Old Dominion University
ODU Digital Commons

Computer Science Faculty Publications | Computer Science

2022

Multi-User Eye-Tracking

Bhanuka Mahanama
Old Dominion University

Follow this and additional works at: https://digitalcommons.odu.edu/computerscience_fac_pubs

Part of the Artificial Intelligence and Robotics Commons, and the Graphics and Human Computer Interfaces Commons

Original Publication Citation

Mahanama, B. (2022). Multi-user eye-tracking. In F. Shic, E. Kasneci, & et al. (Eds.), *ETRA '22: 2022 Symposium on Eye Tracking Research and Applications.* (Article 36) Association for Computing Machinery. https://doi.org/10.1145/3517031.3532197

This Conference Paper is brought to you for free and open access by the Computer Science at ODU Digital Commons. It has been accepted for inclusion in Computer Science Faculty Publications by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.

Multi-User Eye-Tracking

Bhanuka Mahanama
bhanuka@cs.odu.edu
Old Dominion University
Norfolk, VA, USA

ABSTRACT

The human gaze characteristics provide informative cues on human behavior during various activities. Using traditional eye trackers, assessing gaze characteristics in the wild requires a dedicated device per participant and therefore is not feasible for large-scale experiments. In this study, we propose a commodity hardware-based multi-user eye-tracking system. We leverage the recent advancements in Deep Neural Networks and large-scale datasets for implementing our system. Our preliminary studies provide promising results for multi-user eye-tracking on commodity hardware, providing a cost-effective solution for large-scale studies.

CCS CONCEPTS

• **Human-centered computing** → **Interaction techniques**; **Collaborative interaction**; • **Computing methodologies** → *Neural networks*.

KEYWORDS

Eye Tracking, Multi-user, Gaze Tracking, Deep Learning

ACM Reference Format:
Bhanuka Mahanama. 2022. Multi-User Eye-Tracking. In *2022 Symposium on Eye Tracking Research and Applications (ETRA '22), June 8–11, 2022, Seattle, WA, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3517031.3532197

1 INTRODUCTION

The human gaze has a wide range of applications, from human-computer interactions [Mutlu et al. 2009; Palinko et al. 2016; Papoutsaki et al. 2017] to research in psychology[De Silva et al. 2019; Michalek et al. 2019] and behavioral studies [Asteriadis et al. 2009; Mahanama et al. 2021]. Recently, gaze characteristics have been the subject of experiments, and based on the requirements, eye-trackers have evolved into two forms head-mounted and desktop-based[Pathirana et al. 2022]. Even though these eye-trackers perform well for single-user studies, they lack the scalability for multi-user studies mainly because they cannot track more than one person. Further, a multi-user tracking setup will require a dedicated device per participant [Zhang et al. 2017a], making simple experiments costly.

Eye-tracking using commodity hardware provides an alternative for overcoming the scalability issues and has been subjected to many studies. Due to the poor feature extraction techniques [Pathirana et al. 2022], these approaches have exhibited relatively lesser accuracy to be used for experimental setups. However, recent developments such as deep neural network-based techniques [Fischer et al. 2018; Kellnhofer et al. 2019; Krafka et al. 2016; Zhang et al. 2020, 2015, 2017b,c] and large-scale datasets [Fischer et al. 2018; Kellnhofer et al. 2019; Krafka et al. 2016; Sugano et al. 2014; Zhang et al. 2020, 2017c] have countered these issues. Regardless of these advancements, the potential of using them in creating scalable multi-user eye-tracking systems remains unexplored.

2 PROBLEM STATEMENT

In general, an eye-tracking system works in two steps; (1) Identify and localize facial/ocular regions, and (2) Estimate gaze directions/positions. In a multi-user eye-tracking system, we would require to extend the idea to multiple users by estimating gaze information for all the users detected in the frame. Despite the solution appearing trivial, there are several problems associated with it.

Most importantly, the camera setup we use would restrict the number of users the system can support. The number of users a single camera can track depend on the camera's resolution, the field of view, and the experimental setup. In order to be cost-effective, we will focus on extending the range by incorporating multiple cameras. Moreover, the computations associated with the overall system need to be optimized for different computational devices such as GPU-accelerated, non-GPU accelerated, and mobile. In the case of using a single computational setup based on minimum hardware, the system would not enable us to exploit the computing capabilities of the higher-end system, whereas targeting higher-end systems would restrict the accessibility of the eye-tracking system. As a result, the target system requires scalability based on the hardware setup.

Further, compared with traditional eye-trackers, a multi-user eye-tracking system will pose restrictions on calibration. Explicitly considering using the system in the wild, the calibration per user is not an ideal implementation. Instead, the system needs to be capable of estimating gaze with few to no calibration samples. Finally, to be used by a practitioner, the overall system requires including commonly used gaze metrics such as fixational and saccadic measures. Since the system is concurrently used by multiple users, in addition to measures per user, the system also requires multi-user gaze measures.

During the study, I will investigate the following research questions,

(1) Identify scalable commodity hardware-based setup(s) for multi-user eye-tracking.
(2) Develop a deep learning-based gaze estimation architecture suitable for gaze estimation in the wild.

This work is licensed under a Creative Commons Attribution International 4.0 License.

ETRA '22, June 8–11, 2022, Seattle, WA, USA
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9252-5/22/06.
https://doi.org/10.1145/3517031.3532197

(3) Determine approaches for integrating the identified architectures in a multi-user eye-tracking setup.
(4) Incorporate data from gaze estimation to form commonly used eye-tracking metrics.

3 APPROACH AND METHODS

I plan to implement a multi-user eye-tracking system with multi-cameras capable of scaling based on the hardware setup. The overall system will comprise a multi-step pipeline, from capturing the users/ gaze subjects to estimating gaze positions/ metrics for each user (see Figure 1).

The first step of the pipeline is capturing participants through a multi-camera setup. As highlighted earlier, here, we use an array of commodity cameras. Then we use the frames from the cameras to detect users in the environment. An issue here is that there can be overlapping regions, and the system might perform redundant computations due to the overlapping regions. In order to overcome the issue, we assume the camera array to be stationary relative to each camera causing overlapping regions not to vary across steps. Therefore at the initialization step, the system can identify overlapping regions and perform computations once per overlapping region in subsequent steps. Moreover, we plan to include multiple models for user detection, each optimized for each different type of platform used.

The next step of the pipeline is to estimate the gaze of each detected face expressed as pitch and yaw angles concerning the detected face. Since the face patches can be of different sizes depending on the distance from the camera, we will use bilinear interpolation to resize images to the size used by the model. We ensure scalability like the face detection model by including a series of optimized models.

The final step of the pipeline is to estimate the gaze positions or metrics depending on the requirement. In the case of gaze positions, we can incorporate a world view as seen by the participants and estimate the gaze positions. Further, using gaze positions, we can compute fixational and saccadic metrics.

4 RESULTS

We created a prototype application capable of eye-tracking two users using a web camera to conduct a preliminary study. The prototype application used the FaceMesh [Kartynnik et al. 2019] model to detect faces and an EfficientNet [Tan and Le 2019] based gaze estimation model trained using the XGaze [Zhang et al. 2020] dataset for estimating gaze. We assessed the prototype on gaze estimation and an experimental task.

During the experiments, our Gaze estimation model achieved a gaze error of 5.22 for XGaze testing examples. Even though our model exhibited a low accuracy compared to the baseline of 4.5, our model used a significantly lesser number of parameters, 4.3M, compared to 26 M in the baseline model. We consider the results to provide potential room for improvement through architectural modifications of the neural network.

Moreover, to test the utility of the proposed approach, we attempted to replicate a joint attention experiment [Guo et al. 2018]. We experimented with five volunteers on joint attention with a proctor instructing users to look at an on-screen target(out of three). During each session, we gathered information on the gaze positions of each user with the corresponding timestamp and derived fixational eye movement measures. Based on computations, we identified the average time to the first fixation as 1.25 seconds and the average fixation duration as 5.45 seconds.

5 FUTURE WORK

We explore deep neural models for a potential model candidate for the gaze estimation network. Here we plan to thoroughly investigate the gaze estimation capabilities across datasets, the scalability, and throughput on different platforms. Further, we plan to extend the current prototype to a multi-camera environment with more than two users and overlapping regions. Through the prototype's extension, we plan to investigate potential approaches we can follow in multi-camera environments. Then we will integrate eye-tracking metrics for our setup, including the most commonly used fixational and saccadic measures. Finally, we plan to conduct a comprehensive user study to determine the approach's utility. In this study, we will evaluate the performance of our system against commonly used eye trackers.

6 BROADER IMPACT

The multi-user eye-tracking using commodity hardware will contribute to behavioral studies in the wild, often confined to controlled environments and requires costly hardware. Moreover, we expect these studies to help better understand human behavior. Further, the study will improve the accessibility to eye-tracking research in the community by eliminating potential barriers by providing cost-effective and scalable eye-tracking systems.

ACKNOWLEDGMENTS

This work is supported in part by the U.S. National Science Foundation grant CAREER IIS-2045523. Any opinions, findings and conclusion or recommendations expressed in this material are the author(s) and do not necessarily reflect those of the sponsors.

REFERENCES

Stylianos Asteriadis, Paraskevi Tzouveli, Kostas Karpouzis, and Stefanos Kollias. 2009. Estimation of behavioral user state based on eye gaze and head pose—application in an e-learning environment. *Multimedia Tools and Applications* 41, 3 (2009), 469–493.

Senuri De Silva, Sanuwani Dayarathna, Gangani Ariyarathne, Dulani Meedeniya, Sampath Jayarathna, Anne MP Michalek, and Gavindya Jayawardena. 2019. A rule-based system for ADHD identification using eye movement data. In *2019 Moratuwa Engineering Research Conference (MERCon)*. IEEE, 538–543.

Tobias Fischer, Hyung Jin Chang, and Yiannis Demiris. 2018. Rt-gene: Real-time eye gaze estimation in natural environments. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 334–352.

Shengyao Guo, Eric Ho, Yalun Zheng, Qiming Chen, Vivian Meng, John Cao, Si Wu, Leanne Chukoskie, and Pamela Cosman. 2018. Using face and object detection to quantify looks during social interactions. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 1144–1148.

Yury Kartynnik, Artsiom Ablavatski, Ivan Grishchenko, and Matthias Grundmann. 2019. Real-time facial surface geometry from monocular video on mobile GPUs. *arXiv preprint arXiv:1907.06724* (2019).

Petr Kellnhofer, Adria Recasens, Simon Stent, Wojciech Matusik, and Antonio Torralba. 2019. Gaze360: Physically unconstrained gaze estimation in the wild. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 6912–6921.

Kyle Krafka, Aditya Khosla, Petr Kellnhofer, Harini Kannan, Suchendra Bhandarkar, Wojciech Matusik, and Antonio Torralba. 2016. Eye tracking for everyone. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2176–2184.

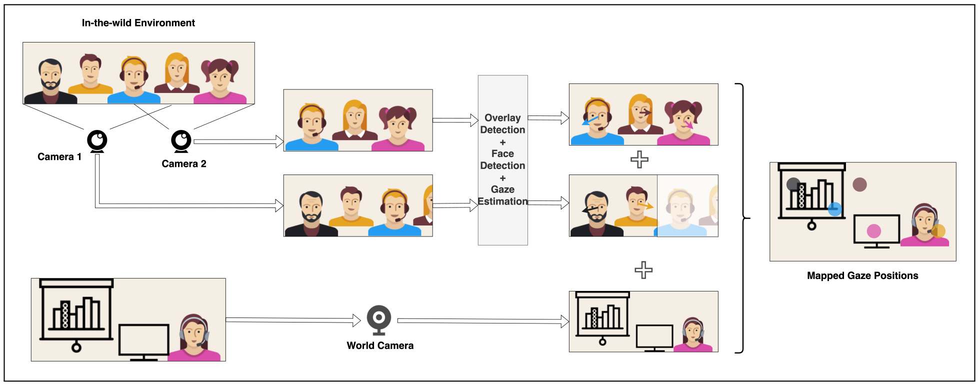

Figure 1: Proposed architecture of the multi-user eye-tracking system. We start by capturing the participants/users through a multi-camera setup (camera 1 and camera 2). Then we detect the overlapping regions, faces/ ocular regions and estimate the gaze directions. By incoporating with a world view (world camera), we can identify real-world gaze positions corresponding to each user.

Bhanuka Mahanama, Gavindya Jayawardena, and Sampath Jayarathna. 2021. Analyzing Unconstrained Reading Patterns of Digital Documents Using Eye Tracking. In *2021 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*. IEEE, 282–283.

Anne MP Michalek, Gavindya Jayawardena, and Sampath Jayarathna. 2019. Predicting ADHD using eye gaze metrics indexing working memory capacity. In *Computational Models for Biomedical Reasoning and Problem Solving*. IGI Global, 66–88.

Bilge Mutlu, Toshiyuki Shiwa, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. 2009. Footing in human-robot conversations: how robots might shape participant roles using gaze cues. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*. 61–68.

Oskar Palinko, Francesco Rea, Giulio Sandini, and Alessandra Sciutti. 2016. Robot reading human gaze: Why eye tracking is better than head tracking for human-robot collaboration. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 5048–5054.

Alexandra Papoutsaki, James Laskey, and Jeff Huang. 2017. Searchgazer: Webcam eye tracking for remote studies of web search. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*. 17–26.

Primesh Pathirana, Shashimal Senarath, Dulani Meedeniya, and Sampath Jayarathna. 2022. Eye gaze estimation: A survey on deep learning-based approaches. *Expert Systems with Applications* (2022), 116894.

Yusuke Sugano, Yasuyuki Matsushita, and Yoichi Sato. 2014. Learning-by-synthesis for appearance-based 3d gaze estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1821–1828.

Mingxing Tan and Quoc Le. 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning*. PMLR, 6105–6114.

Xucong Zhang, Seonwook Park, Thabo Beeler, Derek Bradley, Siyu Tang, and Otmar Hilliges. 2020. ETH-XGaze: A large scale dataset for gaze estimation under extreme head pose and gaze variation. In *European Conference on Computer Vision*. Springer, 365–381.

Xucong Zhang, Yusuke Sugano, Mario Fritz, and Andreas Bulling. 2015. Appearance-based gaze estimation in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4511–4520.

Xucong Zhang, Yusuke Sugano, Mario Fritz, and Andreas Bulling. 2017b. It's written all over your face: Full-face appearance-based gaze estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 51–60.

Xucong Zhang, Yusuke Sugano, Mario Fritz, and Andreas Bulling. 2017c. Mpiigaze: Real-world dataset and deep appearance-based gaze estimation. *IEEE transactions on pattern analysis and machine intelligence* 41, 1 (2017), 162–175.

Yanxia Zhang, Ken Pfeuffer, Ming Ki Chong, Jason Alexander, Andreas Bulling, and Hans Gellersen. 2017a. Look together: using gaze for assisting co-located collaborative search. *Personal and Ubiquitous Computing* 21, 1 (2017), 173–186.