TrackThink Camera: A Tool for Tracking Facial and Body Information while Web Browsing

Ko Watanabe[1,2], Seiya Tanaka[3], Andrew Vargo[3], Koichi Kise[3] and Shoya Ishimaru[3]

[1]RPTU Kaiserslautern-Landau, Pfaffenbergstrasse 95, 67663 Kaiserslautern, Germany
[2]DFKI GmbH, Tripstadter 122, 67663 Kaiserslautern, Germany
[3]Osaka Metropolitan University, 1-1 Gakuen-cho, Naka, Sakai, Osaka 599-8531, Japan

Abstract
In a world where the amount of information is increasing daily, estimating cognitive and affective states in web search activity is essential for self-management. Previous studies have investigated methods for collecting logs of web search activity, such as URLs of web pages visited, tab activation, and clipboard copying. However, none of the work focuses on external information, such as collecting the person's facial expressions or body movements while searching. In this paper, we extend the existing system to enable synchronous collection of web search logs and web camera recordings. Through this research and development, we aim to analyze the cognitive and affective states during search behavior and help improve the efficiency of web search. Our new software will be available to anyone with research purposes.

Keywords
thought process, browsing behavior, web search, cognitive metrics, camera as a smart sensor

1. Introduction

Web search is an essential part of our daily lives. It supports our productivity, creativity, recreation, and even socialization. Previous studies have proposed tools for collecting or visualizing web search logs [1, 2, 3]. These approaches successfully understand users' search activity within a computer or web browser. A study by Aula et al. found that most people, novices and experts alike, produce a certain body language when they get stuck in a search [4]. Their research then turned to detecting frustration, or so-called behavioral changes, from the activity inside the computer. However, we can better understand their cognitive/affective states by synchronously capturing their frustration directly from webcam recordings. Therefore, the discovery of cognitive/affective states from web camera browsing is fascinating.

Estimating cognitive/affective states is a crucial variable affecting human performance in various tasks, including puzzle solving, scuba diving, public speaking, education, fighter aircraft operation, and driving [5]. Understanding the cognitive/affective state during exploration helps

AIED 2023 Workshop: Automated Assessment and Guidance of Project Work, July 03–07, 2023, Tokyo, Japan
ko.watanabe@dfki.de (K. Watanabe); sb22873y@st.omu.ac.jp (S. Tanaka); awv@omu.ac.jp (A. Vargo); kise@omu.ac.jp (K. Kise); ishimaru@omu.ac.jp (S. Ishimaru)
https://ko-watanabe.github.io/ (K. Watanabe); https://www.kise.info/ (K. Kise); https://shoya.io/ (S. Ishimaru)
0000-0003-0252-1785 (K. Watanabe); 0009-0005-7077-2756 (S. Tanaka); 0000-0001-6605-0113 (A. Vargo); 0000-0001-5779-6968 (K. Kise); 0000-0002-5374-1510 (S. Ishimaru)

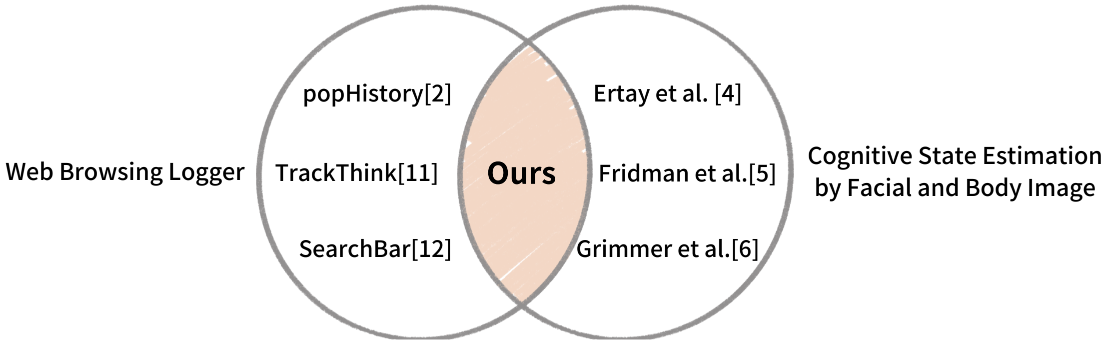

Figure 1: Position of our proposed system.

us understand how efficient our performance was in finding solutions. The concept of estimating cognitive load in the wild, or Automatic Emotion Recognition (AER) technologies [6], needs to gain attention in research and industry. There is work on estimating the affective state while using a smartphone in the wild [7]. This system lets users get feedback on affective states while using the phone. One of the core techniques for understanding cognitive/affective states is using facial or body information [8, 9, 10].

Determining cognitive states while reading digital textbooks has been done by several researchers [11, 12]. The approaches mainly use an eye tracker or a heat sensor to estimate cognitive states. Previous research has found that body temperature and blink frequency best estimate engagement and can classify users independently as engaged or disengaged [11]. Researchers have also found that pupil diameter and nasal temperature changes correlate with cognitive state [12]. These studies contribute to the understanding of cognitive states during reading in particular. However, all of these researches use additional sensors to detect cognitive states. In addition, these studies focused on something other than reading activity during a web search.

This paper proposes a system that synchronously uses a web camera to collect web search behavior logs and camera-recorded facial and body information. At the same time, a user performs a web search. Figure 1 shows the intersection of previous work in the Venn diagram. Our position is to implement an all-in-one technology for tracking web search behavior and synchronizing web camera recordings. The system extends the web search logger TrackThinkTS [3]. This system works with the Google Chrome extension. It does not require any additional sensors or hardware devices. Our main contribution is that we can obtain cognitive/affective information from the user's facial and bodily behaviors without sacrificing this feature. Instead of installing additional sensors or a system to log the video information, anyone can use the system just by installing the Google Chrome extension in the web browser. To the best of our knowledge, we are the first to research and develop an all-in-one system to collect web searches and facial and body behaviors.

2. Related Work

In this section, we focus on three key areas of literature: (1) prior work on web search loggers, (2) prior work on cognitive/affective state estimation by facial and body information, and (3) prior work on cognitive/affective state estimation during a web search. Understanding the related work highlights the significance of discovering facial and body movements while web browsing.

2.1. Web Search Logger

Web search logger is a topic that has been gaining interest in Human-Computer Interaction (HCI), with most studies focusing on discovering the thought process. Morris et al. has proposed a search-centric web history logger *SearchBar* [2]. *SearchBar* is a system implemented as a plugin for Internet Explorer. The system can retrieve, store, and present an annotated integrated search history. Over time, the search history collection will be extended to the mobile scene. Kamvar and Baluja worked in the field of mobile [13]. They used the XHTML search interface to collect and set the experimental condition to analyze the web search behavior. Nagano et al. proposed the concept of TrackThink, a system for tracking human thought processes from web search logs [14]. This system is then extended by Makhlouf et al. and becomes publicly available for anyone to use as a Google Chrome extension [3]. Carrasco et al. has proposed a popHistory system that collects web search history and visualizes it in a bubble interface [1]. It can segment the result into a specific time range. In conclusion, according to our survey, none of the previous web search behavior loggers used a web camera to record data outside the computer.

2.2. Cognitive/Affective State Estimation from Facial and Body Information

Various approaches have been used for cognitive/affective state estimation. This section focuses on camera-based or facial image-based cognitive state estimation. Kunze et al. works on estimating the level of engagement during reading using a Tobii eye tracker and temperature sensors attached to the nose and ear. They found that temperature and blink frequency are the best ways to estimate the engagement and can classify engaged and disengaged users independently. Fridman et al. proposed a system to estimate cognitive load in the wild by a camera while driving a car [5]. Their approach is to extract discriminative signals in the eye movement dynamics to detect the level of cognitive load. Using the 3D-CNN approach, they achieved 86.1% on 3-class cognitive load estimation. Grimmer et al. also used pupil information to estimate cognitive load. They found that pupil diameter decreased over time during the control condition and the 1-back task. However, they also mentioned that the differences in pupil diameter remained the same between the 2- and 3-back tasks. Watanabe et al. proposed engagement level estimation in remote communication [9]. With respect to these works, facial and body information play an important role in understanding human cognitive and affective states.

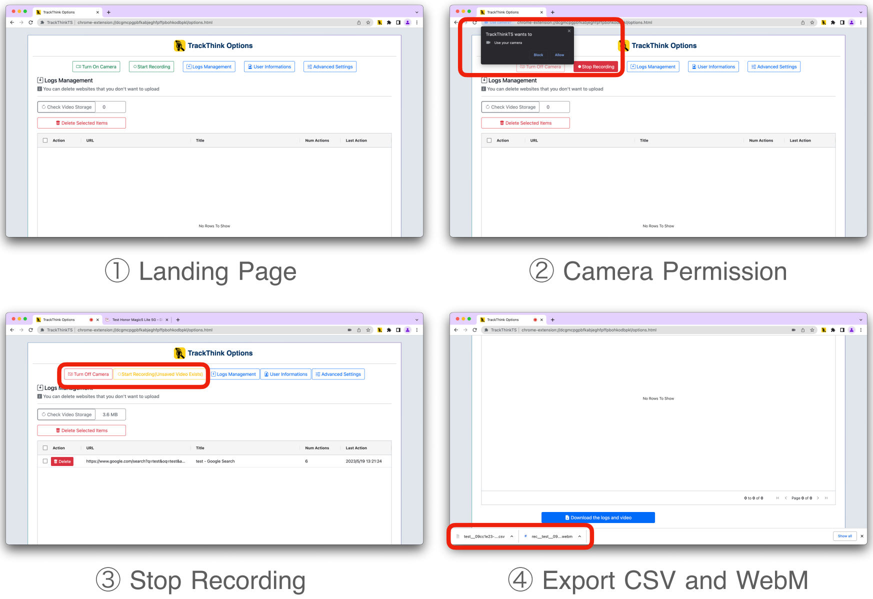

Figure 2: System Workflow. All-in-one collect web search, facial, and body behavior logs. The system work as a Chrome extension. No additional sensors or devices are required.

2.3. Cognitive/Affective State Estimation during Web Search

Whether novice or expert, searchers need help finding the information they want. Aula et al. have discovered that when searchers are frustrated, there are observable changes in their behavior [4]. Specifically, their body language changes (e.g., they frown, start biting their nails, and many lean closer to the screen to ensure they are not missing something obvious), and they begin to sigh. In read-aloud studies, they tend to forget to read aloud. These actions can trigger an estimation of which website they have been frustrated or struggled with during web searches.

3. System Overview

This study will use TrackThinkTS [3] as a baseline system. It collects various information about web search behavior. In this section, we will explain the baseline system in detail. Then we explain two new specific functions: a recording segmentation function and a web camera recording function. The new system TrackThink Camera operation flow is presented in Figure 2.

3.1. TrackThinkTS: Baseline System

Makhlouf et al. proposed TrackThinkTS as a privacy-aware browsing log tracker. It monitors the following browsing actions and collects logs [3]. First, it collects information about the visited website (a website title, URL, HTML content, viewport width/height, and document width/height). Each time a tab-related operation occurs, such as creating a tab, launching a tab, reloading a tab, or deleting a tab, the system keeps a log. Second, specific user actions while browsing the website are collected, such as scrolling logs (scrolling speed, scrolling length, and visible text after scrolling is finished) and clipboard logs (the clipboard contents). Compared to typical browsing history, it is possible to collect information such as which parts of the page the user looked at in detail and how long they stayed on each page. Another advantage of TrackThinkTS over other browser loggers is its user-friendly interface for filtering each log. Therefore, study participants using TrackThinkTS can delete privacy-sensitive information before submitting their log files to an experimenter.

3.2. Recording Segmentation Function

The traditional TrackThinkTS system collects web activity logs after the extension is installed. Logs are collected continuously until the CSV file is downloaded. Therefore, the existing system records web activity immediately after installation by the user. To avoid privacy issues from collecting all logs, TrackThinkTS can delete logs before exporting data. In this work, we implement the start and stop buttons to handle the beginning and end of data collection. Therefore, we added a start/stop recording feature to allow users to record web activity data selectively. More precisely, we placed a recording button on the Settings tab of the extension to manage the start and stop.

3.3. Web Camera Recording Function

This section explains the Web camera recording function. Figure 2 shows an overview of the system. Each person follows the operation as shown below.

1. **Start Recording** - Click *Turn on Camera* and *Start Recording* buttons for logging.
2. **Allow Recording** - Click *Allow* button for giving permission for camera recordings.
3. **Stop Recording** - Click *Stop Recording* button to stop logging.
4. **Download Logs** - Click *Download the logs and video* button to download files.
5. **Restart Recording** - Click *Start Recording* and logs will be refreshed once.

The new TrackThink Camera allows users to record facial and body recordings using the web camera. Users can also check storage the storage of video recordings while logging. The video can be downloaded locally in WebM format, and a web search logs in CSV. The WebM file name will be generated based on the user ID, name, and recording end time.

4. Discussion & Limitations

This paper proposes a new system that synchronously collects search behavior logs and facial and body recordings. The system can now control the start and end of data collection and store video recordings. When we conducted a pilot study, we discovered that participants looked closer at the screen when they found the page difficult to understand, as mentioned in the previous paper [4]. For the next step, some limitations need to be covered.

First, we did not perform a validation experiment with multiple users. We confirmed that the system works well in collecting web search behavior and video recording. However, we did not collect data on cognitive activity during browsing. Experimental design will also be a topic of discussion when conducting a validation experiment. It is necessary to determine how to design tasks with different cognitive loads during the search. Discussing the collection of multiple cognitive states in the search activity would be desirable.

Second, in this system, the challenges observed from our prototype implementations are the highlighting function. In order to select facial and body images while viewing each web page, it is necessary to refer to the time stamps collected in the search logs. Specifically, it is necessary to determine the starting position of the video when viewing a particular page using the timestamp at the start of data collection and the timestamp at the time the web page is viewed. Creating highlights when downloading and adding them to the output video would be desirable.

Lastly, regarding the earlier question, another challenge is to cut out sensitive scenes from the recordings. There are functions for deleting the search log data, but the video recordings are not trimmed before downloading. When implementing the highlight functions, it is necessary to make the video trimming function allow deleting the facial and body shots while viewing the personal website.

5. Future Work

In our future work, we want to implement some systems using the TrackThink Camera. One of the ideas is to use appearance-based eye-tracking techniques. We are interested in adding this technology to identify what part of the web page the person was looking at within the web content. This approach will further support understanding the cognitive status of where specifically the user was looking within the web page. Another idea is to quantify the cognitive load of the web page. By doing so, people can choose to read the web page according to the enumerated cognitive load information of the web page. Considering GDPR (General Data Protection Regulation), our future work includes secureness on privacy issues. We will implement additional options for exporting data, such as allowing to export of cognitive load estimation results only. Finally, we would also like to create a dashboard for web searchers to retrieve cognitive lifelog data. This way, users can understand which web pages have a high cognitive load. These features will significantly benefit all web search users in the future.

6. Conclusion

This paper proposes a Google Chrome extension tool, TrackThink Camera, to collect web search activity and facial/body behavior logs. This system supports collecting web search activity and facial/body video recording. Our proposal is the first to collect internal and external information about people while working on web search activity. The new software will be available to anyone with research purposes.

Acknowledgments

This work was supported in part by JST Trilateral AI Research (JPMJCR20G3).

References

[1] M. Carrasco, E. Koh, S. Malik, Pophistory: Animated visualization of personal web browsing history, in: Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA '17, Association for Computing Machinery, New York, NY, USA, 2017, p. 2429–2436. URL: https://doi.org/10.1145/3027063.3053259. doi:10.1145/3027063.3053259.

[2] D. Morris, M. Ringel Morris, G. Venolia, Searchbar: A search-centric web history for task resumption and information re-finding, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '08, Association for Computing Machinery, New York, NY, USA, 2008, p. 1207–1216. URL: https://doi.org/10.1145/1357054.1357242. doi:10.1145/1357054.1357242.

[3] J. Makhlouf, Y. Arakawa, K. Watanabe, A privacy-aware browser extension to track user search behavior for programming course supplement, in: International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services, Springer, 2021, pp. 783–796.

[4] A. Aula, R. M. Khan, Z. Guan, How does search behavior change as search becomes more difficult?, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '10, Association for Computing Machinery, New York, NY, USA, 2010, p. 35–44. URL: https://doi.org/10.1145/1753326.1753333. doi:10.1145/1753326.1753333.

[5] L. Fridman, B. Reimer, B. Mehler, W. T. Freeman, Cognitive load estimation in the wild, in: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18, Association for Computing Machinery, New York, NY, USA, 2018, p. 1–9. URL: https://doi.org/10.1145/3173574.3174226. doi:10.1145/3173574.3174226.

[6] H. Kaur, D. McDuff, A. C. Williams, J. Teevan, S. T. Iqbal, "i didn't know i looked angry": Characterizing observed emotion and reported affect at work, in: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, CHI '22, Association for Computing Machinery, New York, NY, USA, 2022. URL: https://doi.org/10.1145/3491102.3517453. doi:10.1145/3491102.3517453.

[7] R. Wampfler, S. Klingler, B. Solenthaler, V. R. Schinazi, M. Gross, C. Holz, Affective state prediction from smartphone touch and sensor data in the wild, in: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, CHI '22, Association for Computing Machinery, New York, NY, USA, 2022. URL: https://doi.org/10.1145/3491102.3501835. doi:10.1145/3491102.3501835.

[8] J. F. Cohn, F. De la Torre, Automated face analysis for affective computing. (2015).

[9] K. Watanabe, T. Sathyanarayana, A. Dengel, S. Ishimaru, Engauge: Engagement gauge of meeting participants estimated by facial expression and deep neural network, IEEE Access (2023).

[10] J. Grimmer, L. Simon, J. Ehlers, The cognitive eye: Indexing oculomotor functions for mental workload assessment in cognition-aware systems, in: Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems, CHI EA '21, Association for Computing Machinery, New York, NY, USA, 2021. URL: https://doi.org/10.1145/3411763.3451662. doi:10.1145/3411763.3451662.

[11] K. Kunze, S. Sanchez, T. Dingler, O. Augereau, K. Kise, M. Inami, T. Tsutomu, The

augmented narrative: Toward estimating reader engagement, in: Proceedings of the 6th Augmented Human International Conference, AH '15, Association for Computing Machinery, New York, NY, USA, 2015, p. 163–164. URL: https://doi.org/10.1145/2735711.2735814. doi:10.1145/2735711.2735814.

[12] S. Ishimaru, S. Jacob, A. Roy, S. S. Bukhari, C. Heisel, N. Großmann, M. Thees, J. Kuhn, A. Dengel, Cognitive state measurement on learning materials by utilizing eye tracker and thermal camera, in: 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), volume 08, 2017, pp. 32–36. doi:10.1109/ICDAR.2017.378.

[13] M. Kamvar, S. Baluja, A large scale study of wireless search behavior: Google mobile search, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '06, Association for Computing Machinery, New York, NY, USA, 2006, p. 701–709. URL: https://doi.org/10.1145/1124772.1124877. doi:10.1145/1124772.1124877.

[14] K. Nagano, Y. Arakawa, K. Yasumoto, Trackthink: A tool for tracking a thought process on web search, in: Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers, UbiComp '17, Association for Computing Machinery, New York, NY, USA, 2017, p. 681–687. URL: https://doi.org/10.1145/3123024.3129267. doi:10.1145/3123024.3129267.