

GIUPlayer: A Gaze Immersive YouTube Player Enabling Eye Control and Attention Analysis

Ramin Hedeshy

University of Koblenz, Germany
hedeshy@uni-koblenz.de

Raphael Menges

University of Koblenz, Germany
raphaelmenges@uni-koblenz.de

Chandan Kumar

University of Koblenz, Germany
kumar@uni-koblenz.de

Steffen Staab*

University of Stuttgart, Germany
steffen.staab@ipvs.uni-stuttgart.de

ABSTRACT

We developed a gaze immersive YouTube player, called GIUPlayer, with two objectives: First to enable eye-controlled interaction with video content, to support people with motor disabilities. Second to enable the prospect of quantifying attention when users view video content, which can be used to estimate natural viewing behaviour. In this paper, we illustrate the functionality and design of GIUPlayer, and the visualization of video viewing pattern. The long-term perspective of this work could lead to the realization of eye control and attention based recommendations in online video platforms and smart TV applications that record eye tracking data.

CCS CONCEPTS

• **Human-centered computing** → **Interactive systems and tools.**

KEYWORDS

Video player, eye tracking, Web accessibility

ACM Reference Format:

Ramin Hedeshy, Chandan Kumar, Raphael Menges, and Steffen Staab. 2020. GIUPlayer: A Gaze Immersive YouTube Player Enabling Eye Control and Attention Analysis. In *Symposium on Eye Tracking Research and Applications (ETRA '20 Adjunct)*, June 2–5, 2020, Stuttgart, Germany. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/3379157.3391984>

1 INTRODUCTION

Watching entertainment programs and movies are one of the most common and time-consuming activities of our everyday life. In recent years, YouTube and other online services have gained immense popularity and preference over conventional TV broadcasts. As per 2019 statistics, over 500 hours of video are uploaded to YouTube every minute, and more than 5 billion videos are watched on YouTube every single day [Statista, Inc. 2020]. Given the significance of these online platforms in our modern digital world, it is imperative that these platforms are accessible to all people. However, several individuals with motor-impairment, who lost the ability to operate

*Also with University of Southampton, UK

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

ETRA '20 Adjunct, June 2–5, 2020, Stuttgart, Germany

© 2020 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-7135-3/20/06.

<https://doi.org/10.1145/3379157.3391984>

mouse and keyboard, are not able to interact with computing devices that provide digital content as required by standard interfaces. More specifically, this excludes those individuals from entertainment and educational content on online video platforms that could enrich their social environment and general quality of life.

In this regard, eye tracking as a novel interaction techniques enables hands-free interaction using eye gaze as input control to support people with severe motor disabilities [Jacob and Stellmach 2016; Kumar et al. 2016]. To access the online video platforms, the generic method for eye gaze control at operating system level (emulation approaches [Lankford 2000; Sweetland 2016]) could be employed. However, emulation approaches do not have access on the underlying data structures and interaction possibilities for most applications or websites [Menges et al. 2019]. Thus, a more appropriate solution is to design special graphical user interface adapted for eye gaze input. There have been several approaches to adapt the native application for eye gaze, e. g., eye-controlled gaming [Isokoski et al. 2009; Menges et al. 2017b], drawing [van der Kamp and Sundstedt 2011], browsing [Menges et al. 2017a]. More specifically, in the context of video interaction, the control over the watched content, including pausing, resuming, forwarding, and control over the volume is crucial for the comfort and pleasure of the user. In this work, we demonstrate GIUPlayer, which retrieves YouTube data and adapts interface elements for eye gaze control to provide intuitive video interaction functionality.

Besides providing input control, another major application of eye tracking has been in the field of attention analysis, i.e., to understand user behavior. In the context of video interaction, eye gaze attention can provide implicit understanding of user interest while users are watching the videos. The implicit feedback has been investigated in different analytical scenarios such as to judge video quality [Sawahata et al. 2008], to summarize and compress videos [Gitman et al. 2014; Katti et al. 2011], characterize users [Tanisaro et al. 2015], or to assess relevance of advertisement banners [Tangmanee 2016]. Therefore to support such analysis, in GIUPlayer we record and visualize the gaze data while video is being played.

2 THE GIU PLAYER INTERFACE

GIUPlayer content layout is inspired by the YouTube website since its functionality is well known, thus intuitive for end users. However, several interface adaptations are done following user centered design principles for eyes interaction [Kumar et al. 2016; Menges et al. 2019]. Interface components such as size, shape, appearance, feedback etc., are customized since it is crucial in enhancing eye

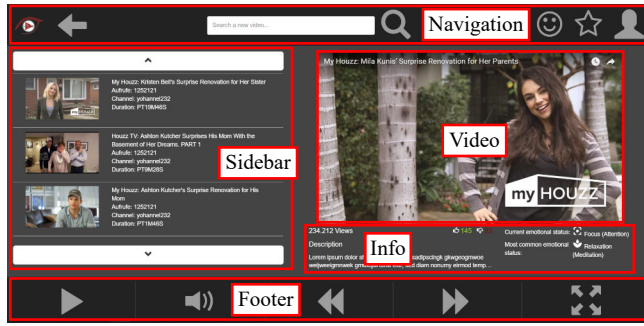


Figure 1: Areas in GIUPlayer watch page interface

tracking accuracy for input control. For example, sizes of buttons are enlarged, their placement is based upon element usage frequency and application control, so natural viewing behavior does not interfere. Several alternatives for eye gaze input like blinking, gesture, pupil dilation [Jacob and Stellmach 2016] had been explored, however dwell time based selection was found most suitable, as it was the most natural and widely accepted mode of input. Visual feedback is embedded to guide the selection process.

GIUPlayer consists of three content related pages including the watch page, a home page, and a search result page.

Figure 1 shows the watch page of GIUPlayer. Every time when a user selects a video this page appears and the video can be played. The place for displaying the video takes the most considerable part of the page on the right side, with controls below (marked as footer in Figure 1) and suggested videos on the left side (marked as sidebar). The suggestions are obtained by using the YouTube API. Dwelling at one of these suggestions (on the image thumbnail part) causes the played video to change. The text description part is not clickable so the user can comfortably read without unintentional selections. Looking at the buttons with arrows above and below suggestions causes changing of the set of proposed videos. The footer area serves the control function of the video. If user activates the fullscreen mode the video size will maximize and the footer is by default visible to provide the control over the video. However in fullscreen mode, a small gray rectangle with an arrow inside appears on the right side over the footer. By looking at this rectangle, a user can hide the bottom navigation bar to have a complete fullscreen view without further distractions by the control buttons.

The home page includes the tiled list of videos with each row representing different categories such as top rated, last watched, and recommended videos. Similarly to the suggestions on the watch page, the tiled lists can be scrolled with corresponding buttons. The search result page contains the list of videos with respect to search query. Additionally, the user settings and profile information can be accessed from the button on the top right of the page.

The entire interface was developed as a Web application using JavaScript, CSS, Ajax, and Bootstrap. The Youtube API [Google LLC 2020] is used to access video information and other meta-data. MySQL database is used to store user credentials and gaze data stream of fixations on the videos. As we implemented GIUPlayer as Web application, an eye-controlled Web browser like GazeTheWeb [Menges et al. 2017a] can be used to display it. Then, the integrated dwell-time-based keyboard of GazeTheWeb is used



Figure 2: Displaying fixations of five users on a video

to enter search queries. Moreover, the system works with a variety of eye tracking peripherals from SMI, Tobii, and Visual Interaction.

3 ATTENTION ANALYSIS

Whenever a user plays any video, the recording of the gaze data starts automatically. The gaze data is processed on-the-fly to generate a list of fixation-points. After each video, these fixation-points are stored in a database in relation to the user id and the video id.

The analysis mode allows the visual comparison of the fixation points. A user can activate the analysis mode, in order to visualize the individual fixation points. To analyze and compare the viewing behaviour, user can choose to display the data of the last five persons with their different fixation points for each video. A unique color represents the fixation per user. See Figure 2 for a screenshot from the analysis mode. The visualization is realized by multiple canvases on top of the video. For each user, a new layer with a canvas is added. The fixations are visualized as points around the central point of the fixation with a 50 px radius. Fixations were defined with minimum time duration of 200 ms [Blaschek et al. 2014; Nyström and Holmqvist 2010]. A fixation has a spatial position (x- and y-coordinate) and a start- and end-time in milliseconds. With these timestamps, the duration of the fixation is calculated and displayed accordingly. We have also implemented the function to calculate the similarity between gaze data stream of different users on a particular video, i. e., to showcase most similar users based on viewing pattern. The similarity function incorporates common fixations (temporal and spatial overlap); common amount of fixations and common duration of fixations.

4 CONCLUSION AND FUTURE WORK

The proposed GIUPlayer provides intuitive hands-free interaction with YouTube platform. We have performed in-lab usability testing of the interface, and plan to conduct an evaluation with end-users.

For attention analysis, currently we only provide fixation visualization and user similarity function, however, the system offers numerous possibility to exploit attention data for interaction optimization and personalized user experience. For example, the user recommendations and filtering on current online platforms are merely based on the explicit feedback from user, e. g., what users have searched and clicked. A platform with eye tracking can build a network of users providing eye gaze data. The implicit feedback can be used to characterize user interests for personalized ranking, and collaborative filtering based on the similar gaze-pattern of users.

ACKNOWLEDGMENTS

We acknowledge the work by students from University of Koblenz: Denise Dünnebier, Mariya Chkalova, Min Ke, Yessika Legat, Arsenii Smyrnov, Benjamin Becker, Daniyal Akbari, Matthias Greber, Steven Schürstedt, and Jannis Eisenmenger, who were involved in a project¹ that contributes towards results reported in this paper.

REFERENCES

- Tanja Blascheck, Kuno Kurzhals, Michael Raschke, Michael Burch, Daniel Weiskopf, and Thomas Ertl. 2014. State-of-the-Art of Visualization for Eye Tracking Data.. In *EuroVis (STARs)*.
- Yury Gitman, Mikhail Erofeev, Dmitriy Vatolin, Bolshakov Andrey, and Fedorov Alexey. 2014. Semiautomatic visual-attention modeling and its application to video compression. In *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE, 1105–1109.
- Google LLC. 2020 (accessed March 5, 2020). *YouTube API*. <https://developers.google.com/youtube>
- Poika Isokoski, Markus Joos, Oleg Spakov, and Benoît Martin. 2009. Gaze Controlled Games. *Univ. Access Inf. Soc.* 8, 4 (Oct. 2009), 323–337. <https://doi.org/10.1007/s10209-009-0146-3>
- Rob Jacob and Sophie Stellmach. 2016. What You Look at is What You Get: Gaze-based User Interfaces. *interactions* 23, 5 (Aug. 2016), 62–65. <https://doi.org/10.1145/2978577>
- Harish Katti, Karthik Yadati, Mohan Kankanhalli, and Chua Tat-Seng. 2011. Affective video summarization and story board generation using pupillary dilation and eye gaze. In *2011 IEEE International Symposium on Multimedia*. IEEE, 319–326.
- Chandan Kumar, Raphael Menges, and Steffen Staab. 2016. Eye-Controlled Interfaces for Multimedia Interaction. *IEEE MultiMedia* 23, 4 (Oct 2016), 6–13.
- Chris Lankford. 2000. Effective Eye-gaze Input into Windows. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications (ETRA '00)*. ACM, New York, NY, USA, 23–27. <https://doi.org/10.1145/355017.355021>
- Raphael Menges, Chandan Kumar, Daniel Müller, and Korok Sengupta. 2017a. GazeTheWeb: A Gaze-Controlled Web Browser. In *Proceedings of the 14th Web for All Conference (W4A '17)*. ACM. <http://dx.doi.org/10.1145/3058555.3058582>
- Raphael Menges, Chandan Kumar, and Steffen Staab. 2019. Improving User Experience of Eye Tracking-based Interaction: Introspecting and Adapting Interfaces. *ACM Trans. Comput.-Hum. Interact.* (2019). Accepted May 2019.
- Raphael Menges, Chandan Kumar, Ulrich Wechselberger, Christoph Schaefer, Tina Walber, and Steffen Staab. 2017b. Schau genau! A Gaze-Controlled 3D Game for Entertainment and Education. In *Journal of Eye Movement Research*, Vol. 10. 220. <https://bop.unibe.ch/JEMR/article/view/4182>
- Marcus Nyström and Kenneth Holmqvist. 2010. An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behavior research methods* 42, 1 (2010), 188–204.
- Yasuhiro Sawahata, Rajiv Khosla, Kazuteru Komine, Nobuyuki Hiruma, Takayuki Itou, Seiji Watanabe, Yuji Suzuki, Yumiko Hara, and Nobuo Issiki. 2008. Determining comprehension and quality of TV programs using eye-gaze tracking. *Pattern Recognition* 41, 5 (2008), 1610–1626.
- Statista, Inc. 2020 (accessed March 5, 2020). *STATISTA*. <https://www.statista.com/topics/2019/youtube>
- Julius Sweetland. 2016. Optikey: Type, Click, Speak. <https://github.com/OptiKey/OptiKey>.
- Chatpong Tangmanee. 2016. Fixation and recall of YouTube ad banners: An eye-tracking study. " *International Journal of Electronic Commerce Studies* " 7, 1 (2016), 49–76.
- Pattreeya Tanisaro, Julius Schöning, Kuno Kurzhals, Gunther Heidemann, and Daniel Weiskopf. 2015. Visual analytics for video applications. *it-Information Technology* 57, 1 (2015), 30–36.
- Jan van der Kamp and Veronica Sundstedt. 2011. Gaze and Voice Controlled Drawing. In *Proceedings of the 1st Conference on Novel Gaze-Controlled Applications (NGCA '11)*. ACM, New York, NY, USA, Article 9, 8 pages. <https://doi.org/10.1145/1983302.1983311>

¹<https://github.com/MAMEM/GazeTheWeb/tree/master/Watch>