Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 35 (2014) 1073 – 1081

Procedia Computer Science

18th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems - KES2014

Towards accurate eye tracker calibration – methods and procedures

Katarzyna Harezlak *, Pawel Kasprowski, Mateusz Stasch

Institute of Informatics, Silesian University of Technology, Gliwice, Poland

Abstract

Eye movement is a new emerging modality in human computer interfaces. With better access to devices which are able to measure eye movements (so called eye trackers) it becomes accessible even in ordinary environments. However, the first problem that must be faced when working with eye movements is a correct mapping from an output of eye tracker to a gaze point – place where the user is looking at the screen. That is why the work must always be started with calibration of the device. The paper describes the process of calibration, analyses of the possible steps and ways how to simplify this process.

Peer-review under responsibility of KES International.

Keywords: eye movement; eye trackers; calibration; regression

1. Introduction

Eyes are the most important input device of human brain. Most of the information that is acquired by a human comes through eyes. The main problem of visual perception is that eyes register scene with uneven acuity. Only the part of the scene that falls on the fovea – region in the middle of the retina – is seen with correct sharpness. All other regions of retina are able to register only contours and fast movements. Therefore, eye movements are very important for correct recognition of objects in visual field. That is why the way that eyes work determines our perception and may reveal our intentions.

Eye tracking devices collect information about eye movements. The first eye trackers were built in the beginning of 20 century[1] but the last decade made eye tracking technology accessible in ordinary personal computer interfaces. Nowadays building eye trackers that achieve 1-2 degrees of accuracy using a simple web camera is possible owing to existence of many image processing algorithms that popularized video based eye trackers - so called video oculography (VOG) eye trackers. VOG eye tracker returns information about a position of an eye within an eye's image registered by a camera. This raw data must be somehow translated into a gaze point. The gaze point may be defined as the point on the screen where a person is currently looking at. To obtain the function mapping eye tracker output to a gaze point, nearly every eye tracking experiment starts with so called calibration procedure[2]. There are some

* Corresponding author.
E-mail address: katarzyna.harezlak@polsl.pl.

Peer-review under responsibility of KES International.
doi:10.1016/j.procs.2014.08.194

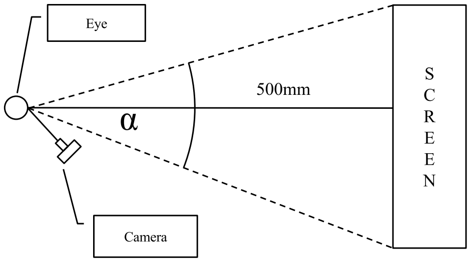

Fig. 1. Experimental setup

systems that don't require calibration, however building such a system requires special care and applying a proper architecture[3,4].

The calibration may be done when an actual gaze point of a person being calibrated is known. The solution is to show a user a point on a screen and to register eye tracker output when he gazes at it. Of course one point is not enough to obtain a correct calibration. There must be several points displayed, preferably in different parts of a screen. Every point must be presented long enough to gather sufficient amount of data[5]. The main problem of the calibration process is that it takes time and is not convenient for users that are not used to staring at the same point for a long time. That is why there is a need to shorten calibration procedure duration. On the other hand, the more points are analyzed and the more recordings are gained during the calibration, the more reliable mapping function is produced.

The paper focuses on exploring possibilities how to simplify and shorten calibration process without losing accuracy of the mapping function. The contribution of the paper is a set of guidelines how to prepare the calibration procedure. It extends the research presented in[5] with usage of both two additional methods – ANN and SVR – for building calibration models and with more diverse sets of calibration points.

Section 2 presents the setup of the experiment. Section 3 describes, which aspects should be taken into account when building calibration model, including initial delay due to saccadic latency, incorrect samples removal and length of registration. Section 4 analyzes the impact of calibration point layout in results. The last section summarizes the obtained results.

2. Experimental setup

Calibration data was captured using a VOG head-mounted eye tracker developed with single CMOS camera with USB 2.0 interface (Logitech QuickCam Express) with 352x288 sensor and lens with IR-Pass filter. 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 causes "dark pupil" effect, which was useful during pupil detection. The system generates 20 - 25 measurements of the center of the pupil per second. The experimental setup of tracking system is shown in Fig. 1. The calibration was done on a 1280x1024 (370mm x 295mm) flat screen. The eye-screen distance was 500mm and vertical gaze angle was 40° and horizontal gaze angle was 32°. It is a usual condition when working with the computer. To avoid head movements, the head was stabilized using chin rest.

The calibration procedure was done using a set of 29 dark points distributed over a white screen as in Fig 2. Points were displayed in each session in the same predefined order. Each point was displayed for 3618 msec. To keep user attention on the selected point it was pulsating. There were 26 participants and 49 sessions registered. Before the experiment, participants were informed about the general purpose of the experiment after which they signed a consent form. The time interval between two sessions of the same user was at least three weeks to avoid the learning effect – when user

learns the order of the points and is able to anticipate the next point position. All images for which it was impossible to find eye center were removed. It was 1% of all samples with average value of removed samples from 0% to 5% for separate calibrations.

The next step after collecting eye positions related to points of regard (PoR) shown on a screen was creating a function that correctly maps eye positions to PoRs for unknown samples. There should be in fact two functions

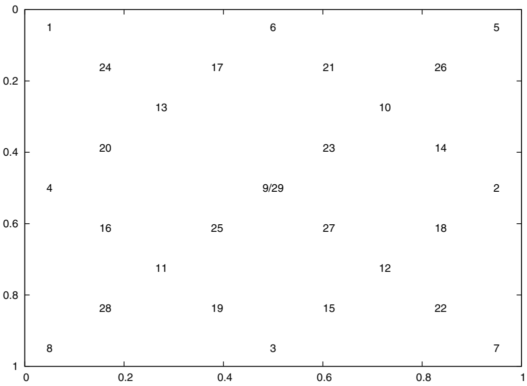

Fig. 2. Points used during sessions

created – separately for each axis. It may be defined by equations:

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

where x_e and y_e represent data obtained from eye tracker and x_s and y_s are estimated gaze coordinates on a screen. All of them are considered in the cartesian coordinate system.

There may be any regression function used. Therefore three different functions were chosen in this study. The first and the obvious one was a second order polynomial function in form:

$$x_s = A_x x_e^2 + B_x y_e^2 + C_x x_e + D_x y_e + E_x, \tag{2}$$

$$y_s = A_y x_e^2 + B_y y_e^2 + C_y x_e + D_y y_e + E_y.$$

The values of $A_x ... E_x$ and $A_y ... E_y$ parameters were calculated using a classic Levenberg-Marquardt optimizer[6]. Polynomial regression is one of the most popular and the fastest mapping functions used in many eye tracking applications[7 8]. The second type of function was an artificial neural network (ANN). An activation network with sigmoid function as an activation function was used. Network was trained 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. The network was trained until the total train error was lower than 0.1. ANN has been already used in several eye tracing applications[9 10]. The third type of function was Support Vector Regression (SVR)[11]. It was RBF kernel used with parameters $C = 10$ and $\gamma = 8$. Similar function has been used for eye tracker calibration in[12] and[13] but in completely different setups.

3. Aspects of building calibration model

One of the basic issues regarding the eye trackers calibration is ensuring its shortest possible accomplishment time with simultaneously guaranteeing high accuracy of a defined calibration model. Because this process is influenced mainly by a number of calibration tasks, their complexity and duration, this is the area in which the desirable solution of the presented problem should be searched for. Many scientific works studied this subject, yet so far no single unequivocal method was found, so searching for the new solutions is still valid. When dealing with this task, some questions have to be answered. Among them there can be listed:

- how to handle the fact that eyes react to stimulus change after some amount of time,
- how long each calibration point should be presented to a user,
- what is the least number of calibration points allowing to achieve the required accuracy of a calibrated system.

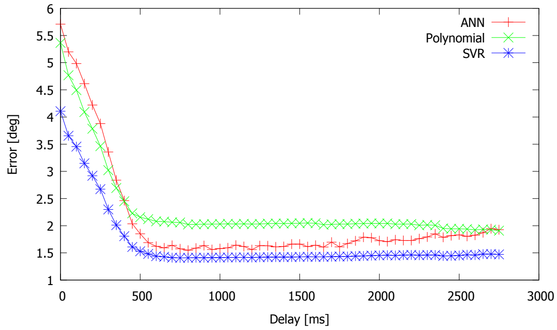

Fig. 3. The accuracy obtained for different delays represented in degrees. The accuracy was calculated using the same points that were used for building the model.

3.1. Initial delay

The first of the aforementioned problems - making a decision on which samples of an eye movement signal use to build a calibration model - results from the often discussed phenomena of *saccadic latency*. This occurrence is understood as a time from a stimuli presentation to the commencement of a saccade [14]. When the point on the screen changes its position, it takes some time for the human brain to react and to initiate eye movement. Additionally, it happens quite often that the first saccade (fast eye movement to a new location) is misplaced and the fixation point must be corrected.

Taking this findings into account, it becomes evident that data recorded during initial phases of the measured samples should be labeled as useless for calculating a calibration model. The length of these phases was studied during the presented research. To achieve a goal, the total display time for each of 29 points was divided into slots of 50 msec. Subsequently, 55 sets of meaningful points for each user were defined. Each set contained samples recorded after the passage of time equal to multiple of 50 msec. The sets were built for delays ranging from 0 to 2750 msec. They were then used for building a calibration model with usage of three methods: polynomial regression, ANN and SVR. The quality of obtained solutions was checked using the same set of points. For this purpose, the E_{deg} error represented by a degree distance between calibration points and their locations calculated by the three given methods was defined (Equ. 3).

$$E_{deg} = \frac{1}{n} \sum_{i} \sqrt{(x_i - \widehat{x_i})^2 + (y_i - \widehat{y_i})^2}, \tag{3}$$

where x_i, y_i represent the observed values and $\widehat{x_i}$, $\widehat{y_i}$ values calculated by the model.

Obtained results are presented in Figure 3. It can be noticed that the best results were achieved for the SVR method. However, the main goal of this part of the research was not to validate system by comparing accuracy of methods used but to point out the best delay, after which it is feasible to state that user's eyes are already directed at the required point. Thus, analyzing results from this point of view, it can be concluded that samples recorded during first 600-700 msec. should not be taken into account in constructing a calibration model.

3.2. Removing incorrect samples

After deciding what delay should be applied to build a proper calibration model, the following issue was analyzed: how long the registering process needs to be in order to correctly estimate an eye position for a particular calibration point. However, before this step of the research was started, the data gathered on the eye movement signals had to be filtered. Analysis of the earlier obtained results showed that some of the participants had problems with completion of the calibration tasks. There may be several reasons for bad quality sessions.

- Problems with acquiring the image of an eye with sufficient quality. The reason may be mascara, blink and so on.

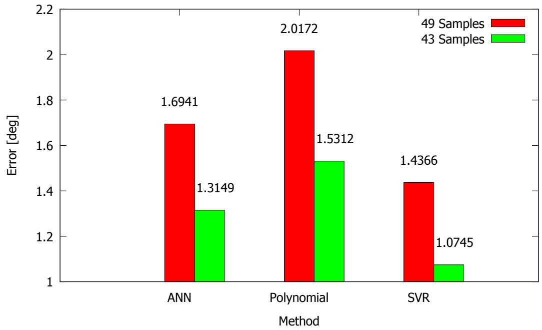

Fig. 4. Comparison of error values obtained before and after removing incorrect samples being about 10% of all registered one.

- Problems with participant's focus on task. Some people are not able to stare at the same point for longer time and their eyes are in constant movement.
- General problems with lighting conditions, software or hardware.

Due to that fact coefficients of determination were calculated for all collected samples for both vertical and horizontal axes (equ. 4).

$$R_x^2 = 1 - \frac{\sum_i (x_i - \widehat{x_i})^2}{\sum_i (x_i - \bar{x})^2}, \tag{4}$$

$$R_y^2 = 1 - \frac{\sum_i (y_i - \widehat{y_i})^2}{\sum_i (y_i - \bar{y})^2},$$

where x_i, y_i represent observed values, $\widehat{x_i}$, $\widehat{y_i}$ represent values calculated by model and $\bar{x}$, $\bar{y}$ are the means of observed values.

Defining calibration models and their verification were done using all calibration points. Sessions for which R_x^2 or R_y^2 were lower than 0.85 were removed from further studies. This was the case for six sessions. The $\widehat{x_i}$ and $\widehat{y_i}$ values were evaluated using a polynomial regression function, however the correctness of the choice of the samples, which were to be withdrawn from subsequent tests, was confirmed using the ANN method and the SVR one. It can be noticed that by removing incorrect samples a substantial accuracy improvement was obtained for all types of calibration methods with the biggest improvement for the polynomial function (25%). The conclusion, which can be drawn from the achieved results states that a special care of samples quality is required as calibration process is very sensitive for poor quality data (Fig 4).

3.3. Registration lengths

The filtered set of samples was utilized for determining the shortest possible registration lengths that gives an acceptable accuracy of a calibration. Analyzed intervals of time varied from 150 msec to 2850 msec. starting with delay of 700 msec., accordingly to the previous conclusions. As the sampling frequency was about 20-25Hz, 2-3 samples were gathered in each 150 msec. of registration time. Samples from outside of a currently analyzed time range were used for checking accurateness of chosen models. The models were verified by determining the deviation of estimated values from accurate point coordinates, by the use of equation 3. The results are presented in Figure 5. It can be observed that in case of the polynomial model the best result (the lowest error) is achieved with the time window length of 1250 msec. Similarly, for two other functions 1250 msec. seems to be a good tradeoff between accuracy and length. Summarizing the findings of this research stage - it can be concluded that the display time of a calibration point can be reduced to 1950 msec. (1250 of the meaningful time + 700 msec. of the delay), but its first 700 msec. should not be used for building calibration model.

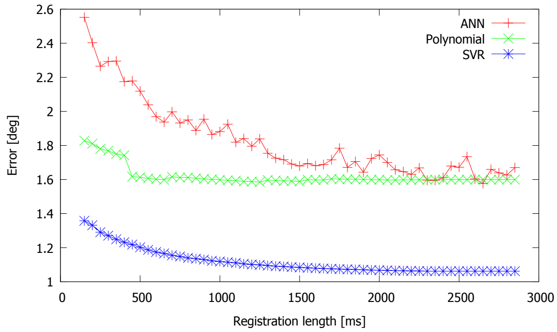

Fig. 5. Error rates for various registration lengths.

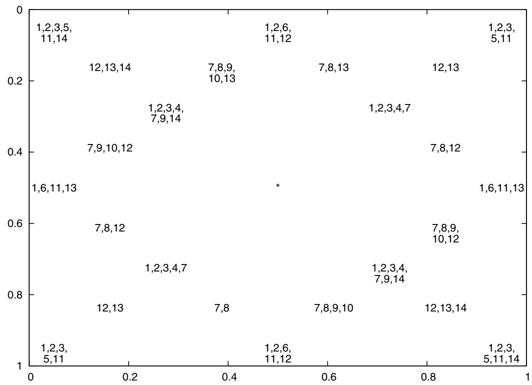

Fig. 6. Points layout in 14 groups.

4. Choosing calibration points

While the previous chapter answered two out of tree of the aforementioned questions regarding length and meaningful points of registration, this section describes the studies aiming at finding how the number and locations of calibration points influence the accuracy of the model built. From the set of 29 calibration points, 14 groups of points were chosen for the calibration. There were 4 groups consisting of 5 points, 1 of 7 points, 3 of 9 points, 3 of 11 points and 3 of 13 points. The groups with the same number of points differed in points layout which was asymmetrical for some groups. The arrangement of points in each group is presented in Figure 6. Numbers in cells represent points in particular group which were marked by successive, monotonically increasing numbers starting from 1. The order of point presentation was the same as shown in Figure 2. And so, the following groups were constituted - in accordance with notation *group number(number of points)* - 1(13), 2(11), 3(9), 4(5), 5(5), 6(5), 7(13), 8(9), 9(7), 10(5), 11(9), 12(11),13(11), 14(13) points respectively. The point in the center of the screen was included into all groups. The points that weren't used in any group were included into test sets.

Based on samples gained for each group, calibration models were built using all three methods. To test particular models, a set of 14 points different from the training points was defined. This set for majority of groups was constant, however in few cases two or three points had to be changed, because of belonging to a calibration group. The points were exchanged by points located close to their positions. For each group and for each method, the E_{deg} error (eq 3) was determined.

Obtained results are presented in the table 1. They were sorted by error values with ascending order. The best results (the lowest error values) were achieved for groups with higher number of points, accordingly to the preliminary assumptions. However, some of the results indicate that it doesn't necessary need to be a rule and an accuracy of a calibration depends on stimuli layout as well. Groups marked with number 2(11) and 13(11) ensure better precision of the estimating gaze points then group with number 7(13). What is more, there are groups marked with numbers 3(9), 8(9) and, in case of polynomial function, even the group number 6(5) with much lower point number but still providing lower error value then the 7(13) one. The reason of such poor results of group 7(13) is probably that the points are concentrated near the middle of the screen and there are no points at the edges of the screen in this group. Another interesting finding is that differences between the best and the worst group is the lowest for polynomial regression model and the highest for SVR model. Additionally, SVR gives results with the highest inter-group deviation (1.95 on average) comparing to ANN (1.09) and polynomial (1.51). Both these findings influence the following analyses.

Polynomial			**ANN**			**SVR**		
Group No.	E_{deg}	S_d	Group No.	E_{deg}	S_d	Group No.	E_{deg}	S_d
13(11)	1.80	1.17	1(13)	1.90	0.93	1(13)	1.71	1.16
2(11)	1.89	1.17	2(11)	2.13	1.06	2(11)	2.16	1.35
14(13)	1.90	1.28	13(11)	2.27	0.77	13(11)	2.24	1.61
1(13)	1.91	1.20	12(11)	2.61	0.95	8(9)	2.86	1.77
3(9)	2.05	1.26	8(9)	2.67	1.01	3(9)	2.88	1.57
12(11)	2.09	1.23	3(9)	2.80	1.43	7(13)	2.95	1.63
8(9)	2.17	1.49	7(13)	2.96	0.99	12(11)	3.02	1.49
6(5)	2.27	1.55	9(7)	2.98	1.24	14(13)	3.09	1.26
11(9)	2.28	1.43	14(13)	3.03	1.22	11(9)	3.17	1.42
7(13)	2.34	1.51	10(5)	3.04	1.36	6(5)	3.37	1.82
4(5)	2.37	1.54	4(5)	3.09	0.82	4(5)	3.38	1.95
5(5)	2.47	1.40	6(5)	3.30	0.87	9(7)	3.72	2.73
9(7)	2.91	2.47	11(9)	4.20	1.25	5(5)	3.86	1.96
10(5)	2.95	2.50	5(5)	4.75	1.40	10(5)	5.45	5.62

Table 1. Errors calculated for different functions and groups of points.

The obtained results were compared using a paired Student test to check if the differences in error rates are significant. The comparison was organized as follows:

- Inter-class - groups of equal point numbers were compared with each other,
- Between-classes - the best member of each class (with lowest E_{deg} values) was compared with the best members of other classes.

Such scenario was applied for all studied methods.

4.1. Inter-class comparisons

In case of the polynomial method the first of mentioned tests, in all comparisons, provided outcomes that did not allow to reject the hypothesis H0. It means that differences between groups in the same class were not significant. On the contrary, there were significant differences found using for the ANN and SVR methods. For instance group 5(5) gave significantly worse results than groups 4(5) and 6(5) for ANN (with $p < 0.001$). Similarly 11(9) group was significantly worse than 3(9) and 8(9) groups (with $p < 0.001$ in both cases). It may be a bit surprising because for both 5(5) and 11(9) the layout was symmetrical with even distribution of points over the screen. The probable reason for high errors is the usage of points located in the corners of the screen. People have problems with stable focusing at the points near the edge of their vision so the registration of these points may be of lower quality. When taking into account 13 points class, the group 1(13) with even distribution of points on the screen proved to be significantly better than groups 7(13) and 14(13) with $p < 0.001$. Similarly to the previous findings group 14(13) consisted of points in corners and group 7(13) of points laying near to each other. This finding was confirmed for SVR method as well.

4.2. Between-classes comparisons

The differences between the best groups of each class showed that there are no significant differences between the results for 11 and 13 calibration points. In fact, the result for group 13(11) was the best one for polynomial method. When comparing the best groups of 11 and 9 points the results showed significant differences for all three methods (with $p < 0.05$). But it should be emphasized that the best group of 9 points was – for every method – better than group 12(11). The differences between the best 9 points group and 5 points group were not significant for all three methods.

The most important finding, which can be read from the obtained data, is the confirmation that accuracy of calibration model depends on both number and a layout of the stimuli. There are some results indicating that it is possible to achieve better estimation using a smaller amount of points but applying their arrangement better. The finding described above in conjunction with the E_{deg} error analysis allows to conclude that the good quality of the calibration process and its acceptable time of duration is possible to achieve with relatively small calibration points number but having an appropriate layout.

5. Summary

Hardware and software development causes a constant advance of eye tracking systems. On one hand, this development regards expensive specialized devices. On the other hand, due to specialized algorithms, tracking eye movement became possible using cameras, which are accessible for ordinary users. Obtaining good accuracy in environments containing cameras of that type is nowadays of a big importance. It would allow to vastly expand their usability.

The research presented in the paper is one of the studies concerning the eye-tracking system accuracy problem. It concentrated on specifying phases of a calibration process, which can be simplified and shortened to make this process more convenient, yet still preserving sufficient system accuracy. The first test concerned a period of time which should be skipped when building a calibration model. Analysis of all collected samples allowed to indicate a precise time delay. This finding doesn't truncate total time of system calibration, although it influences the correctness of a calibration model. It gives a possibility to reduce a number of meaningful samples and therefore to make calibration procedure shorter. It was confirmed that, for sampling frequency lower than 25Hz, the display time of a calibration point can be as short as 1250 ms after a moment when an eye signal becomes stable. This outcome was verified using various types of calibration.

Another important outcome of the research is a conclusion that good layout of calibration point provides an opportunity to decrease a number of calibration points and thereby to shorten the calibration process, without diminishing the accuracy level. Three different regression methods were compared using significant amount of data. The comparisons revealed that classic polynomial method is more sensitive to bad samples while ANN and SVR methods handle the bad samples better. However, when the quality of samples is assured, the polynomial method is able to produce low error models using less number of points and is not so sensitive to the points layout as ANN and SVR methods. The results of all three methods were comparable for the best groups of points but when the number of calibration points was reduced the accuracy of polynomial method became better.

The presented results were gathered using only one device and 49 calibration sessions (of 26 subjects). Confirming all findings using different devices could be beneficial. Additionally, deeper analysis of the best layout of calibration points could be a possible further work.

References

1. E. B. Huey, The Psychology & Pedagogy of Reading, The Macmillan Company, 1908.
2. A. T. Duchowski, A breadth-first survey of eye-tracking applications, Behavior Research Methods, Instruments, & Computers 34 (4) (2002) 455–470.
3. A. Villanueva, R. Cabeza, Models for gaze tracking systems, Journal on Image and Video Processing 2007 (3) (2007) 4.
4. Y. Sugano, Y. Matsushita, Y. Sato, Calibration-free gaze sensing using saliency maps, in: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, IEEE, 2010, pp. 2667–2674.
5. N. Ramanauskas, Calibration of video-oculographical eye-tracking system, Electronics and Electrical Engineering 8 (72) (2006) 65–68.
6. J. J. Moré, The levenberg-marquardt algorithm: implementation and theory, in: Numerical analysis, Springer, 1978, pp. 105–116.

7. J. J. Cerrolaza, A. Villanueva, R. 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, ACM, New York, NY, USA, 2008, pp. 259–266.
8. P. Blignaut, D. 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, ACM, New York, NY, USA, 2013, pp. 39–46.
9. K. Essig, M. Pomplun, H. Ritter, A neural network for 3d gaze recording with binocular eye trackers., IJPEDS 21 (2) (2006) 79–95.
10. Z. Zhu, Q. Ji, Eye and gaze tracking for interactive graphic display, Machine Vision and Applications 15 (3) (2004) 139–148.
11. A. J. Smola, B. Schölkopf, A tutorial on support vector regression, Statistics and computing 14 (3) (2004) 199–222.
12. B. Noris, J.-B. Keller, A. Billard, A wearable gaze tracking system for children in unconstrained environments, Computer Vision and Image Understanding 115 (4) (2011) 476–486.
13. F. Martinez, A. Carbone, E. Pissaloux, Gaze estimation using local features and non-linear regression, in: Image Processing (ICIP), 2012 19th IEEE International Conference on, IEEE, 2012, pp. 1961–1964.
14. J. H. Darrien, K. Herd, L.-J. Starling, J. R. Rosenberg, J. D. Morrison, An analysis of the dependence of saccadic latency on target position and target characteristics in human subjects, BMC neuroscience 2 (1) (2001) 13.
15. X. Brolly, J. Mulligan, Implicit calibration of a remote gaze tracker, in: Computer Vision and Pattern Recognition Workshop, 2004. CVPRW '04. Conference on, 2004, pp. 134–134.
16. R. 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 (1991) 152–169.
17. C. H. Morimoto, M. R. M. Mimica, Eye gaze tracking techniques for interactive applications, Comput. Vis. Image Underst. 98 (1) (2005) 4–24.