Investigating Eye-Tracking in 3rd Party Off-the-Shelve Software

Mary Igbudu [1], Johannes Tümler [2]

Abstract: Eye-tracking and its uses to assess visual attention in head-mounted displays opens many possibilities for virtual reality experiences. Eye-tracking in virtual reality can reveal what is behind an individual's subconscious reaction and behavior when interacting with environments

Several Virtual Reality headsets have integrated eye-tracking capabilities. In order to use the eye-tracking feature, it usually is necessary to implement specific SDKs into the Virtual Reality software. The focus of this paper is to present an approach to use such systems to investigate eye-tracking data in third party off-the-shelve Virtual Reality software without integration of specific SDKs. Also investigated is the measurement and forecasting of user performance based on the eye-tracking data.

A randomized controlled study was conducted with 20 participants. They used the Varjo VR-2 head-mounted display with integrated eye-tracking capabilities in a biomedical teaching application. The method proved to be successful. A link was found between visual attention and educational outcomes in biomedical training with a positive and strong correlation.

Keywords: Eye-tracking, Virtual reality, cardiac anatomy, Varjo VR-2, Head mounted display, Fixation, Gaze.

1 Background, Motivation and Goals

Eye-tracking has become a valuable tool in numerous research areas such as cognitive psychology, usability marketing, information visualization research, on-road driving applications, and assistive technology. Eye-tracking analysis has helped researchers determine what an individual sees, providing clues to what a person could be cognitively engaged in and supplying an informative window into a person's thoughts and intentions. Eye-tracking is the measurement of eye movement and gaze activity. The eye is one of the most acute and critical senses in humans [Ch13]. Nearly 80% of all sensory impressions are delivered to the brain via the visual channel. Vision also provides information at the highest speed and the visual cortex, in turn, accounts for about 60% of the whole cerebral cortex, considering all areas responsible for visual stimuli [Pl18]. Furthermore, cognitive behavior involves measuring visual, verbal, and motoric operations.

[1] Hochschule Anhalt, Biomedical Engineering, Bernburger Straße 55, 06366 Köthen, Deutschland, igbudu54@gmail.com, https://orcid.org/0000-0002-2849-8993
[2] Hochschule Anhalt, Ingenieurinformatik/Augmented und Virtual Reality, Bernburger Straße 55, 06366 Köthen, Deutschland, johannes.tuemler@hs-anhalt.de, https://orcid.org/0000-0002-4788-2667

doi: 10.18420/delfi2022-ws-14

Eye movements are human behavior that occurs averagely three times per second [Tr18]. It is an indicator of the data which is being received at the actual moment by the subject. It allows to observe and analyze the way a person looks at an object. This makes it possible to see in detail what is at the central direction of gaze as well as follow the visual attention path of the user [Ho17]. By tracking and analyzing these eye movements, one can gain valuable insights into human behavior, physiology, psychology, perception, and visual attention. For teaching scenarios, it can aid to detect and guide students' attention. Modern eye-tracking involves an array of infrared or near-infrared light sources and cameras that track the gaze of one (monocular) or both (binocular) eyes [Tr18].

In most modern systems, an array of non-visible light sources illuminates the eye and produce a corneal reflection (the first Purkinje image); the eye tracker monitors the relationship between this reflection and the center of the pupil to compute vectors that relate eye position to locations in the perceived world [Ho17]. As the eyes move, the computed point of regard in space also moves. Due to the complexity of the spatiotemporal gaze data, we usually must derive other, more simplified dependent variables from the raw gaze data to perform eye data analysis. An important class of analysis approaches is based on eye-tracking metrics derived from the processed eye-tracking data. Objects or specific regions on a stimulus can be of special interest for easy analysis [Ku16]. By defining boundary shapes around these areas of interest (AOI), fixation data can be mapped to these areas. Common eye-tracking metrics involved in eye-tracking studies and relevant to this study are fixation, gaze point, areas of interest, time spent, dwell time, revisit time, and fixation sequence. Others like saccades, smooth pursuit and respondent count are also important metrics when carrying out eye-tracking studies.

The effectiveness of eye-tracking technology in various fields, its application in medical education, and the possibility of using it to examine healthy people, make eye-tracking technologies suitable as a forecasting tool when investigating an individual's decision-making process or cognitive behavior in a virtual environment. Medical education is associated with several problems which lead to a limited spatial experience obtained from lectures and anatomic dissection [Pl18]. However, VR has proved a valuable tool for training in all aspects of health sciences, particularly in anatomy medical education.

Virtual Reality has offered educational potentials in numerous areas such as data gathering and visualization, project planning and design, the design of interactive training systems, virtual field trips and the design of experiential learning environments [Mc03]. However, in medicine, it offers tremendous potential, both as a tool for medical practice and for training medical students. In terms of medical education, numerous businesses have introduced VR surgical simulators that provide both visual and tactile feedback [Ba13]. One particularly interesting use of eye-tracking in Virtual Reality (VR) is foveated rendering where the system will render higher-quality graphics at the gaze point and lower quality at the periphery [Mi22]. Compared to real-world eye-tracking, VR has the advantage of the more straightforward definition of the regions that user had looked at.

This paper presents a study carried out together with another previously reported study (see [Od22]). It assessed the effectiveness of using non-customized, predesigned off-the-shelve Virtual Reality software to teach specific biomedical knowledge. Therefore, the experimental setup, materials and methods, and part of the procedures used in this research are the same as that from the previously reported study. This paper aims at presenting an approach to investigate eye-tracking data in third party off-the-shelve software. To be able to discuss the expected results, the following research questions served as a compass:

> **RQ 1.** Is it possible to record eye-tracking data from off-the-shelve VR software without implementation of specific eye-tracking SDKs?
>
> **RQ 2.** Is it possible to see gaze replays and determine the participants' fixation from the eye-tracking data?
>
> **RQ 3.** From the analysis of eye-tracking data, is it possible to forecast the user's learning performance?

These questions will help answer if eye-tracking data from an off-the-shelf VR software is usable and if it gives quantifiable results that can be used to forecast user performance in medical education.

2 Use Case and Implementation

Usually, if researchers want to investigate users gaze through eye-tracking in Virtual Reality, it needs software developers to change the source code of the VR application: An eye-tracking software development kit (SDK) needs to be implemented. Such an SDK will allow direct access to eye-tracking data as 3D point sets in the virtual scenery. The gazed-at 3D objects can be identified directly by comparing the 3D gaze point from eye-tracking with the 3D data from the virtual world. But, in most cases, VR users do not have access to the source code of the VR application, thus an integration of the eye-tracking SDK is impossible. In such cases it is very difficult to use eye-tracking at all. We present an approach, which allows a context-based eye-tracking analysis in off-the-shelve software.

This study implemented a biomedical use case where users studied the human heart in Virtual Reality. We used the Varjo VR-2 head mounted display, which has eye-tracking capabilities. It offers dual 1920 x 1080 px display resolution, combined with secondary dual 1440 x 1600 px displays, IPD range of 61-73 mm, and 87° horizontal field of view [Va22]. The software used is *ShareCareYOU VR* purchased from Steam. It is a real-time simulation software, which allows free navigation and exploration of 3D models of the human body, its anatomical structures, organs, and their natural functions [Sh22]. It provides 3D models, videos, informative text, interactive tools, and functionalities to dissect to see internal object structures. It also has an audio interpretation for easy assimilation and is compatible with all major VR headsets. The content of the learning materials was verified for correctness and relevance by an expert surgeon with several years of medical experience [Sh22], [Od22]. Figure 1 provides an overview of a complete setup showing a user with the Varjo VR-2 HMD and the *ShareCare YOU VR* software.

Figure 1 - A user with complete eye-tracking / VR setup

3 Procedure

Although the main procedure was reported in a previous publication (see [Od22]), it has been summarized here again for completeness. Twenty participants from the Biomedical Engineering Master's degree program at Anhalt University of Applied Sciences took part in this study. They belonged to the experimental group of the previously reported study. They were volunteers and received no compensation for their participation. Three (15%) were between the ages of 21-25 and 17 (85%) between the ages of 26-30. Before the study, the participants were given a demographic survey to determine their neurological, mental, and visual states. Information on if they had prior cardiac knowledge or VR experience were collected. Also, the students had to undergo a pre-intervention test consisting of twenty multiple-choice questions: Eleven non-visual spatial and nine visual-spatial cardiac anatomy questions.

In difference to the previously reported study, here the eye-tracking data was recorded during the VR learning phase. Because access to the software source code was not possible, 2D eye-tracking data of participants were collected with a corresponding VR-video from user's perspective. Participants stood at a defined position in space, looking into a defined direction in the VR scene to study the human heart.

A subjective questionnaire was filled out by the participants after the learning session to gain more insight into their experience and degree of satisfaction with using VR to study the heart. It was also done to obtain and assess the participant's thoughts on VR implementation in the study of the heart and assist in further analysis of the eye-tracking data derived from the study. This helped to analyze and evaluate students perceived learning and attention capabilities and if it had any influence on their performance.

It was possible to define specific areas of interest (AOI) the participant would focus on in the 2D screen of each VR display (left and right eye, see Figure 2). Since the users were able to move their heads freely, these regions had to be increased in 2D, to match the approximate 3D area of interest even if the users moved their heads a bit. This approach is valid, because the users had to look straight towards the 3D heart anyways in order to learn and understand it. Teleportation or walking was not allowed. We estimated that users would rotate their heads only approximately 20° horizontally or vertically while looking at the heart. This resulted in areas of interest that had a circular or ellipsoid shape. Semantically, the AOIs were placed around the following structures of the heart: AOI-1 is the Left Atrium, AOI-2, the Left Ventricle, and AOI-3, the Aortic Valve.

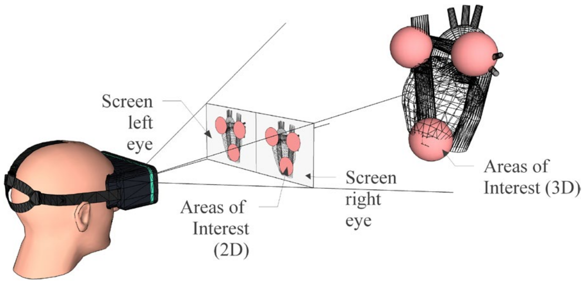

Figure 2 - Idea of AOI in 2D and 3D space. Only 2D can be used from eye-tracking data because 3D position of objects in the used software is not available.

A post-processed image extraction procedure from the recorded video was done after the teaching phase using Python[3]. A video overlay was created for each data point on the corresponding image frames to match the data points. Figure 3 shows an example.

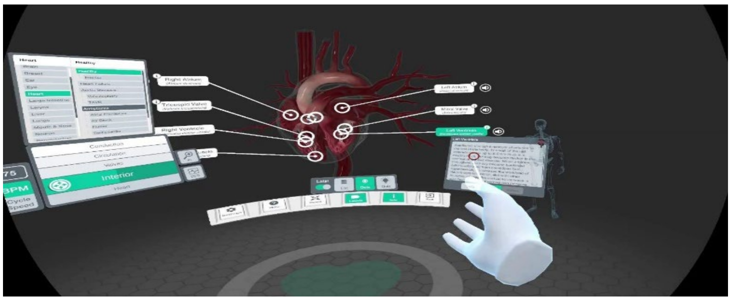

Figure 3 - Representation of Gaze point overlay. The gaze point is visible in the right textual area as a small red circle.

[3] The Python script files are available here: https://gitlab.com/xRLab_HSAnhalt.de/igbudueyetrackinganalysis

4 Results

According to the observation made from eye-tracking data, Table 1 represents the order in which AOIs and matrices were observed and recorded for analysis for each participant.

Table 2 - Showing exemplary recorded fixation data from AOIS of one participant

Metrics	AOI-1 (ms)	AOI-2 (ms)	AO1-3 (ms)
Time To First Fixation	88200	150000	243600
Dwell-Time	102000	49000	36000
Revisits	_	2000	4000
Revisit- dwell time	_	4000	8000

Evaluation of differences in mean average scores of the AOIs viewed and the statistical significance of the differences in the mean scores of the test before and after the learning session for the different participants were analyzed with (Wilcoxon signed-rank test). The Pearson Correlation coefficient (r) was also calculated. All analysis in this study were done with Python, SPSS Statistics 28.0 software, and Microsoft Excel.

An analysis from the demographic survey was carried out for the performance and knowledge assessment to check if prior cardiac knowledge influenced the participant's test scores. Twelve students (60%) reported having low cardiac knowledge, seven students (35%) had medium cardiac knowledge, and one student (5%) had high cardiac anatomy knowledge. This data was analyzed by comparing participants performance in specific questions of the test with their prior cardiac knowledge to see if this had any significant influence on their performance.

Fixation data showed a significant influence in the students post intervention test performance. In terms of the Pearson's correlation test, fixation (visual attention) is relative to participants' performance with a positive correlation of (r=0.68).

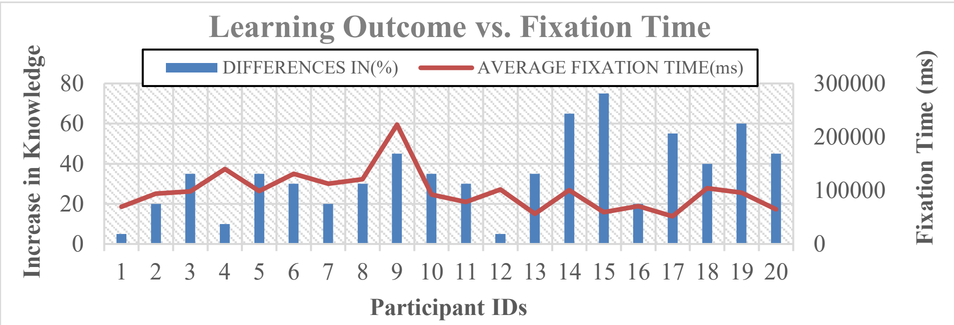

Figure 4 - Participants fixation data and post intervention scores

Figure 4 shows participant's average fixation time and post intervention increase in knowledge. It also shows how much time a person spent on the AOIs, and it demonstrates that regardless of how much time was spent, there is a positive result which reflected in the Quiz scores of each participant.

From the knowledge assessment, participants post test scores were compared to their prior cardiac knowledge to see if it had any influence in their performance in the test. As seen in Figure 4, a clear observation on participant (no.9) can be seen to have the longest fixation time in the chart as compared to all other participants. This participant had low prior cardiac knowledge. So, a longer fixation time during the study, is clearly an indication that he had focus and attention and obviously wanted to have a clear understanding of what was being taught. It resulted in knowledge increase and a positive performance in the post test of 85%. This pattern was noticed also in other participants, who scored between 15% and 35% in their pre-test and an improved score 90% to 100% post-test.

5 Discussion

From the analysis done in the previous study, based on test scores and performance of the participants from the experimental group, the average pre-intervention score was 47 %, with an average score increment of 34.75% to reach an average of 81.75% at the post-intervention scores ($Z = 3.927$, $p < 0.001$). This finding was already reported in the previous publication (see [Od22]) and indicates a statistically significant difference in quiz scores after the post-test.

The possibility of mapping AOIs on the 2D screen provided the fixation data of each participant and based on these AOIs, we were able to record and calculate the average fixation data on those AOIs of each participant. This approach best suited the scenario as it provided answers to our **RQ1** and **RQ2**: Recording eye-tracking data from an off-the-shelve VR software is possible without implementing software SDKs. The possibility of determining fixation of participants based on focus areas makes this approach valid.

However, data from the study shows students might have been hesitant about choosing the correct options during the quiz and might have chosen the right answer based on luck. Some may have relied on their prior cardiac knowledge and did not pay much attention to what was being taught; some could have been fatigued from the session and, as such, couldn’t spend much time answering the question correctly. Since the participants' behavior while answering the test questions was not tracked, we cannot predict their behavior in answering each question, so the eye fixation data becomes the only determinant to access them. Following this logic, a Pearson's correlation analysis presented an outcome that the forecast of students' performance can be done based on the correctness of their answers in co-relation to their cognitive level and based on their fixation duration respectively. Although, the 3D points of interest viewing still poses a challenge, this aspect of the study is open for more research.

As seen in Figure 4, participants fixation is relative to increase in knowledge gained after the quiz. The relationship ($r = 0.686$, $p < 0.05$) between the students responded post intervention scores and their fixation duration shows a strong positive correlation. This conclusively, established and answered **RQ3**: Is it possible to forecast the user's performance from the eye-tracking data. Participants gaze data proved their attention on the specific AOIs marked out; their performance in the test did not occur by random chance and visual attention did play a vital role in their performance as there was a significant increase in knowledge and performance of the participants due to their fixation on the specified AOIs. The results also demonstrated that regardless of the participant's fixation duration, there was still a positive significance in the participant's performance due to their visual attention.

This study shows, that the assessment of students' learning performance using off-the-shelve Virtual Reality software is possible. This can help lecturers better understand how students learn and how to design better teaching scenarios – even if no access to specific source code of the used VR software is available.

6 Outlook

The approach reported here leaves room for significant improvements to alleviate some of the issues of analyzing eye-tracking in virtual space. In this specific situation, where users have to look into a defined direction, without being able to walk or move freely, the idea of AOIs might be sufficient. However, when it comes to VR scenes with more freedom of movement this approach will fail. Therefore, we plan to develop two new methods to use VR eye-tracking with off-the-shelve VR software:

1. Using AI methods to automatically detect semantic features within a VR scene and then re-detect those from any position within the 3D space.
2. Use a 3D-3D point set registration between derived 3D-eye-tracking points, live positional tracking data of the VR headset and the gazed-at point within the 3D scene.

It is important to follow this path to allow future human factors and teaching studies to be carried out even if the source code of VR software is not available.

Conflict of Interest Statement

The software and hardware used in this study were purchased at the regular price. The authors have no connection to Varjo or Sharecare, Inc. The results presented here are based on the sole actions of the authors and are in no way related to Sharecare, Inc or Varjo.

References

[Ba13] Bamodu, Oluleke, and Xu Ming Ye. "Virtual reality and virtual reality system components." *Advanced materials research*. Vol. 765. Trans Tech Publications Ltd, 2013.

[Ch13] Chennamma, H. R., and Xiaohui Yuan. "A survey on eye-gaze tracking techniques." *arXiv preprint arXiv:1312.6410,* 2013. [Online]. Available: http://www.tobii.se

[Ho17] Holmqvist, Kenneth, and R. Andersson. "Eye-tracking: A comprehensive guide to methods." *paradigms and measures*, 2017.

[Ku16] Kurzhals, K., Fisher, B., Burch, M., & Weiskopf, D. "Eye-tracking evaluation of visual analytics." *Information Visualization* 15.4 (2016): 340-358.

[Mc03] H. Mclellan, "VIRTUAL REALITIES," 2003. Accessed: Feb. 08, 2022. [Online]. Available: http://www.cyberedge.com/4f.html.

[Mi22] Mikhailenko, Maria, Nadezhda Maksimenko, and Mikhail Kurushkin. "Eye-tracking in immersive virtual reality for education: a review of the current progress and applications." *Frontiers in Education*. Vol. 7. Frontiers Media SA, 2022.

[Od22] Odame, Angela, Tümler, Johannes: "Is Off-the-Shelf VR Software Ready for Medical Teaching?" *Proceedings of the HCII 2022 Virtual, Augmented and Mixed Reality: Design and Development*, part of the Lecture Notes in Computer Science book series (LNCS, volume 13317), 2022.

[Pl18] Płużyczka, Monika. "The first hundred years: A history of eye-tracking as a research method." Applied Linguistics Papers 25/4 (2018): 101-116.

[Sh22] "Sharecare YOU." https://www.sharecareyou.com/ (accessed Feb. 12, 2022).

[Tr18] Andrychowicz-Trojanowska, Agnieszka. "Basic terminology of eye-tracking research." *Applied Linguistics Papers* 25/2 (2018): 123-132.

[Va22] "Varjo VR-2: Full Specification – VR compare." https://vr-compare.com/headset/varjovr-2 (accessed Feb. 16, 2022).