Scalable Webcam Eye Tracking by Learning from User Interactions

Alexandra Papoutsaki
Department of Computer Science
Brown University
Providence RI 02912 USA
alexpap@cs.brown.edu

CHI'15 Extended Abstracts, April 18–23, 2015, Seoul, Republic of Korea.
ACM 978-1-4503-3146-3/15/04.
http://dx.doi.org/10.1145/2702613.2702627

Abstract
Eye tracking systems are commonly used in a variety of research domains, but cost thousands of dollars. In my thesis I investigate a new approach to enable eye tracking for common webcams. The aim is to provide a natural experience to everyday users that are not restricted to laboratories and highly controlled studies. The accuracy of eye tracking webcams will be improved by user interactions which continuously calibrate the eye tracker during regular usage. Eye tracking can become a reality for many potential applications such as large-scale naturalistic user studies, online gaming, or enabling people to perform hands-free navigation of websites.

Author Keywords
eye tracking; webcam; user interactions

ACM Classification Keywords
H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous.

Research Situation
I am a fourth year PhD student in the Department of Computer Science of Brown University. I have successfully completed the required coursework, and passed the comprehensive exams on a research project I undertook

over my first two years in my program, fulfilling the requirements to acquire a Masters degree.

I spent my first three years of my PhD working in the area of Computational Biology. My background is in Computer Science, but I am interested in applying my theoretical knowledge on more human centric research. I joined the newly founded lab on Human Computer Interaction in my department last year and since then I have been exploring different research projects. I anticipate that I will graduate in my sixth year.

My research focuses on eye tracking using webcams and user interactions. I have created a recommendation system that uses click interactions to recommend areas of interest in search engine web pages and plan to include gaze data captured by webcams. In the future I aim to incorporate more user interactions, like touch activity to improve my eye tracker.

By participating in the Doctoral Consortium I expect to receive feedback by experts in eye tracking. I anticipate pointers to related work and ideas for potential extensions that I could use to better shape my thesis. Apart from senior researchers I hope to meet fellow PhD students that can bring a different perspective in my work. Since my department is quite small and I am the most senior PhD student in my lab, it would be greatly beneficial for me to be exposed in different types of research. That would not only allow me to get a better grasp of my thesis but to also contemplate on my more general trajectory as a scientist working in HCI.

Context and Motivation

Eye tracking systems are commonly used for psychology experiments, human computer interaction studies, medical research, marketing studies, and many more applications that seek to measure where the user is looking. Typical eye trackers today are passive, remote, and comprise a horizontal bar with an infrared camera that is placed in a fixed distance from the subject. They are highly specialized, and cost thousands of dollars, thus being widely unavailable. In my work I plan to investigate a new approach to eye tracking for common webcams already present in laptops and mobile devices.

Webcams for eye tracking have been studied before using offline software, and unsurprisingly webcams have been found to be less accurate than specialized eye tracking equipment. However, recently just over 50% of web browsers support the HTML5 API for accessing the webcam, typical laptop webcams now support higher resolutions of capture, and laptops are now more capable of computing realtime eye-gaze positions. These advances make real-time webcam eye-tracking possible online, and thus enable applications that scale to large numbers of users; but these advancements do not solve the problem of poor accuracy due to poor lighting and low frame rate.

Huang et al. [4] showed that when a user clicks on a web page, they will first look at the target they intend to click. Inhoff and Gordon [5] describe a study that shows that the eye is most likely to be 2–4 characters to the right of the last typed character. We hypothesize based on preliminary work [3] that a person would need to look at an area in a touchscreen about a second before tapping it, especially because of the finger occlusion problem [1, 2]. Webcam images during these interactions can be collected and used as cues for how the user's pupil looks like when interacting with a particular location. Future and past observations of similar pupils can be matched as the proposed eye tracking system collects mappings of pupil features to eye-gaze locations on the page. These and other user

interactions can be collected by the website, and used as ground truth, i.e. the more a person uses the website, the more training data can be collected for calibrating the eye tracker. The calibration data can be collected globally the entire time a user interacts with a website, and without disrupting the user experience. Furthermore, the calibration data can be captured at the beginning of a browsing session to provide training data that better match the local environment in terms of ambient lighting, position, and background, solving the problem of having to recalibrate each time a user sits down at a computer.

Thesis or Problem Statement

By making eye tracking accessible to a web browser and continuously improving the tracking accuracy, eye tracking becomes a reality for many potential applications such as online gaming, large-scale naturalistic user studies, or even navigation of web interface only by gaze. The user's gaze could be used as an input technique for people with hand motor impairments. This eye tracking procedure is opt-in as browsers request access to the webcam, and the website captures data only if the user agrees. For usability testing, the users may be employed by the web company to help improve the quality of the website. As another example, search engines can reward users to do eye tracking in order to learn which search results and advertisements the user has viewed (and perhaps did not click) and help them improve their relevance. In my thesis I plan to develop an experiment that uses the webcam eye tracker to make in-page recommendations of what areas of a web page to examine, based on the behavioral patterns of similar users.

Research Goals and Methods

Firstly, I have focused on developing a robust eye tracker that utilizes webcams and captures gaze activity. I have been working on developing a JavaScript library, which upon user's consent, activates the webcam and receives video stream. I have been implementing different face and eye detection algorithms that identify the location of the user's face and eyes. This has already been demonstrated to be possible in real-time [6] by using Haar-like features provided by OpenCV, and by using classic object detection algorithms like Viola-Jones [7].

Next, with the location of the eye within the captured image, I use corner detection algorithms to identify the location of the pupils. I have already preliminary results that show successful capture of the pupil, but I aim to refine the model I am using so that to capture in real-time and with higher accuracy the position of the pupil.

I have been using basic trigonometry to combine several variables as the eye and head position, the angle of their gaze trajectory, and the user's distance to the screen, in order to compute the gaze location. I plan to investigate more advanced techniques that could potentially improve the accuracy of the predicted gaze.

Soon I will be focusing in capturing more general user interactions such as clicking, tapping (for touchscreens), entering text, scrolling, etc.. I plan to use heuristics informed from prior user interaction experiments (e.g. [4, 5]) to match the pupil features with the location of interaction.

As the user interacts more with the website, the eye tracking model's parameters are updated. Over time, the model will gain additional calibration data. As part of this research, I will evaluate how robust these model parameters are over time, and when the eye tracking model needs to be recalibrated if the utility of the parameters gathered from past observations start to decay.

Finally, I have conducted preliminary research on pointing interactions to recommend areas of interest in search engines. I identify behaviorally similar users based on the similarity of their cursor activity. With the webcam eye tracker, I will be able to identify past users with similar gaze examination patterns. These users are even more likely to be similar in mindset to the active user. Future users need only a webcam for semi-accurate eye tracking to gain useful visual recommendations, and having more accurate eye tracking through the user interaction cues would improve the quality of the visual recommendations.

Dissertation Status

I have some preliminary results on the development of an eye-tracking suite but I plan to improve the accuracy of the predicted gaze. I have designed a recommendation system that points search engine users to potential regions of interest, based on their click activity. Once I finalize my eye tracker I will incorporate webcam data in the recommendation system. The problem I have not worked so far and I anticipate that will be the most time consuming and challenging is to build a learning model that will utilize general user interactions in order to to automatically calibrate the webcam eye tracker.

Expected Contributions

The resulting contributions will be: First, the research, development, and evaluation of a real-time online eye tracking system using the typical webcam available in laptops and mobile devices (with source code made available publicly). Second, research into how user interactions such as cursor clicks, text entry, touches can be used to automatically train eye tracking algorithms. Finally, an evaluation of recommendation algorithms that use webcam data on a controlled experiment of navigating search engine result pages.

Acknowledgements

I would like to thank my advisor Jeff Huang for his help, support, and academic advising.

References

[1] Albinsson, P.-A., and Zhai, S. High precision touch screen interaction. In *Proceedings of CHI* (2003), 105–112.

[2] Benko, H., Wilson, A. D., and Baudisch, P. Precise selection techniques for multi-touch screens. In *Proceedings of CHI* (2006), 1263–1272.

[3] Huang, J., and Diriye, A. Web user interaction mining from touch-enabled mobile devices. In *HCIR Workshop* (2012).

[4] Huang, J., White, R. W., and Buscher, G. User see, user point: Gaze and cursor alignment in web search. In *Proceedings of CHI* (2012), 1341–1350.

[5] Inhoff, A. W., and Gordon, A. M. Eye movements and eye-hand coordination during typing. *Current Directions in Psychological Science* (1997), 153–157.

[6] Lundgren, E., Rocha, T., Rocha, Z., Carvalho, P., and Bello, M. tracking.js: A modern approach for Computer Vision on the web. http://trackingjs.com/, 2014. [Online; accessed 2014-08-15].

[7] Viola, P., and Jones, M. Rapid object detection using a boosted cascade of simple features. In *Proceedings of CVPR*, vol. 1 (2001), 511–518.