Comparison of eye tracking devices used on printed images

Barbora Komínková[ab], Marius Pedersen[b], Jon Y. Hardeberg[b], Marie Kaplanová[a]
[a]University of Pardubice, 532 06 Pardubice, Czech Republic;
[b]Gjøvik University College, P.O. Box 191, N-2815 Gjøvik, Norway

ABSTRACT

Eye tracking as a quantitative method for collecting eye movement data, requires the accurate knowledge of the eye position, where eye movements can provide indirect evidence about what the subject sees. In this study two eye tracking devices have been compared, a Head-mounted Eye Tracking Device (HED) and a Remote Eye Tracking Device (RED). The precision of both devices has been evaluated, in terms of gaze position accuracy and stability of the calibration. For the HED it has been investigated how to register data to real-world coordinates. This is needed since coordinates collected by the HED eye tracker are relative to the position of the subject's head and not relative to the actual stimuli as it is the case for the RED device. Result Results show that the precision gets worse with time for both eye tracking devices. The precision of RED is better than the HED and the difference between them is around 10 - 16 pixels (5.584 mm). The distribution of gaze positions for HED and RED devices was expressed by a percentage representation of the point of regard in areas defined by the viewing angle. For both eye tracking devices the gaze position accuracy has been 95-99% at 1.5-2° viewing angle. The stability of the calibration was investigated at the end of the experiment and the obtained result was not statistically significant. But the distribution of the gaze position is larger at the end of the experiment than at the beginning.

Keywords: Eye tracking, precision, gaze position, stability of calibration.

1. INTRODUCTION

Eye tracking is the process of measuring either the point of gaze or the motion of an eye relative to the head. Collected data such as eye positions and eye movement can be statistically analyzed to determine the pattern and duration of eye fixations and the sequence of scan paths as a user visually moves through a page or screen.

Eye tracking is a technique used in different field such as vision, cognitive science, psychology, human-computer interaction, marketing research and medical research to provide useful information. Specific areas include web usability [1, 2], advertising [3], reading studies [4], and evaluation of image quality [5, 6, 17]. To perform these kinds of studies with valid data a precise eye tracker is needed.

Accurate knowledge of eye position is often desired, not only in research on the oculomotor system itself but also in many experiments concerning visual perception, where eye movements can provide indirect evidence about what the subject sees [7]. Research works are focused on the accuracy of the image of the eye pupil, because the accuracy of gaze tracking greatly depends upon the resolution of the eye images. The detection of pupil center in the image of the eye is the most important step for video-based eye tracking method [8]. If good accuracy is required, there is a method that uses edges and local patterns to obtain detection of eye features with subpixel precision. This algorithm can robustly detect the inner eye corner and the center of an iris with subpixel accuracy [9].

Different light conditions are also considerably influencing the eye tracking methods. The high contrast between the pupils and the rest of the face can significantly improve the eye tracking robustness and accuracy [10]. Very small pupil sizes make it difficult for the eye tracking system to model the pupil center. It depends on the brightness and size of the pupils; therefore light conditions are required to be relatively stable and the subjects close to the camera.

The aim of the work presented in this paper is to compare eye tracking devices used on printed images. From the literature review we have identified prior work that compares three different eye tracking devices in psychology of programming experiment. Nevalainen and Sajaniemi [11] studied the ease of use and accuracy of the three devices by having observers examine short computer programs using a program animator. The results showed that there were significant differences in accuracy and ease of use between the devices.

Human Vision and Electronic Imaging XIII, edited by Bernice E. Rogowitz, Thrasyvoulos N. Pappas,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 6806, 68061I, © 2008 SPIE-IS&T · 0277-786X/08/$18

As it is known, the head-mounted systems (HED) are effective for studies which require the head to move freely. On the other hand, with RED systems the observer has to keep the same position and avoid large movements most of the time. Head-mounted systems, within traditional usability testing, is useful for paper prototype studies or out-of-the-box studies, and also typically used in studies, where head or body movement is required of users (automobile drivers, airplane pilots or even athletes practicing). This project investigates if the HED can be used on printed images as the RED is used in some cases. One of the advantages is that the observer has larger possibility of freedom than in the RED.

In case of HED, data analysis is performed on data collected by a video camera. The problem is how to register data from HED to real-world coordinates. The system creates a superimposed image of a dot representing the participant's point of regard (exactly where they are looking), laid over the top of the image of their field of vision [12]. The coordinates collected by the eye tracker are thus relative to the position of the subject's head and not relative to the actual stimuli as in the remote eye tracking case. Hence, this method requires not only analysis of the coordinates generated by the eye tracker but also analysis of the recorded video [13, 14].

In order to investigate advantages and disadvantages of both eye tracking devices, an experiment has been designed and carried out for determining the precision of both devices in different directions (precision in the time aspect, precision on the edges of the image, etc.) and obtaining a percentage representation of the fixation point of the eye in certain areas (circles), which matches different viewing angles.

2. EXPERIMENTAL SETUP AND METHODOLOGY

2.1 Equipment

The main experiment hardware equipments were Remote Eye tracking Device (RED) and Head Mounted Eye tracking Device (HED) from SensoMotoric Instruments (SMI) (Figure 1a and b). The software equipment used was iView X system combining all required components for efficient high-quality eye movement and scene video recordings into a single high-performance computer.

a) b)

Fig 1. a) Remote Eye Tracking Device (RED); b) Head-mounted Eye Tracking Device (HED)

The calibration of the equipment is a very necessary and critical part of the experiment. The calibration establishes the relationship between the position of the eye in the camera view and a gaze point in space, the so-called point of regard. At the same time the calibration establishes the plane in space where eyes movements are rendered. Poor calibration can invalidate an entire eye tracking experiment, because there will be a mismatch between the participant's point of regard, and the corresponding location on a display [15].

There are many calibration methods which are possible to use, these usually differ in the number of points calibrated The calibration method named "9 Point with Corner Correction" was used in this experiment for both eye tracking devices [16]. Calibration of the system was done for each observer before commencing the experiment.

2.2 Experiment

The experiment was modeled in the same way for both eye tracking devices, and divided into two parts. One part was done with one of the eye tracking devices on one day and the second part was done with the second eye tracking device on the second day.

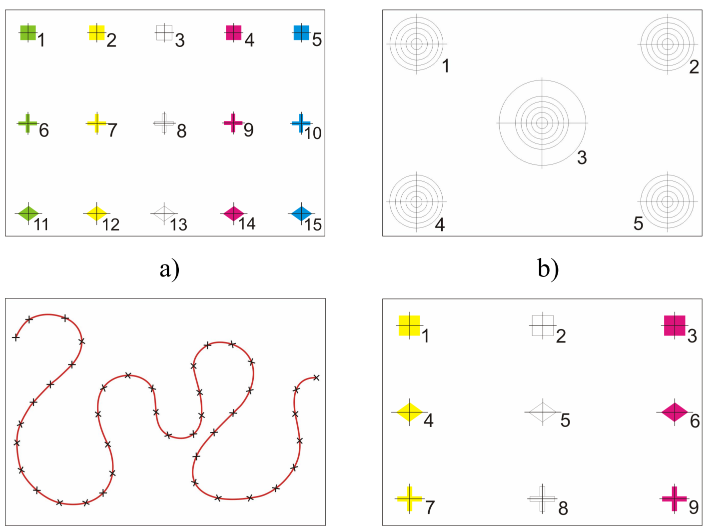
The experiment was carried out with 20 observers. Observers included a mix of students and no observers provided any evidence of color blindness. None of the participating observers used glasses, in order to achieve precise data. The dominant eye was found before the experiment [15], and this eye was tracked. Observers were randomly divided into groups, the first group used RED first then HED, and the second group used the devices in the reverse order. Six images were shown to the observers in a given sequence in the both case with RED and HED. In case of RED it was important to keep the same position of the images, and the same distance between image, observer and RED. Each observer did the experiment twice (20 observers resulted in 40 values). Instructions for observers were presented before and also during the experiment. After the experiment observers were asked to fill out a questionnaire.

2.3 Viewing and light conditions

The distance from the observers to the board with the images was approximately 80 cm for all the observers and the viewing angle for the whole image was about 32×24 degrees. The remote eye tracking camera had placement below the eye and the location of the camera depended on the dominant eye of the observer. A chair was selected that minimized the amount of upper body movements made by the observer. This decreased the possibility that the observer will change position in a way that causes gaze inaccuracies and prevented the observer from changing the distance from the eye to the image during the experiment. The images had to be highly visible and the eye could not have a lot of reflections. Hence the experiment was carried out in standard D50.

2.4 Images

Six images, of format A3 were created in a Adobe Illustrator, represented very simple test fields with different symbols, and the symbols were predominantly in a shape of a cross (Figure 2a, b, c and d). In order to better distinguish crosses in evaluation part each cross was numbered. To make a simple understanding and survey, we can say that four different images were made and one of them was created three times. From now we can call them as image A, B, C and D.



a) b) c) d)

Fig 2. a) One of the images A with numbering of the crosses; b) The image B with numbering of the crosses; c) The image C with 42 crosses on the curve; d) The image D with numbering of the crosses.

In 4 images (3 images A with 15 crosses and 4th image D with 9 crosses), for better recognition of the crosses, a unique combination of the shape and the color was given to the each cross. Further image B contains 5 crosses and each cross certain number of circles (4 crosses in corners and 5th cross in a middle of the image). The distances between these circles were created by 0.5 viewing angles. The first circle is starting 0.5 degrees from the centre of the cross and the last circle is 4 degrees from the centre of the cross. In the case of 4 crosses in the corners, the last circle is far 3 degrees from the center. On the last image C a simple curve with small crosses was created.

The sequence of the images was very important, because each image had its own meaning for the evaluation of the experiment. Their sequence is as follows: A, B, A, C, A, D. The image A was created for finding the precision of a gaze position on each symbol in the image and also for finding a possible change of precision with time. Therefore their sequence was not immediately consecutive. The image B with 5 simple targets should demonstrate a percentage representation of the points of regard in circles. By the curve with small crosses on the next image C had been expressed the total precision in whole and a subjective aspect. As the last image D should indicate if the calibration of the eye tracking device is stable when an observer moves and leaves his or her position in front of the device, and returns to sit in the same position. In the case of the HED, the observer took off the helmet and after a while the helmet was refitted on the observer's head. This experiment will show whether the devices are able to correctly calculate point of regard even though the observer has moved or changed positions. Image D contains 9 crosses, which are distributed in the same way as the image A, but only with less number of crosses.

2.5 Real-world coordination

For this kind of studies, using eye tracking devices (RED and HED) on printed images, output data from the realized experiment could not be immediately used for a data evaluation, especially for HED.

The obtained data from RED was recorded like a data file designed for BeGaze evaluation software from the company SensoMotoric Instruments (SMI). But for our purposes these data files were converted into a text file for evaluation outside BeGaze. The most important data in this file was the coordinates of the point of regard and the time stamp. For the HED experiment, videos were recorded in MPEG format. At the same time we could get a similar text file as for RED. The coordinate of the point of regard are here not relative to the actual stimuli, but to the position of the subject's head. Therefore we could not work with the same text file as in case of RED, thus the recorded MPEG video was just one possibility how to work on.

The proposed process used to find the real-worl coordinates can be divided roughly into four steps, as described below.

The model can perform stabilization in MPEG format, but to improve the speed of this stabilization model we converted each MPEG video into AVI. Otherwise the model has to encode the video to the AVI itself. Next step concerns about the stabilization of AVI video. In order to stabilize the video we used the Simulink demo model created by the Mathworks team. This model has been modified by adding a "Stop Simulation" Block, which stopped the process of the stabilization at the end of the video. This model uses the Video and Image Processing Blockset to remove the effect of camera motion from a video stream. The video stabilization process used the first frame of the stream as a reference. It means that each frame relates to this frame as reference. In this reference image the target needed to track the stabilization process was defined (in our case, it was the corners of the image). Simply, the model searched for the target and in each frame determines how much the target has moved relative to the reference frame. It uses this information to remove unwanted shaky motions from video and generate a stabilized video.

After the stabilization of the video, an algorithm was developed for finding real-world coordinates of the point of regard. The algorithm includes three phases. The first one calculated all coordinates (referential to the image in the video) from the video (one frame – one coordinate). The target as a circle indicates the point of regard in the video. By applying an image processing algorithm the center was found, which corresponds with the coordinate of the point of regard. By the second phase; the mentioned stabilization tries to find the best transformation (translation, rotation and scale), but when we saw on the results, the transformation still had some shift in the image due to the presence of distortion of the image caused by the lens of the camera. Therefore we chose to implement another stabilization algorithm to eliminate this shift. In order to make a new video stabilization, first, we need to find and calculate the four corner coordinates of the image. These coordinates are used to find the best transformation (translation and rotation) of the current image in order to match with the first frame. Just by minimizing the distance between the four corners coordinates (of the reference and the current frame), we are able to find the best one. The third phase; we got the coordinates which were related to the image in the video and not to the original image (420x297). Finally, we could find the transformation of the image coordinates in the video into the original image (the width and length of the image are compute thanks to the four corners) and the result was the real-world coordinates of the points of regard, which are relative to the actual stimuli. With these coordinates we could work on evaluation of the data.

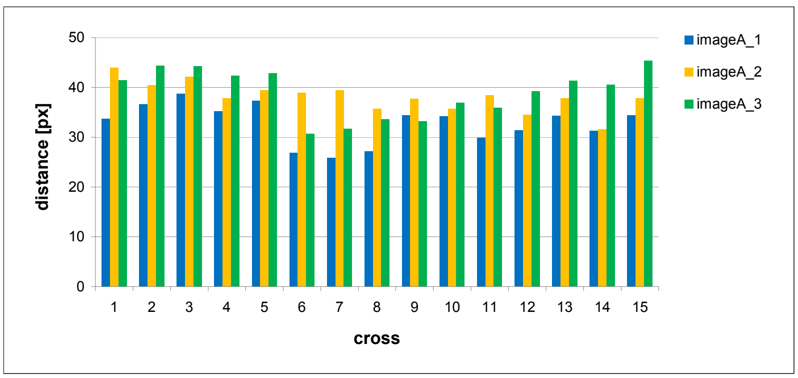
3. RESULTS AND ANALYSIS

The main values which were calculated from the points of regard were mean and median values. These values were calculated for each image. Only one value, mean or median, was chosen for the further data evaluation. In order to find out whether mean or median will better characterize of the resulting data, several values of mean and median were randomly chosen and were statistically analyzed. Standard deviations were computed for each value of mean and median and represented in graphs. We saw that the values overlapped in most cases in the graph and so it was not statistical significant, therefore were chosen lower values that were median values. Reason for this was that some point of regard could occur outside of observed object, and contribute to a high mean. This could be caused by passing between one cross to another cross by the eye, by blinking or misregistration from the eye tracker.

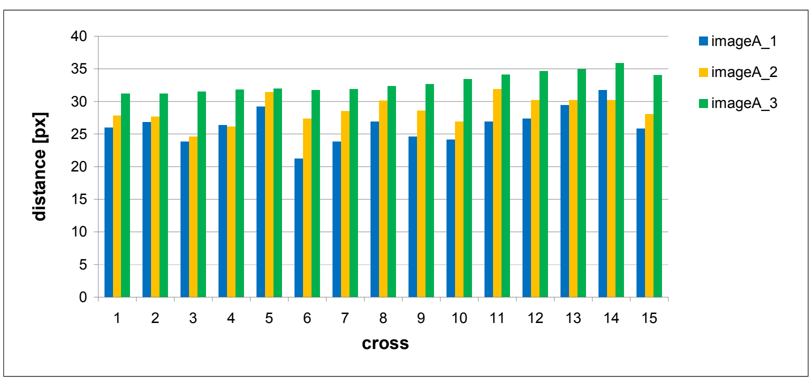
3.1 Evaluation and statistical analysis of first of three same images A

The process of the data evaluation was the same for all three images. As it was mentioned above, these images obtain 15 crosses. For each cross in the image and all observers, median of distances between points of regard and center point of the cross was calculated.

The aim was to investigate how precise the eye tracking devices are on each symbol (cross) in the image and how precise they are in the time aspect. The results are shown in Figure 3a and b. Figure 3a show that HED is more precise in the middle of image and in the right down corner and less precise in the first line of crosses (crosses 1-5 on Figure 2). In the Figure 3b the precision of RED is not clearly seen as in the case of HED. But RED is most precise in the middle of the image and the worst in the crosses on the bottom. As the time elapses the RED's precision is getting worse and worse. The HED is getting worse as well as RED, but in the middle of the image the precision is kept. The time interval between image A_1 and image A_3 is approximately 5 minutes.



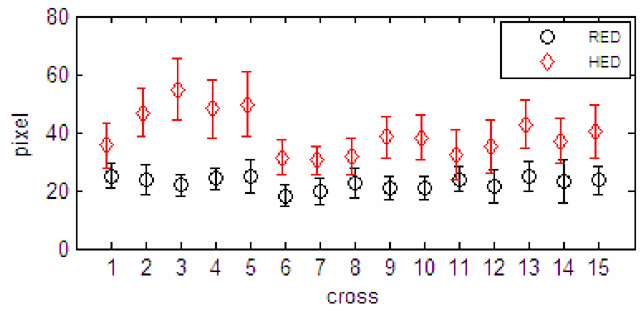
a)



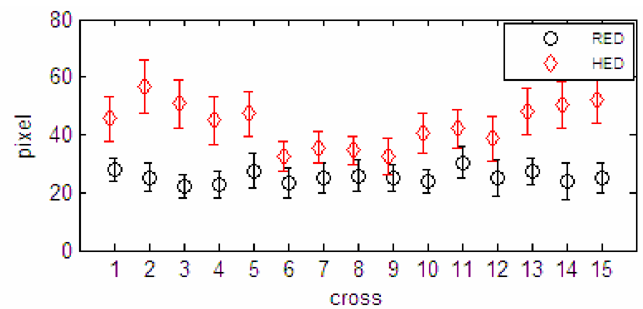
b)

Fig 3. a) Median distance between points of regard and centre of the cross for HED;
b) Median distance between points of regard and centre of the cross for RED

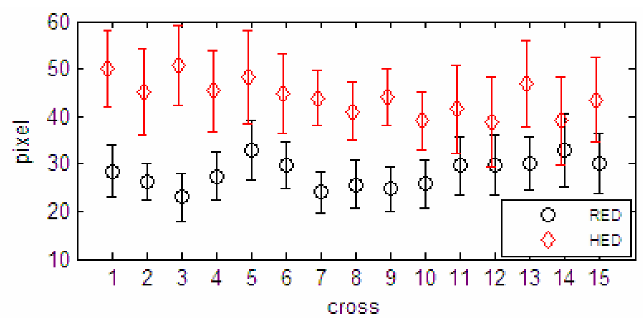
Figures 4a, b, and c are representing the standard deviations of the median distance between points of regard and centre of each cross. These standard deviations are compared between RED and HED, but for each image separately. In a most case the values are statistically non-significant. But small statistical significance is seen in the middle (crosses 6 – 9) of the image on imageA_1 and imageA_2 (Figure 4a and b) and in the lower line (crosses 11 – 15) of the image on the imageA_3 and bit on the imageA_1 (Figure 4c and a).



a)



b)

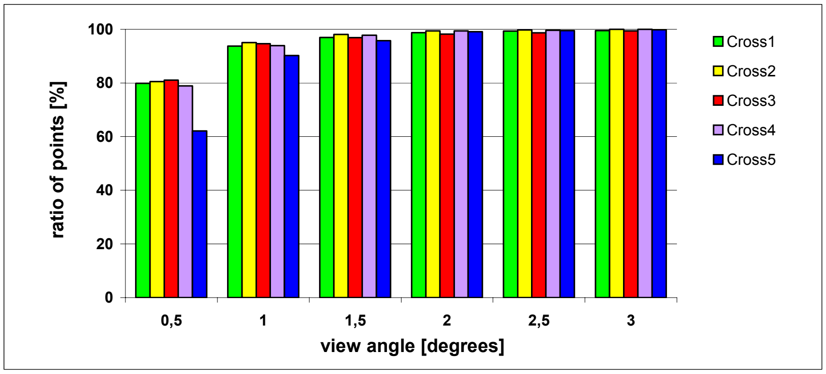


c)

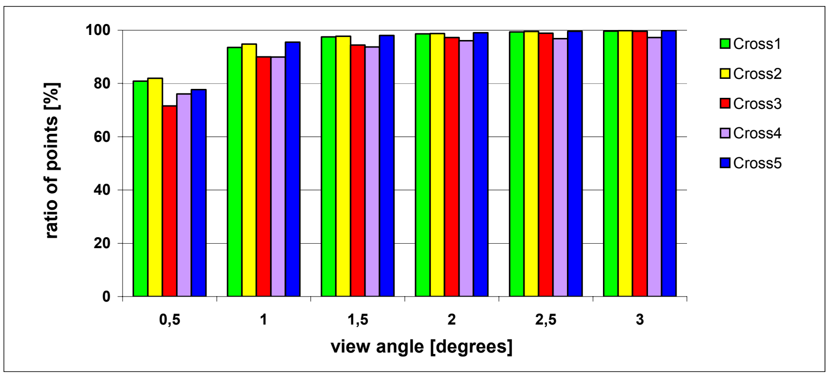
Fig 4. a) The standard deviations of the median of distances between points of regard and centre point of crosses in the image A_1;
b) The standard deviations of the median of distances between points of regard and centre point of crosses in the image A_2;
c) The standard deviations of the median of distances between points of regard and centre point of crosses in the image A_3

3.2 Evaluation and statistical analysis of second image B

This image had 5 crosses with determinated number of circles. In the first step we computed median of all points of regard for each cross in the image, then a distance between this median and the centre of the cross. In the second step we took the median of the points of regard and substitute it as centre of the new notional cross. Because the precision of fixation points of the eye tracking methods is not 100% perfect, then the center of the new notional cross is not on the same place as the original. But for the observers this place was as the center of the cross. Circles with view distance 0.5 degrees were subtracted from the center of the new notional cross. The percentage representation of points of regard in each circle of the notional cross is shown in Figure 5a and b. The RED and HED points of regard are almost 100% occupying 1.5° – 2° viewing angle.



a)



b)

Fig 5. a) Percentage representation of points of regard in each circle represent by viewing angle for HED;
b) Percentage representation of points of regard in each circle represent by viewing angle for RED.

When the values of RED and HED were subtracted from each other, we got the results shown in Figure 6. This figure shows that RED is better on target 5 and HED on target 3 and 4.

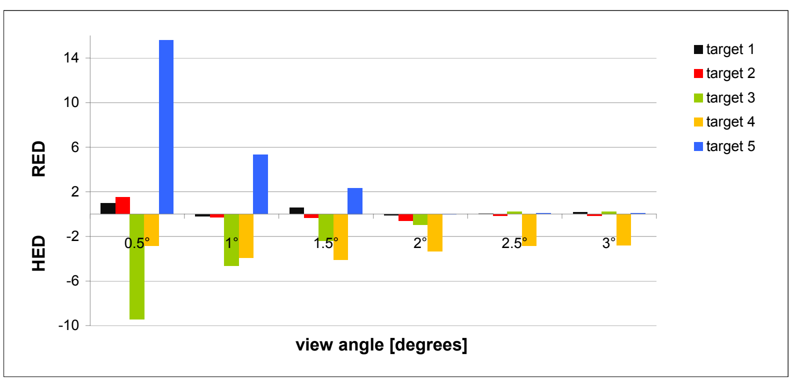


Fig 6. With a positive value the RED has a higher ratio of points inside the different viewing angles, and with a negative value HED has a higher ratio of points inside the different viewing angles.

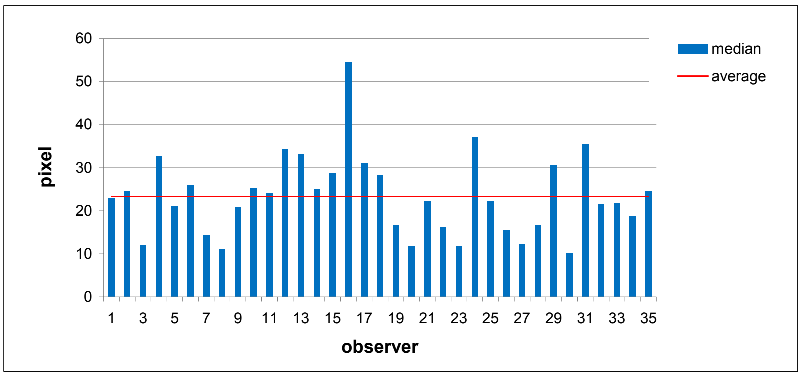
3.3 Evaluation and statistical analysis of third image C

This image represented a curve over whole image and 42 small crosses on this curve. In the first step we found every point of regard in the image and computed Euclidean distances between every point of regard and every of the 42 crosses on the curve. These Euclidean distances will constitute matrixes (one matrix, one observer) about 42 lines (42 crosses) and columns corresponding to a number ΔE of points of regard. In the second step, a local minimum is found corresponding to the minimum of each line of this matrix and is kept for further evaluation. By the evaluation as third step, mean, median, max and min distances were computed for every observer and subsequently computed average distances of mean, median, max and min of all observers (Table 1, figures 7a and 7b).

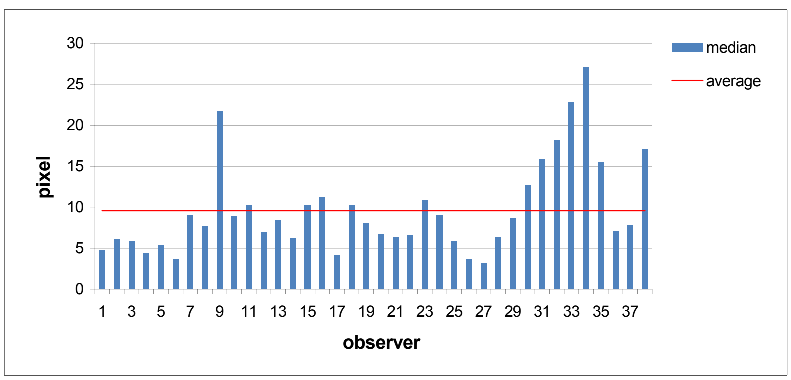
Because not all 40 values could be used, especially in case HED with video that was not very quality for the stabilization, in really there were used 38 values for RED and 35 values for HED.

Table 1. Overview of average distances of whole image for all observers

	Mean	Median	Max	Min
HED	27,24	23,30	82,52	1,74
RED	12,74	9,59	49,22	0,60



a)

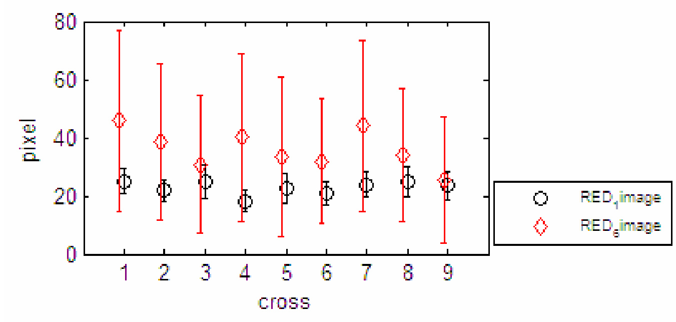


b)

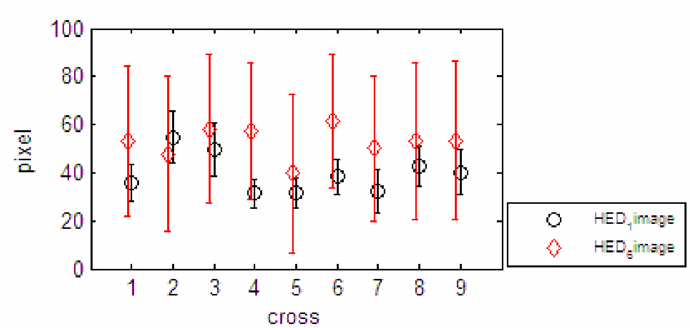
Fig 7. a) Median values of whole image C for every observer with average median values, for HED;
b) Median values of whole image C for every observer with average median values, for RED.

3.4 Evaluation and statistical analysis of fourth image D

Image D had 9 crosses. The process of the data evaluation was very similar as for image A. For each cross in the image and all observers, median of distances between points of regard and center point of the cross was calculated. This median of image D was compared with median of image A for same 9 crosses (Figures 8a and b). From these figures we can say the calibration is stable when an observer moves and leaves his or her position in front of the device, and returns to sit in the same position. But there is a question, why the standard deviation of the image D is so higher than standard deviation of the image A_1.



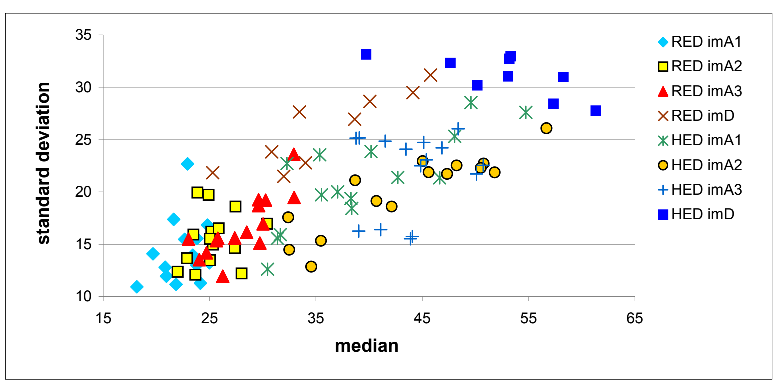
a)



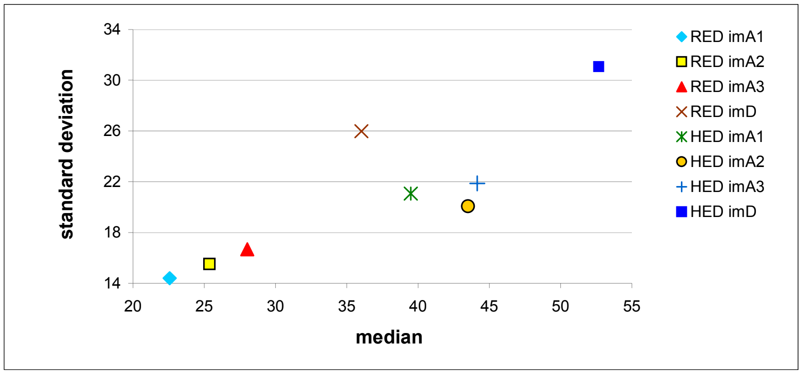
b)

Fig 8. a) The standard deviations of the median of distances between points of regard and centre point of crosses in the image A_1 and image D of RED.
b) The standard deviations of the median of distances between points of regard and centre point of crosses in the image A_1 and image D of HED.

For this understanding we did further Figures 9a and b. These figures show that when median is increasing than the standard deviation is increasing as well, indicating a larger spread of points not only a moved cluster of points. With time aspect the median of RED is continuously increased as the standard deviation, but the median of HED is increased continuously with standard deviation as well except image A_2.



a)



b)

Fig 9. a) Dependence of average standard deviations and median distances of all observers for 15 crosses of image A_1, 2, 3 and for 9 crosses of image D; for RED and HED values;
b) Average values of values from Figure 9a) in dependence of standard deviation on the median distance of image A_1, 2, 3 and for 9 crosses of image D; for RED and HED values.

It is evident from the figures, that RED is the best method for use on printed images. It can be seen in Figures 9a and 7b. The RED has lower median and standard deviation. The HED seems to be the best in the Figure 4, where is obviously seen a preponderance of values in all degrees and on target 3 almost too. Only for target 5 the RED is evidently better. For both devices, when observers get longer into the experiment, the median and standard deviation increases.

On the end of the experiment the observers filled out a questionnaire. The observers were asked to say which image was most problematic to carry out the experiment. The most problematic image was the image A and the image C. Image A was longest task and image C was the most difficult because it required a lot of concentration of the observers. Next important question has been built in a psychological way. The observers should answer which eye tracking device was more comfortable for realization of the experiment. Most observers felt more comfortable with RED than HED.

Causes why the observers felt worse with HED than RED were:

- Size of the helmet (too small or too big). Observer worried about helmet movement instead of concentrating on the experiment.
- Observer did not feel free; the weight of the helmet was noticeable by the observers and after a while it became uncomfortable to wear and to heavy. In case of RED observer did not have to wear anything and did not realize that the camera was present.

Causes why the observers felt better with HED than RED were:

- Observers had more possibility to move by the head, because the helmet was as a part of the observer and therefore they could concentrate better.
- The position of sitting was better to look at the image and observers had a feeling that did not have to focus so much with the helmet.

At the start of this paper there was said, that the calibration of the eye tracking equipment is very necessary and critical part of the experiment. If an eye was difficult to track, the calibration was very difficult to manage as well, it took too long and observer became tired. It could be a cause of bad results of the experiment.

4. CONCLUSION

If HED can be used on printed images and its comparison with RED was investigated. The stabilization and transformation to get the real-world coordinates was also found. The results show, that RED is the best for printed image evaluation in most of the cases and HED was the best only in one case. The difference of the precision between HED and RED is about 10 – 16 pixels, it is approximately 3.49 – 5.584 mm. Attempt at a better stabilization and transformation, which was done just for two directions (translation and rotation) instead of three directions (translation, rotation and scale), can acquire better results in the HED case. The change of the lens could be the next cause of getting better results.

But in the first step, how to get the real-world coordinates, the results are not so different. Majority of the observers said that the RED was more comfortable than HED, but it is also observer dependent.

ACKNOWLEDGMENTS

We would like to thank Faouzi Alaya Cheikh for the help with stabilization of the video. We would also like to thank Damien Lefloch for the processing of the algorithm for the transformation of the images.

REFERENCES

[1] Goldberg, J. H., Stimson, M. J., Lewenstein, M., Scott, N. and Wichansky, A. M., “Eye Tracking in Web Search Tasks: Design Implications,” ETRA ’02. New Orleans, Louisiana USA, (2002).

[2] Cowen, L., Ball, L. J. and Delin, J., “An eye movement analysis of webpage usability,” In Proceedings of the annual HCI 2002, 317-335 (2002).

[3] In Hyönä, J. R. and Deubel, H. (Eds.), [The mind's eye: cognitive and applied aspects of eye movement research], Radach, R., Lemmer, S.,Vorstius, Ch., Heller, D. and Radach, K., “Eye movement in the processing of print advertisements,” Elsevier Science Ltd., 609-632 (2003).

[4] In Hyönä, J. R. and Deubel, H. (Eds.), [The mind's eye: cognitive and applied aspects of eye movement research], Holmqvist, K., Holsanova, J., Barthelson, M. and Lundqvist, D., “Reading or scanning? A study of newspaper and net paper reading,” Elsevier Science Ltd., 657-670 (2003).

[5] Babcock, J. S., Pelz, J. B. and Fairchild, M. D., “Eye tracking observers during rank order, paired comparison, and graphical rating tasks,” IS&T PICS Conference, Rochester, 10-15 (2003).

[6] Ninassi, A., Le Meur, O., Le Callet, P., Barba, D. and Tirel, A., “Task Impact on the Visual Attention in Subjective Image Quality Assessment,” In The 14th European Signal Processing Conference, (2006).

[7] Mulligan, J. B., “Image processing for improved eye-tracking accuracy,” Behav Res Methods Instrum Comput., 54-65 (1997).

[8] Gintautas, D. and Ramanauskas, N., “The accuracy of eye tracking using image processing,” Proceedings of the Third Nordic Conference on Human-Computer Interaction October 23-27, 2004, Tampere, Finland, 377-380 (2004).

[9] Zhu, J. and Yang, J., “Subpixel Eye-Gaze Tracking,” Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition, May 20-21, 131 (2002).

[10] Zhu, Z., Fujimura, K., and Ji, Q., “Real-Time Eye Detection and Tracking Under Various Light Conditions,” Proc. ETRA 2002, ACM Press, 139-144 (2002).

[11] Nevalainen, S. and Sajaniemi, J., “Comparison of Three Eye Tracking Devices in Psychology of Programming Research,” Proceedings of the 16th Annual Workshop of the Psychology of Programming Interest Group, Carlow, Ireland, 151-158 (2004).

[12] DeSantis, R., Zhou, Q. and Ramey, J., “A Comparison of Eye Tracking Tools in Usability Testing,” STC Proceedings, Usability and Information Design, (2005).

[13] Li, F., Kolakowski, S. and Pelz, J. B., “A model-based approach to video-based eye tracking, ” Journal of Modern Optics (2007).

[14] Kolakowski, S. M. and Pelz, J. B., “Compensating for Eye Tracer Camera Movement,” Proceedings of the 2006 Symposium on Eye Tracking Research & Applications, San Diego, California USA, (2006).

[15] In Hyönä, J. R. and Deubel, H. (Eds.), [The mind's eye: cognitive and applied aspects of eye movement research], Goldberg, J. H. and Wichansky, A. M., “Eye tracking in Usability Evaluation: A Practitioner’s Guide, North-Holland, 493–517 (2003).

[16] SMI (SensoMotoric Instruments). iViewX system manual, document version 1.05.09, (2005).

[17] Pedersen, M., Hardeberg, J. Y. and Nussbaum, P., "Using gaze information to improve image difference metrics," Human Vision and Electronic Imaging XIII, Proceedings of SPIE Vol. 6806-35, (2008).