location comparison. As it was shown in the previous section, groups with lower number of points may be as good as groups with higher number of points when a correct function is used. In this section different sets with the same number of points were compared to find out if the points location influences error rates. For this purpose the average error for all analyzed functions and for each set of points was calculated. The results for the VOG eye tracker are gathered in table 8. Analysis of these results showed that in case of the VOG

Set	E_{deg}	Set	E_{deg}	Set	E_{deg}
5_7	2.81	7_2	2.79	9_2	2.56
5_2	3.27	7_7	2.83	9_8	2.64
5_8	3.28	7_18	2.89	9_1	2.64
5_15	3.31	7_13	2.90	9_6	2.68
5_1	3.39	7_14	2.92	9_7	2.73
5_19	3.74	7_8	2.93	9_5	2.78
5_17	3.86	7_15	2.93	9_9	2.97
5_3	3.87	7_17	2.93	9_10	3.07
5_18	3.93	7_1	2.94	9_3	3.28
5_12	3.93	7_10	3.04	11_9	2.28
5_4	4.12	7_16	3.05	11_1	2.36
5_11	4.17	7_9	3.07	11_10	2.37
5_13	4.31	7_6	3.11	11_8	2.42
5_16	4.40	7_5	3.22	11_2	2.48
5_14	4.47	7_4	3.32	11_7	2.48
5_10	4.60	7_3	3.62	11_5	2.59
5_9	6.13	7_19	3.71	11_12	2.62
5_6	7.38	7_12	3.72	11_11	2.63
5_5	7.75	7_20	3.76	11_6	2.66
		7_11	4.06	11_4	2.69
				11_3	2.70

Table 8. Errors for sets ordered by group and error rate (VOG)

eye tracker there were significant differences for all five points group (19 sets). Among them the best set 5_7 (points: 3,6,9,12,13) giving average error 2.8 deg and the worst set 5_5 (points: 1,7,9,12,13) with an average error equal to 7.74

This is a pre-print. The original version was published in Springer, Advances in Intelligent Systems and Computing Volume 284, 2014, and is available via http://dx.doi.org/10.1007/978-3-319-06596-0_21

deg can be mentioned. After deeper layout examination it occurred that the lowest errors are calculated for sets containing point number 3 and 6 (north and south positions on the screen) and additionally of some points in the middle (10,11,12,13). Utilizing points in corners of the screen resulted in significantly higher errors. Similar situation was observed for the VOG system in case of seven points group (20 sets) although the differences were not so significant (from 2.78 deg for set 7_2 to 4.06 for set 7_11). There were no significant differences in points layout for sets with 9 and 11 points noticed.

Studying results for the second eye tracker Jazz-novo it was observed that there were significant differences in 5-points group as well. However, the results varied from that obtained in the VOG case. Although the same two sets (5_5 and 5_6) were the worst ones, there were sets for which Jazz-novo worked much better than VOG eye tracker. The 5_9 set with points on the left side of the screen can be taken as an example. In general, points located in the corners of the screen did not influence significantly Jazz-novo outcome - 5_19 set with points 1,5,7,8,9 was the fourth of best models. As it was stated above, their usage in case of the VOG system entitled higher error rates.

Differences for all other groups (7,9 and 11) were not significant but similar trends to the VOG system findings were observed.

5 Conclusions

The main goal of the research was to check repeatability of results for both various regression functions being used and various numbers and layouts of calibration points. To achieve this goal two different eye trackers were used. Using them two environments were developed to test five various functions operating on 61 sets of calibration points. These sets differed with numbers of points and their position on the screen. The obtained results were compared in terms of type of device, type of function and type of calibration points sets. There were some significant differences found. For instance simpler regression functions like x^1 or x^2 operated better than more complicated (like x^3 or SVR) on sets with lower number of points. Some specific sets with 7 points gave results comparable to 13 points sets. In the same time it was showed that calibration results highly depend on the type of eye tracker: (1) Points locations in screen corners were not a good solution for VOG while gave very good results for Jazz-novo. (2) Regression functions SVR and ANN worked very good for VOG eye tracker and very bad for Jazz-novo. (3) x^2 function outperformed x^1 function for VOG but the results of both were similar for Jazz-novo. Therefore, pointing out the best calibration set, which worked well for all functions and for both devices, turned out to be difficult.

Because the presented studies focused on making the comparison described above, the further research are planned to be aimed at an improvement of the particular results separately for each eye tracker.

This is a pre-print. The original version was published in Springer, Advances in Intelligent Systems and Computing Volume 284, 2014, and is available via
http://dx.doi.org/10.1007/978-3-319-06596-0_21

References

1. Pieter Blignaut and Daniël Wium. The effect of mapping function on the accuracy of a video-based eye tracker. In *Proceedings of the 2013 Conference on Eye Tracking South Africa*, ETSA '13, pages 39–46, New York, NY, USA, 2013. ACM.
2. X.L.C. Brolly and J.B. Mulligan. Implicit calibration of a remote gaze tracker. In *Computer Vision and Pattern Recognition Workshop, 2004. CVPRW '04. Conference on*, pages 134–134, 2004.
3. Juan J. Cerrolaza, Arantxa Villanueva, and Rafael Cabeza. Taxonomic study of polynomial regressions applied to the calibration of video-oculographic systems. In *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications*, ETRA '08, pages 259–266, New York, NY, USA, 2008. ACM.
4. Jennifer H Darrien, Katrina Herd, Lisa-Jo Starling, Jay R Rosenberg, and James D Morrison. An analysis of the dependence of saccadic latency on target position and target characteristics in human subjects. *BMC neuroscience*, 2(1):13, 2001.
5. Andrew T Duchowski. A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers*, 34(4):455–470, 2002.
6. Kai Essig, Marc Pomplun, and Helge Ritter. A neural network for 3d gaze recording with binocular eye trackers. *IJPEDS*, 21(2):79–95, 2006.
7. Edmund Burke Huey. *The Psychology & Pedagogy of Reading*. The Macmillan Company, 1908.
8. Robert J. K. Jacob. The use of eye movements in human-computer interaction techniques: What you look at is what you get. *ACM Transactions on Information Systems*, 9:152–169, 1991.
9. Francis Martinez, Andrea Carbone, and Edwige Pissaloux. Gaze estimation using local features and non-linear regression. In *Image Processing (ICIP), 2012 19th IEEE International Conference on*, pages 1961–1964. IEEE, 2012.
10. Jorge J Moré. The levenberg-marquardt algorithm: implementation and theory. In *Numerical analysis*, pages 105–116. Springer, 1978.
11. Carlos H. Morimoto and Marcio R. M. Mimica. Eye gaze tracking techniques for interactive applications. *Comput. Vis. Image Underst.*, 98(1):4–24, April 2005.
12. Basilio Noris, Jean-Baptiste Keller, and Aude Billard. A wearable gaze tracking system for children in unconstrained environments. *Computer Vision and Image Understanding*, 115(4):476–486, 2011.
13. N Ramanauskas. Calibration of video-oculographical eye-tracking system. *Electronics and Electrical Engineering*, 8(72):65–68, 2006.
14. Alex J Smola and Bernhard Schölkopf. A tutorial on support vector regression. *Statistics and computing*, 14(3):199–222, 2004.
15. Yusuke Sugano, Yasuyuki Matsushita, and Yoichi Sato. Calibration-free gaze sensing using saliency maps. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 2667–2674. IEEE, 2010.
16. Arantxa Villanueva and Rafael Cabeza. Models for gaze tracking systems. *Journal on Image and Video Processing*, 2007(3):4, 2007.
17. Zhiwei Zhu and Qiang Ji. Eye and gaze tracking for interactive graphic display. *Machine Vision and Applications*, 15(3):139–148, 2004.
18. Dave M. Stampe. Heuristic filtering and reliable calibration methods for video-based pupil-tracking systems. *Behavior Research Methods, Instruments, & Computers*, 25(2):137–142, 1993.
19. Takehiko Ohno, Naoki Mukawa, and Atsushi Yoshikawa. Freegaze: A gaze tracking system for everyday gaze interaction. In *Proceedings of the 2002 Symposium on*

This is a pre-print. The original version was published in Springer, Advances in Intelligent Systems and Computing Volume 284, 2014, and is available via
http://dx.doi.org/10.1007/978-3-319-06596-0_21

Eye Tracking Research & Applications, ETRA '02, pages 125–132, New York, NY, USA, 2002. ACM.

20. Marcus Nyström, Richard Andersson, Kenneth Holmqvist, and Joost van de Weijer. The influence of calibration method and eye physiology on eyetracking data quality. *Behavior research methods*, 45(1):272–288, 2013.