Webcam as Alternate Option for Eye-Trackers in Gaze Gaming Software: GazePlay

Nairit Bandyopadhyay, Sébastien Riou, Didier Schwab

Webcam as Alternate Option for Eye-Trackers in Gaze Gaming Software: GazePlay

Nairit Bandyopadhyay **Sébastien Riou** **Didier Schwab**
Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, 38000 Grenoble, France
Nairit.Bandyopadhyay@grenoble-inp.org
{sebastien.riou, didier.schwab}@univ-grenoble-alpes.fr

Abstract

Gaze gaming software like GazePlay are often used by people with multiple disabilities especially children as a tool for learning. Interaction with this software can be done with a special type of device known as eye-trackers. These eye-trackers capture the gaze of the player. The eye-trackers are often very expensive and it is not affordable for everyone. This report is aimed at evaluating whether webcams can be used instead of eye-trackers (like Tobii Eye-tacker 5) for gaze interaction. We have developed a methodology that can be extended to test any gaze-tracking device and their related software. We evaluated our methodology by performing remote experiments with ten subjects. We conclude that webcam can be a viable option if the user is willing to accept some trade-offs with respect to performance and also depending on what kind of interaction he usually performs.

1 INTRODUCTION

Gaze is often considered as a natural way by which individuals with multiple disabilities interact with the environment. Gaze gaming software like GazePlay[1] is often used by children as a way to learn as discussed by Schwab *et al.* [2018]. The most common way of interacting with such software is by using gaze tracking systems. Not only children, the gaze tracking system helps any people with severe motor disabilities to communicate only with their eye movements. Among these gaze tracking systems, eye-trackers are the most popular one. These eye-trackers record the movement of the eye and accordingly replicates the movement of the cursor on the screen. It is like instead of using hands, one will be using their eyes to control the cursor. However, the eye-trackers are expensive and thus not affordable for all. The average starting price of an inexpensive eye-tracking hardware is around $250.There are various reasons for an eye-tracker to be expensive like usability, flexibility, support, compatibility with advanced modules etc. In most cases, expensive eye-trackers produce very precise and accurate outputs. As an alternative for such expensive device, low-cost and off-the-shelf hardware are gaining popularity with the improvement in camera technology. One such inexpensive hardware is webcam. It is similar to eye-trackers and it can also record eye movements. The price of an average high-definition (HD) webcam ranges from $8 -$20.Thus it is significantly less expensive compared to eye-tracker. We will try to evaluate the performance of webcam when it is used as an input device for playing the games on the GazePlay gaze gaming software and compare it with the performance of an eye-tracker (like Tobii[2]). For this we will be conducting scientific experiments and evaluate these two devices on the basis of certain performance metrics that we have defined in our research.

2 RELATED WORKS

The scientific community has made consistent efforts to establish gaze-based interactions as means of communication between individuals with disabilities. Previously, Townend *et al*. [2016] shows some work specific to people with Rett syndrome. Also, Maurer *et.al.* [2018] showed how to utilize gaze interaction in online multiplayer gaming.

We also looked for previous works on eye-tracking with webcams. In particular we were interested on how authors have evaluated the performance of webcams as eye-tracking devices. Previously Skovsgaard *et al.* [2011] has shown that precision and accuracy are two very important measurements for an input device like webcam. So, we will be including that in our methodology. Burton *et al.* [2014] has shown the importance of size and the placement of static stimulus on the screen. Apart from that, Papenmeier *et. al* [2010] has shown how dynamic object can be recognized with a head mounted eye-tracker by fixing an area of interest (AOI) on the screen. In our experiments, we will be using the GazePlay software for providing both static and dynamic stimulus to the players. Apart from that the quality of webcam will also play an important factor as shown by Sarkar *et al.* [2017] during the experiments. We will also need to do some calibration procedure. The aim is to map the raw eye-tracking data to the pixels i.e., the coordinates of the screen. However, this step, though important, can sometimes introduce error in the measurement because we have limited control in accurately directing our gaze, as pointed out by Reingold [2014].

1 https://gazeplay.github.io/GazePlay/

2 https://www.tobii.com/

3 PERFORMANCE METRICS

3.1 Accuracy and Precision

Both the eye-tracker and the webcam are sensors and the performance of a sensor is typically measured in accuracy and precision. Accuracy refers to the degree to which sensor readings represent the true value of what is measured, while precision (also known as spatial precision) refers to the extent to which successive readings of the same physical phenomenon agree in value, as described by Wilson [2007].

Skovsgaard *et al.* [2011] has measured the accuracy and precision of an eye-tracking system. Based on that, we have defined accuracy A as the average distance (d_i) between n fixation locations and the corresponding fixation targets (see Equation 1)

$$A = \frac{1}{n}\sum_{i=1}^{n} d_i \tag{1}$$

Spatial precision P is calculated as the Root Mean Square (RMS) of the distance (d_i) between successive samples (see Equation 2)

$$P = \sqrt{\frac{1}{n}\sum_{i=1}^{n} d_i^2} \tag{2}$$

Under ideal situation, when the gaze tracking devices perform optimally, then both A and P should have numerical value of 0 pixels.

3.2 Gaze Speed

Another important feature of a sensory device like webcam or an eye-tracker is latency. It means how much time difference is there between actual occurrence of the event and the manifestation of the event. In our system, as soon as the player moves his eyes, the gaze on the screen should change its position. To measure this latency, we are using a concept known as the Gaze Speed. We define Gaze Speed as S and it is represented as (see Equation 3)

$$S = \frac{Distance_{i,j}}{time_{reaction}} \tag{3}$$

Here $Distance_{i,j}$ is the distance between the initial gaze point and the final gaze point and $time_{reaction}$ is the time needed by the player to select the dynamic object on the screen using his gaze.

4 METHODOLOGY FOR EVALUATION

This section deals with the methodology that we prepared to do the experiments so as to evaluate the performance of the gaze tracking systems. This methodology can be adapted and extended to test any gaze-tracking system using a stimulus providing software, like GazePlay that we have used in our experiments.

4.1 Adapting the GazePlay to work with webcams

Our GazePlay currently supports interaction with eye-trackers so we will be using a separate software known as GazePointer that can work with webcam. The reason for choosing GazePointer is it is stable, open-source, has in-built calibration system and it has the option to control mouse cursor with gaze; exactly what we are looking for. The webcam is the hardware which will capture the eye movements. On software side, the GazePointer will process this data obtained i.e., the eye-movements obtained from the webcam and accordingly control the position of the cursor. Now we can launch the GazePlay and play any games because we have already established the interaction between the player and the Gaze Play through GazePointer.

4.2 Choice of Games

Out of several games provided by the GazePlay, a total of three games were considered for the experiments. This includes the Egg game, Creampie game and the Ninja game. We used the Creampie and the Egg game for evaluating the accuracy and precision of the gaze-tracking systems. In egg game, a static image is shown at the center of the screen which the user has to touch by using his gaze. In the Creampie game, a static image appears randomly at different parts of the screen, which the user has to touch by using his gaze. The Ninja game is used for calculation of gaze speed. In this game a dynamic object moves on the screen which the user has to touch using their gaze.

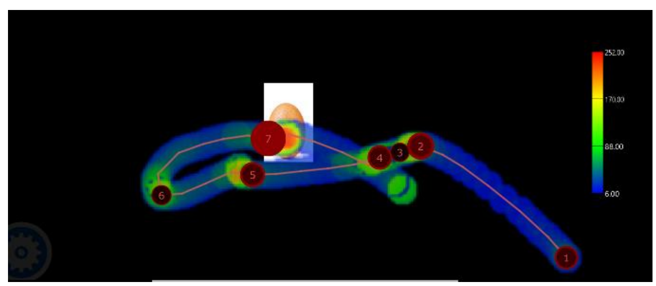

Figure 1: Egg game where the egg image is at the center of the screen and the lines show the gaze points on the screen.

4.3 Experiment setup

Pilot experiment

Initially we did a pilot experiment with three participants. We asked them to play the games that we considered for our experiments. All of them had previous experience with GazePlay. This experiment was done remotely where the participants were asked to download the GazePlay and the GazePointer before the experiment. These three participants had the eye-tracker as well as the webcam with them, and that is why they were selected for the experiments. The target of these pilot experiment was to see what practical challenges were we facing while doing the experiments. Also, since the eye-tracker is an expensive device and it is not available to everyone, we also used the results obtained from the tests

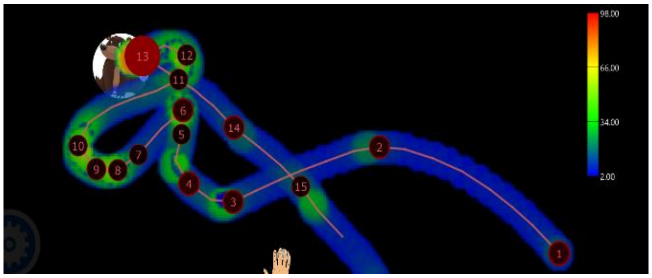

Figure 2: Creampie game where the image appears at random positions of the screen and the lines show the gaze points on the screen.

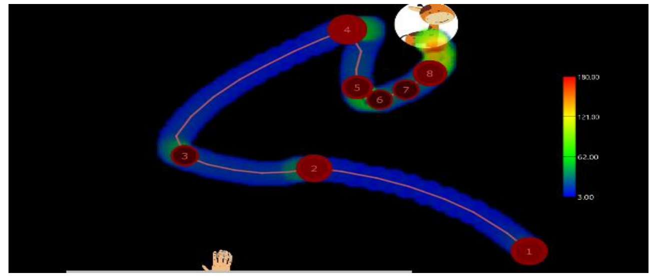

Figure 3: Creampie game where the image appears at random positions of the screen and the lines show the gaze points on the screen.

using eye-trackers during these pilot testing as a benchmark to evaluate the performance of the webcam at the next stage of our testing.

Actual experiment

We selected seven participants for the actual experiments from the age group between 21-28. All of them had at least an undergraduate level of education. The experiments were done remotely. At the beginning of the experiment, they were asked to fill a questionnaire which asked them some basic questions like their age, education level, whether they are wearing any glasses or not, their eye color, whether they have ptosis or droopy eye-lids etc. We also asked them to provide the specifications of their computer and the webcams they are using. Then they were asked to install the GazePlay and the GazePointer. The details of how the experiments were conducted are described in the next section (Section 4.4).

Remote Experiments

All the experiments were done remotely. It not only helped us to avoid the sanitary conditions related to the pandemic but also it helped us in including many participants. Scientific research conducted by including the common people is known as Citizen science which improves the capacity of scientific communities and increases the public's understanding of the science as pointed out by Hand [2010].

4.4 Design and Procedure

As discussed in the previous section, at first the participant is asked to fill the questionnaire. Then they are asked to install and launch the two software: GazePlay and GazePointer separately. The participant has to calibrate their eyes using the Gaze calibration feature provided by the GazePointer. Basically, the player has to look at various positions on the screen as instructed by the GazePointer. After calibration is over, the GazePointer is ready to be used with GazePlay for playing the games.

The participant first plays the Egg game in three sets. In each set, the participant has to select the image of an egg which is on the center of the screen before the timer runs out. For this game, we set the timer for eight seconds. Once the timer runs out or the participant is successful in selecting the image of the egg with his gaze, the game proceeds to the next set. We take note of the center coordinates of the image of the egg and ask the participant to share the JSON file that was generated when the Egg game was being played. The JSON file contains the mouse coordinates at different timestamps. We will use the distance between these mouse coordinates from the center coordinates of the image of the egg to calculate the precision ad accuracy in each of the sets of the Egg game.

Next the participant plays the Creampie game in three sets. Here an image appears on random position of the screen, contrary to the Egg game where the image is always at the center. The participant has to select this image before the timer runs out. For this game, we set the timer for six seconds. Once the timer runs out or if the participant is successful in selecting the image, the game proceeds to the next stage. We take note of the center coordinates of these images and ask the participant to share the JSON file that was generated when the Creampie game was being played. The JSON file contains the mouse coordinates at different timestamps. We will use the distance between these mouse coordinates from the center coordinates of the images to calculate the precision and accuracy in each of the sets of the Creampie game.

During the pilot experiment, we also asked the participant to play the Ninja game. In this game an image appears on the screen and it keeps on moving. We look into the JSON file which has the initial coordinate of the gaze as well as the final coordinate of that gaze (when the player selected that object). Difference between these two coordinates provide us the $Distance_{i,j}$ in Equation 3 (Section 3.2).The JSON file also provides us with the $time_{reaction}$ in Equation 3 (Section 3.2). Using these two values we can calculate the Gaze Speed *(S)* as mentioned in Equation 3 (Section 3.2)

While designing the games for the experiments, we have decided to keep only single image (object) on the screen. This is because we want the user to look at one image only and not get distracted by other images on the screen. So, by placing only one image on the screen for each set, we have actually fixed the Area of Interest (AOI) to one.

4.5 Hardware

Since we did the experiments remotely, we did not arrange for any specific hardware except the Tobii Eye-tracker 5. From the data collected from the participants, we found that four participants had used standard 640x480 VGA webcams provided in the laptops. Remaining participants used 1280x720 High-definition webcams. Three of the participants had Intel i7, two had Intel i5, two had i3 and the remaining had AMD Ryzen 5 as the processing unit.

5 RESULTS

5.1 Accuracy and Precision

When the image is placed on the center of the screen

This scenario corresponds with the Egg game. In the Egg game, the image is always located at the center of the screen. We have obtained the following results from our experiments:

Type	Average Accuracy (pixels)	Average Precision (pixels)
Webcam	465.95	503.77
Eye-tracker (Tobii 5)	432.52	488.74

Table 1: Average Accuracy and Precision for a static image with fixed position

When the image appears randomly on different positions of the screen

This scenario corresponds with the Creampie game. In the Creampie game, the image appears at random positions of the screen each time a game is played. We have obtained the following results from our experiments:

Type	Average Accuracy (pixels)	Average Precision (pixels)
Webcam	435.78	507.56
Eye-tracker (Tobii 5)	304.67	345.18

Table 2: Average Accuracy and Precision for an image appearing randomly on different positions

5.2 Gaze Speed

Here a dynamic object will move on the screen (horizontally or vertically). Our Ninja game corresponds to this scenario. We have obtained the following results from our experiments with the Gaze tracking systems:

Type	Average Speed of Gaze (pixels/millisecond)	Average Reaction time (millisecond)
Webcam	0.178	3465
Eye-tracker (Tobii 5)	0.22	2502.5

Table 3: Average Speed of Gaze and Reaction time

6 DISCUSSIONS

From Table 1 (Section 5.1) we can see that the average accuracy and precision for an eye-tracker is slightly better than the webcam when the image is placed on the center of the screen. For an ideal gaze-tracking system, the accuracy and precision should be 0 pixels, according to the definition and the expression provided in Section 3.1. This means, greater the numerical value of accuracy and precision, poorer will be the performance. The webcam has about 10% lesser accuracy and 3% lesser spatial precision than the eye-tracker

From Table 2 (Section 5.1) we can see that the average accuracy and precision for an eye-tracker is significantly better than the webcam when the image appears on random positions of the screen. The webcam has about 40% lesser accuracy and 47% lesser precision compared to the eye-tracker. This drop in performance of the webcam can be due to certain factors like size of the image as well the webcam being unable to track the gaze when there is a significant eye movement to select the images randomly appearing at the edges and corners of the screen.

From Table 3 (Section 5.2) we can see that the Gaze Speed for the eye-tracker is 24% higher than the Webcam. When the average distance covered by the gaze on the screen remains same, then the Gaze Speed of Eye-tracker is higher compared to the webcam. This is because the reaction time in eye-tracker is lesser. This means that in case of eye-tracker, there is less latency in the interaction. Having low latency provides a smooth interaction between the user and the system.

7 CONCLUSIONS

Our study on the performance evaluation shows that a webcam can provide a performance comparable to expensive systems like eye-trackers. If we take into account the fact that the price of a webcam is significantly lesser than an eye-tracker, then we can consider the performance of the webcam as satisfactory.

However, there are certain **limitations** in our study. We left out other crucial factors that have not been evaluated in this study such as tolerance against head movements, performance under varying degrees of luminosity, distance between the eye and the webcam. Since all the experiments had been done remotely, it was not feasible to create various extreme situations for stress testing the gaze tracking systems.

In our **future work**, we aim to further investigate these issues and would like to perform more complex experiments in a lab setup. We would also like to explore on how we can perform calibrations at specific time intervals so that it can improve the performance. We also plan to conduct tests using Tobii 5 eye tracker and the webcam where the audience will comprise of people with multiple disabilities. The results will be beneficial in creating an inexpensive gaze tracking system which can be afforded by everyone.

ACKNOWLEDGEMENTS

We are thankful to the entire **GazePlay (InterAACtion)** team along with the **GETALP (LIG)** team for their support in this work and also to the users without whom we would have not been able to do the experiments.

References

[Schwab *et al*., 2018] Schwab, Didier & Fejza, Amela & Vial, Loïc & Robert, Yann. (2018). The GazePlay Project: Open and Free Eye-Trackers Games and a Community for People with Multiple Disabilities. 10.1007/978-3-319-94277-3_41.pages 1-2.

[Skovsgaard *et al.*, 2011] Skovsgaard, Henrik & Agustin, Javier & Johansen, Sune & Hansen, John & Tall, Martin. (2011). Evaluation of a remote webcam-based eye tracker. ACM International Conference Proceeding Series. 7. 10.1145/1983302.1983309.page 2.

[Burton *et al.*, 2014] Burton, Liz & Albert, William & Flynn, Mark. (2014). A Comparison of the Performance of Webcam vs. Infrared Eye Tracking Technology. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 58. 1437-1441. 10.1177/1541931214581300.page 2.

[Sarkar *et al.*, 2017] Sarkar, A. & Sanyal, Prof (Dr.) Goutam & Majumder, S. (2017). Performance comparison of different low cost cameras for development of an eye-tracking system. International Journal of Applied Engineering Research. 12. 1819-1825. page 3.

[Reingold, 2014] Reingold, Eyal. (2014). Eye Tracking Research and Technology: Towards Objective Measurement of Data Quality. Visual cognition. 22. 635-652. 10.1080/13506285.2013.876481.pages 635-652.

[Holmqvist *et al*, 2011] Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). Eye tracking: A comprehensive guide to methods and measures. Oxford: University Press. Pages 174-189.

[Townend *et al*.,2016] Townend, Gillian & Marschik, Peter & Smeets, Eric & Berg, Raymond & Berg, Mariëlle & Curfs, L.. (2016). Eye Gaze Technology as a Form of Augmentative and Alternative Communication for Individuals with Rett Syndrome: Experiences of Families in The Netherlands. Journal of Developmental and Physical Disabilities. 28. 10.1007/s10882-015-9455-z.pages 1-11.

[Maurer *et.al.*,2018] Maurer, Bernhard & Lankes, Michael & Tscheligi, Manfred. (2018). Where the Eyes Meet: Lessons Learned from Shared Gaze-based Interactions in Cooperative and Competitive Online Games. Entertainment Computing. 27. 10.1016/j.entcom.2018.02.009.pages 47-59

[Papenmeier *et. al*,2010] Papenmeier, Frank & Huff, Markus. (2010). DynAOI: A tool for matching eye-movement data with dynamic areas of interest in animations and movies. Behavior research methods. 42. 179-87. 10.3758/BRM.42.1.179.pages 179-187

[Wilson, 2007] Wilson, A.D. (2007). Sensor- and Recognition-Based input for interaction: In the Human Computer Interaction Handbook. Lawrence Erlbaum Associates. pages 177–199

[Hand, 2010] Hand, Eric (2010). Citizen science: People power. *Nature* 466. 10.1038/466685a.pages 685-687