SearchGazer: Webcam Eye Tracking for Remote Studies of Web Search

Alexandra Papoutsaki
Brown University
alexpap@cs.brown.edu

James Laskey
Brown University
jlaskey@cs.brown.edu

Jeff Huang
Brown University
chiir@jeffhuang.com

ABSTRACT

We introduce SearchGazer, a web-based eye tracker for remote web search studies using common webcams already present in laptops and some desktop computers. SearchGazer is a pure JavaScript library that infers the gaze behavior of searchers in real time. The eye tracking model self-calibrates by watching searchers interact with the search pages and trains a mapping of eye features to gaze locations and search page elements on the screen. Contrary to typical eye tracking studies in information retrieval, this approach does not require the purchase of any additional specialized equipment, and can be done remotely in a user's natural environment, leading to cheaper and easier visual attention studies.

While SearchGazer is not intended to be as accurate as specialized eye trackers, it is able to replicate many of the research findings of three seminal information retrieval papers: two that used eye tracking devices, and one that used the mouse cursor as a restricted focus viewer. Charts and heatmaps from those original papers are plotted side-by-side with SearchGazer results. While the main results are similar, there are some notable differences, which we hypothesize derive from improvements in the latest ranking technologies used by current versions of search engines and diligence by remote users. As part of this paper, we also release SearchGazer as a library that can be integrated into any search page.

Keywords

web search behavior; remote user studies; online eye tracking; user interactions; gaze prediction

1. INTRODUCTION

Search is a visual activity. Users examine search results to determine what is relevant to them and their task. Knowing what a searcher has examined, or is looking at, has been the focus of numerous studies in information retrieval. Typically, the goal is to understand searcher behavior and apply that information to improve the search systems. Traditionally, these studies are done in lab with specialized eye trackers, or inferred using remotely collected interaction data like clicks or cursor activity.

CHIIR '17, March 07 - 11, 2017, Oslo, Norway

ISBN 978-1-4503-4677-1/17/03…$15.00

DOI: http://dx.doi.org/10.1145/3020165.3020170

We explore a new approach to understanding visual attention in search that leverages the advantages of both types of studies: scalability across millions of users, naturalistic environments, and real webcam-based gaze tracking. SearchGazer is a pure JavaScript eye tracking library that infers the gaze behavior of users in real time. The eye tracking model self-calibrates by watching searchers interact with search pages and trains a mapping between features of the eye and positions on the screen. Huang et al. [15] have shown that when a user clicks on a web page, they will first look at the target where they intend to click. Webcam images during these user interactions can be collected and used as cues for what the user's eyes look like. Future observations of the eye can be matched to similar past instances to infer the eye-gaze, even when the user is not interacting.

SearchGazer can be added to any search page for standard eye tracking and for identifying which search elements the user is looking at. In this paper, we investigate the utility of SearchGazer in the context of web search, assessing its ability to substitute or at least approximate specialized eye trackers. Currently, while there have been software artifacts that use webcams for eye tracking (typically by processing the video stream offline), there is no published research describing the implementation or application of browser-based webcam eye trackers (e.g. [32]). Further, we ask the question, "can SearchGazer *really* be useful for search behavior studies?"

We investigate this by directly replicating some of the main results of three past studies: Cutrell et al. [8] and Buscher et al. [5] presented highly-cited eye tracking studies which investigated searcher behavior on the presentation of search results and search advertisements respectively. Lagun et al. [21] used the cursor as a restricted focus viewer, also a remote behavior capture technique, but which obscures most of the search page, hindering the user experience. Therefore, we directly substitute the specialized eye tracker or cursor-as-a-viewer interface with SearchGazer. While we briefly report on the accuracy of SearchGazer in Section 3, the focus is on the ultimate evaluation: whether researchers conducting a prior study performed with an eye tracker would reach similar conclusions with SearchGazer, performed remotely, online, and in real time without any special equipment.

These studies were conducted simultaneously with crowdworkers, with more participants at a lower cost (in terms of time and money). Indeed, many of the main results are quite similar, and we show the original charts and heatmaps side-by-side with corresponding charts and heatmaps generated by SearchGazer. For the results that are different, we discuss plausible explanations, primarily due to the change in search technology since the original study and differences in the diligence of in-lab participants and remote crowd-workers.

The main contributions of this work are: 1) the description and evaluation of our real-time online webcam eye tracker, SearchGazer (with source code publicly available at `http://webgazer.cs.brown.edu/search`), and 2) the investigation of results from replication of three seminal web search behavior papers, when SearchGazer is substituted for specialized eye trackers or interfaces.

2. RELATED WORK

2.1 Webcam Eye Tracking

There are relatively few academic publications about using webcams for eye tracking, with most webcam eye trackers typically being video streams processed offline (not real-time) and involving an explicit calibration phase. There has been work on eye tracking self-calibration using image saliency to estimate user gaze for calibration purposes [30], but this is a very rough estimate of where a user is looking. Alnajar et al. [1] introduced a webcam eye tracker that self-calibrates with pre-recorded gaze patterns instead of predicted saliency. Although more accurate, it still requires users to view stimuli for which "ground truth" gaze patterns have been recorded. Xu et al. introduced TurkerGaze [32], a webcam eye tracker deployed on Amazon Mechanical Turk for image saliency prediction. It requires calibration which is performed during a game phase where users lock their gaze on a specific target. Moreover, it requires an offline training component. Huang et al. [17] developed PACE, a standalone desktop application that performs eye tracking using user interactions. SearchGazer is distinguished by self-calibrating in real time via gaze-mouse interaction relationships which are readily available, as well as being able to be used on any web page without any software installation or configuration. In addition, it can predict both the absolute gaze location and the corresponding search element on the screen. These attributes make SearchGazer ideal for remote online studies and experiments, especially in the context of web search.

Several software artifacts for eye detection have been made available online, though without much formal evaluation. OpenGazer [34] is an open source desktop application that performs eye detection using computer vision algorithms from OpenCV; it has been abandoned since 2010. Camgaze.js [31] is a JavaScript library that predicts the pupil location and gaze direction in real-time, but does not map it to the screen. Clmtrackr [25] is a JavaScript library that performs facial feature tracking through constrained local models. SearchGazer adopts clmtrackr for eye detection and uses our own algorithms for mapping the gaze on the screen.

There have also been commercial forays into online webcam eye tracking for usability studies. Tobii Technologies has spun off a company called Sticky that helps websites optimize advertisements based on visual behaviors. One of the earlier services to offer webcam eye tracking was GazeHawk, which was acquired by Facebook in 2012 and is now shut down. Like our work, their system tracked the user in their natural environment from the browser, without the need to install software. However, their approach is significantly different as they transmitted the webcam video to their own servers for offline processing. They did so because at that time, laptops were not capable of processing the video data in real time [personal communication with GazeHawk founders]. Additionally, they required a phase of user calibration and did not include user interactions. Finally, a recent startup called Xlabs focuses on head tracking to determine the gaze position and has built a Chrome browser compatible software extension that can be installed by its users.

2.2 Examination Behavior in Web Search

Lab studies involving eye tracking during web search have been commonly used to trace visual attention. Past research has found a correlation between gaze and cursor positions in search behavior [11, 28]. Guo and Agichtein [11] reported that the distance between cursor and gaze positions was larger along the x-axis, and was generally shorter when the cursor was placed over the search results. Their model could predict with 77% accuracy when gaze and cursor were strongly aligned using cursor features. Huang et al. [15] note that the notion of gaze and cursor correlation is overly naive; instead their relationship greatly depends on what the user is doing at that time. They show that the two are highly correlated when users aim at or hit a target, but the correlation is poor when the cursor is idle. Navalpakkam et al. [26] investigate the gaze-cursor relationship on non-linear page layouts which, in search, may represent cases when information or advertisements are shown in a second column. Furthermore, they perform gaze prediction using a non-linear model and identify particular regions of interest. Xu et al. [33] created a computational model that predicts spatio-temporal visual attention on graphical interfaces based on cursor movements, keyboard activity, and the UI components. Boi et al. [2] created a method for segmenting content groups in pages that based on mouse movements identifies the user-perceived group of contents with a 20% mean pixel-based error. Lagun et al. [22] devised an approach that combines user interactions and salience of the web page to infer visual attention in search. Liu et al. [23] extended this work by using visual saliency maps derived from image content to predict users examination behavior on an experimental browser.

Numerous studies of web search use eye tracking or some proxy (like cursor activity) as a tool for understanding searchers and design better search systems. Buscher et al. [4] used eye tracking features to infer user interest and show that this can yield great improvements when personalizing search. Huang et al. [16] have investigated the meaning behind cursor interactions, and how they can improve our understanding of searcher behavior along with the relevance of search results for future users. Buscher et al. [6] and Dumais et al. [9] notice that users have different gaze behavior patterns, but can be clustered into different personalities: exhaustive examiners, economic examiners focused on the organic results or also on the ads. Liu et al. [24] tap into the different phases of gaze behavior in web search by developing a two-stage model that examines the "skimming" and "reading" phases. Finally, beyond traditional web search, Kules et al. [20] understand gaze behavior in a faceted search interface and find that as part of examining results, users spend half the time looking at facets, prompting "task building that incorporates consideration of the dimensions of the task."

3. SEARCHGAZER

We have developed SearchGazer, a self-calibrated client-side eye tracking library that extends WebGazer [27] and trains a regression model which maps eye features to gaze locations and search page elements during user interactions. In addition to predicting the gaze of a user within any device display which has a browser that supports access to the webcam, SearchGazer also identifies gaze periods over regions of interest on the search results page for analysis. A few lines of JavaScript code are enough to integrate SearchGazer in any search page and perform eye tracking once the user starts interacting with the page naturally. The software is open source and available at `http://webgazer.cs.brown.edu/search`.

SearchGazer is relatively simpler than trackers with explicit 3D reasoning [12]. This simplicity allows it to run in real time through browser JavaScript. The primary novelty of SearchGazer is to constantly self-calibrate based on cursor-gaze relationships. Not only does this eliminate the need for explicit calibration sessions, it means users are free to move and SearchGazer will learn new mappings between eye features and screen coordinates.

Any facial feature detection library can be plugged in SearchGazer; it only needs the location of the eyes within the video. Here, SearchGazer uses clmtrackr [25] and accepts as input the smallest rectangle that fits the eye contour.

3.1 Eye Features

SearchGazer relies on a regression model that maps video pixels to gaze locations. To detect the eye features we follow the same procedure with TurkerGaze [32] and represent each eye region as a resized 6×10 image patch. The two eye regions are processed with gray-scaling and histogram equalization, resulting to a 120D feature that will be input to the regression algorithm described below. Unlike TurkerGaze, SearchGazer does not post-process offline and retrains in real time.

3.2 Mapping to Screen and Self-Calibration

Matching the computed eye features to screen coordinates requires a mapping between the 120D vectors and the coordinates of the user's gaze on the device screen. This complex relationship depends on the 3D position and rotation of the head with respect to the camera and screen. These 3D properties can be estimated, but generally require careful calibration and expensive computation. SearchGazer avoids this by using a simpler mapping between eye features and display coordinates. In addition, it relies on continual self-calibration through user interactions that naturally take place in web search and do not disrupt the user experience.

We base our model on the assumption that when a user interaction occurs, the gaze locations on the screen match the screen coordinates of that interaction. In a web search study, Huang et al. showed that the gaze-cursor distance averages 74 pixels [15] during a click. Research on attention control and its allocation mechanisms led to similar findings [10]. For simplicity, we assume that the gaze and cursor align perfectly during clicks.

The click history and the corresponding gaze predictions are stored locally in the browser, thus avoiding privacy concerns of storing eye features in a remote location. No data are transmitted from the user's computer to the website hosting the SearchGazer code, other than the continuous predictions and their corresponding cursor positions in the case of clicks. SearchGazer's predictions are not affected by scrolling and are projected within the window viewport. Whenever the user is not directly interacting with the page, the camera still captures eye features, and applies the regression model.

3.2.1 *Mapping Eye Features*

To map the eye pixels to gaze locations we implement a ridge regression model which maps the 120D eye feature vector to the display coordinates (D_x, D_y) for each click. The simplicity of this regularized linear regression allows it to provide relatively accurate predictions in real time and with only a few clicks needed for training.

Without loss of generality, we consider the ridge regression model function for the x-coordinate prediction: $f(\mathbf{v}) \rightarrow D_x$. This function is $f(\mathbf{v}) = \phi(\mathbf{x})^T\mathbf{w}$, where $\phi(\mathbf{x})$ is a basis function and $\mathbf{w}$ is a vector of weights which satisfy:

$$\min_{\mathbf{w}} \sum_{x_i \in \mathbf{x}} ||D_{x_i} - f(x_i)||_2^2 + \lambda||\mathbf{w}||_2^2 \qquad (1)$$

Here the last term λ acts as a regularization to penalize overfitting. In our study we set $\lambda = 0.00001$.

3.2.2 *Sampling Cursor Movements*

Past studies have shown that when users move their cursor to perform an action there is a strong correlation between cursor and gaze location [13]. This distance grows when the cursor remains idle. We extend the ridge regression model, taking into consideration that the location of the cursor is a good signal for the gaze location when the cursor is active. Our assumption for clicks remains the same. When cursor moves, we assume it matches the true gaze location. Unlike click coordinates though, cursor locations contribute to the regression model for at most 200ms, a duration comparable to that of a gaze fixation. Therefore, when the cursor is idle and no new cursor location has been introduced, our model falls back to the original simple ridge regression where only clicks contribute to the training of SearchGazer.

3.2.3 *Mapping to Search Elements*

The predicted gaze coordinates are combined with the DOM structure of the underlying search page and mapped to examined page elements such as links, snippets, and ads.

4. EVALUATION

In [27], we conducted two user studies—one online and one in-lab—with a total population of 87 participants. Gaze predictions from SearchGazer were compared to those made by the commercial eye tracker Tobii EyeX. The mean error was 128.9 pixels with an average visual angle of 4.17° or 1.6 inches. Figure 1 shows the average distance (in pixels) between the predictions made by SearchGazer's regression model and the true gaze locations across 50 clicks. These results are promising as they show that SearchGazer can predict relatively accurately eye-gaze locations.

To further investigate the applicability and utility of SearchGazer in web search, we replicate the studies found in three seminal papers in the area of information retrieval that used eye tracking to better understand web search behavior. We conducted all three replication studies remotely, recruiting participants through the Amazon Mechanical Turk crowdsourcing platform. All crowd-workers passed a qualification test which ensured they had a webcam and their browser supported the getUserMedia() API that provides access to the webcam stream. To ensure lack of bias each study was conducted with a unique population of crowd-workers.

The gaze predictions made by SearchGazer are not fine-grained enough to identify fixations. Instead, we rely on

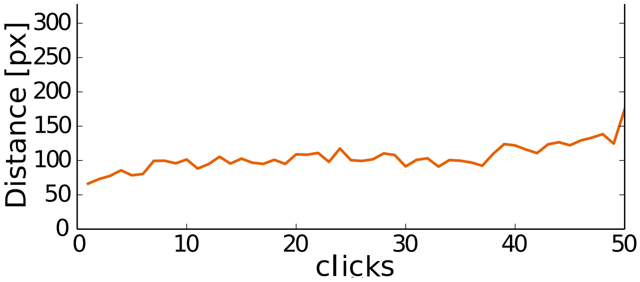


Figure 1: Average distance in pixels between SearchGazer and the true gaze locations across 50 clicks.

raw gaze data when comparing our findings to previous studies. In addition, each of the heatmaps included in the following sections was created according to the color palette of the corresponding original study. In the following sections we report the results when replicating the three seminal studies. Due to lack of the original reference data we cannot perform any rigorous statistical analysis to assess our findings. Instead, we report average differences and high-level metrics that give a general quantified measure of the similarities between SearchGazer and the original studies.

5. RESULT EXAMINATION BEHAVIOR STUDY

The first study we reproduce is [8], a prominent early study that used eye tracking in web search. Cutrell et al. conducted an in-lab user study with 22 participants, exploring the effects of changes in the presentation of search results, i.e. the snippet length, on 6 informational and 6 navigational queries that were submitted to MSN Search. A Tobii x50 eye tracker was used to identify gaze coordinates.

5.1 Experimental Design and Procedure

To replicate this study we performed a few modifications as some of its specifications are outdated. Given that MSN Search does no longer exist as an independent search engine, we used its successor, Bing. One of the navigational queries ("Pinewood") was also changed to a current software company ("Symantec"). In contrast to [8], our participants were only allowed to look for the answer within the first page of returned results and could not manipulate the given query. For each of the 12 queries, the first search engine result page (SERP) was downloaded from Bing, without snippet length manipulation. All ads were removed so that the SERPs resembled as much as possible the MSN Search, leading to 8 instead of 10 organic results. SearchGazer was added on each of the 12 SERPs to predict and log in real time the gaze locations.

Our version of the study started with instructions and a consent form. As tasks designed for crowd-workers tend to be shorter than in-lab studies, we included an explicit calibration step, during which crowd-workers had to click on a circular target that appeared in a 5×3 grid. The calibration step was conducted again at the middle of the study. Crowd-workers proceeded with the 12 search tasks presented in a randomized order. For each task, a description of the search goal and a corresponding query was given. After a search, crowd-workers provided their answer or declared they were unable to successfully acquire the target information. After the study, they filled an online demographic questionnaire.

5.2 Participants

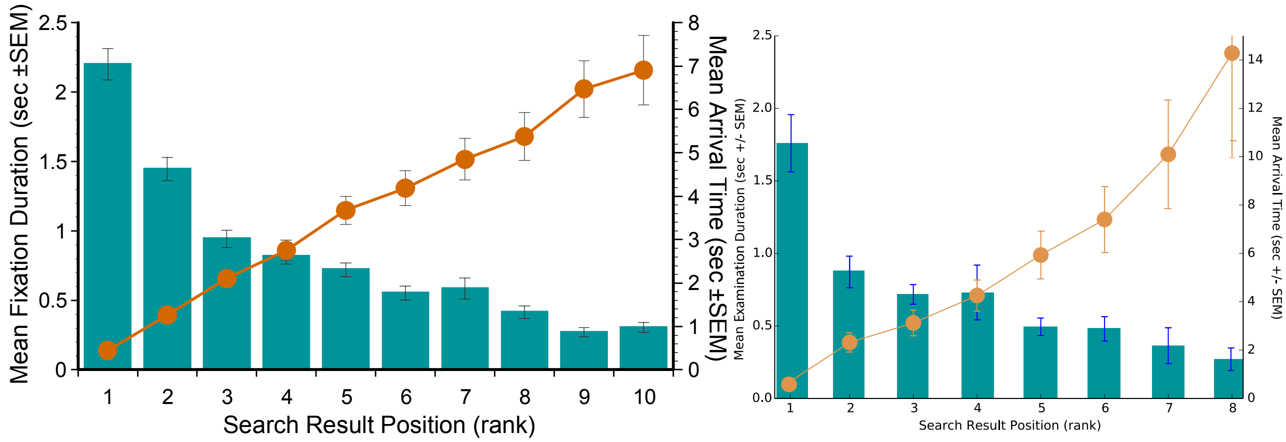
49 crowd-workers performed this study. 13 participants were excluded due to abandoning the task, not following the instructions, or due to technical issues with data logging in our server. The final population consisted of 36 participants (14 female, 22 male). They were 20 to 49 years old (M=30.1, SD=7.24). 22 had normal vision, 8 wore glasses and 6 contact lenses. All participants received 2USD at the end of the experiment. To ensure they completed the whole study, they provided an ID they were handed with the questionnaire. The study lasted on average 10.11 minutes (SD=4.13). In total, there were 610 clicks and 76,389 gaze predictions.

5.3 Results

Cutrell et al. provided preliminary results on the general characteristics of search results, along with changes in the web search behavior when varying the snippet length. We focus on the former, as those findings are more generalizable and applicable to modern search engines.

Viewing order and fixation duration: Research in information retrieval has repeatedly shown that users tend to examine results from top to bottom when presented with SERPs that have a single-column linear layout [18]. Figure 2 shows in circles the mean time for the gaze to arrive at each organic result. Cutrell et al. confirmed previous findings, showing in Figure 2a that the mean time for the gaze to arrive at each result is roughly linear, with lower ranked results attracting attention last. Figure 2b shows the corresponding mean arrival times for the data obtained through SearchGazer. Note that in our replication studies a SERP contained at most 8 organic results. We observe that the arrival times also follow a linear fashion but the slopes are significantly different. For lower ranked results, arrival times are higher in our study, with result 8 being reached on average after 14.2 seconds. We hypothesize that as search engines have become more powerful web searchers tend to trust the first ranks even more, exploring the bottom of the page only after careful consideration of the first results. After normalization, the average difference between the original study and our findings is 14.75%.

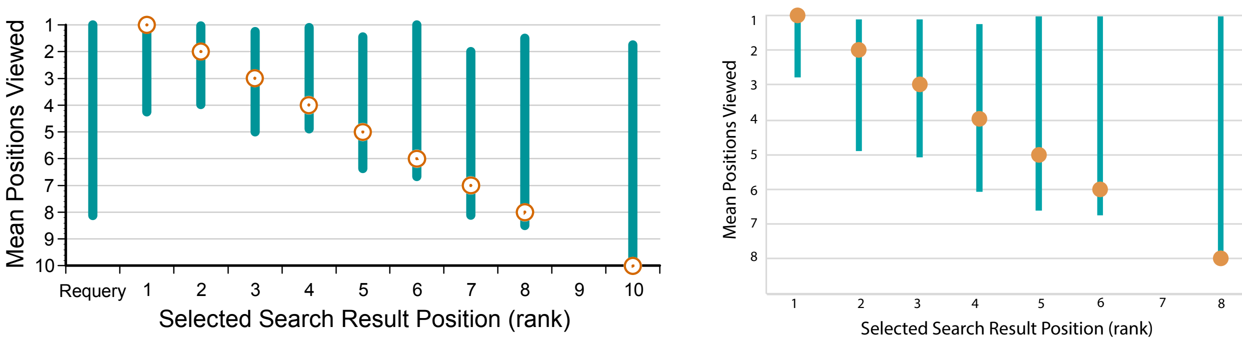
The second component of Figure 2 is the average fixation duration for each result, depicted in bars. As shown in Figure 2a from [8], most gaze activity was directed at the first results, which attracted far more attention. In Figure 2b, SearchGazer's predictions demonstrate similar total visual attention towards lower-ranked results. Following [19], we assume that the power law fits the data better than any other common distribution. Fitting two power law curves on the original and SearchGazer's results, we find that the exponents are 0.7235 and 0.7906, respectively. After normalization, the average difference between the fixations in the original study and our findings is 5.09%. We observe that the curves have similar slopes, although the crowd-workers of our experiment spent less time examining each result. This can be perhaps explained by the difference in nature of a remote and an in-lab study. As crowd-workers are unattended and use Amazon Mechanical Turk as their source of income, they tend to go through the tasks faster and possibly not as diligently. In addition, crowd-workers often come from different countries, with English being a secondary language and therefore they might approach the tasks differently than in-lab participants who are often native speakers.



(a) Result Examination Behavior Study

(b) SearchGazer

Figure 2: Mean fixation/examination duration (bars) and mean time for gaze to arrive at each result (circles).



(a) Result Examination Behavior Study

(b) SearchGazer

Figure 3: Mean number of search results looked at before users clicked on a result (above and below that result).

Results Viewed before a click: An interesting measure of saturation is how many results were viewed on average before a click occurred on a result. Figure 3 shows the average number of results ranked higher and lower than the one that the user clicked on. For example, in Figure 3a, users from [8] who clicked on result 5, on average looked at almost all items above it and about 1.5 results below it. Figure 3b shows our findings when we analyzed SearchGazer's predictions. Crowd-workers that clicked on result 5, on average look at 4 items above and 1.6 below it. The average difference between the two studies is 0.28 results for those ranked higher and 0.61 results for those ranked lower than the one clicked. Missing the reference data prevents us from running any rigorous statistical test, but on average it seems that SearchGazer can replicate similar studies relatively accurately.

6. AD EXAMINATION BEHAVIOR STUDY

Buscher et al. [5] investigated contemporary search engines that contain ads and related searches in addition to organic results. They conducted a lab user study, varying the type of task (informational or navigational) and the quality of ads (relevant or irrelevant). The experiment was conducted with 38 participants that were provided with custom generated SERPs for each of the above combinations, for a total of 32 search tasks. A Tobii x50 eye tracker was used to measure visual attention across different areas of interest (AOIs).

6.1 Experimental Design and Procedure

The authors of [5] provided the list of 32 queries along with the corresponding SERPs that were used in the original study. Each query could return one of two SERPs that varied in the quality of ads (relevant or irrelevant to the query). SearchGazer was added on all SERPs to predict gaze locations in real-time. To keep the study short and remotely practical, we only worked with 12 of the original queries (6 informational, 6 navigational). The experimental design was identical to the procedures followed in the "result examination behavior study" described in Section 5. The ordering of tasks was also randomized across all participants but for the ad quality we followed the same protocol as [5]. Each participant was assigned to one of three ad quality blocks. Each block contains 12 trials, with ads being mostly good (relevant), bad (irrelevant) or randomly selected.

6.2 Participants

Forty-four crowd-workers performed this study. Nine were filtered out due to incomplete or abandoned tasks. The resulting 35 participants consisted of 17 female and 18 male, with ages ranging from 21 to 59 years (M=30.4, SD=9.1).

Of these participants, 19 had normal vision, 13 wore glasses and 3 contacts. The study lasted on average 9.75 minutes (SD=7.05) and in total there were 88,438 gaze predictions.

6.3 Measures

The following two measures were used as defined in [5]:
AOIs: Each SERP was broken into separate AOIs that correspond to 10 organic results, 3 top ads, and 5 rail ads.
Fixation Impact: Buscher et al. used the measure of fixation impact [3] which spreads the duration of a fixation to all AOIs that fall close to the fixation center using a Gaussian distribution. They used Tobii Studio to detect fixations, for which the exact technique is unknown. We instead used raw gaze predictions and a smaller radius Gaussian.

6.4 Results

General Gaze Distribution on SERPs: The gaze heatmap in Figure 4a demonstrates the distribution of visual attention across all participants in [5]. The data have been aggregated across all queries and the background SERP serves as an example. Figure 4b shows the corresponding heatmap created by the predictions of SearchGazer across all 12 queries and 35 participants. We observe that both heatmaps follow the golden triangle, with the majority of visual attention focused in the first three organic results. For SearchGazer there is a wider spread of predictions across the whole SERP, as it is not as stable as a commercial eye tracker. Nevertheless, it is worth noting that the aggregated data can lead to similar conclusions between the two studies.

Figure 5 shows the mean fixation impact for each AOI across all participants and tasks. The results from [5] in Figure 5a show that most visual attention falls on the top results. Figure 5b shows the corresponding SearchGazer results. Our findings have the same linear decay across organic results, with the exception of the 7th and 8th which are examined for longer. This is perhaps due to the page-fold falling near them or because crowd-workers are more deliberate in the examination of lower results. In addition, our study shows smaller overall examination durations. Surprisingly, the five right ads attract much higher visual attention. SearchGazer predictions could lack precision, as we noticed that the inferred gaze positions were more scattered along the x-axis. After normalization, the average difference between the mean fixation impact as seen across the two studies is 28.78%.

Effects of Task Type: Figure 6 shows the mean fixation impact for AOIs, split between informational and navigational tasks. Both [5] and our results show in Figures 6a and 6b respectively that participants spent more time on SERPs for informational tasks. This additional time was mostly spent on the organic results. After normalization, the average difference between the mean fixation impact is 21.85% for informational and 29.52% for navigational tasks.

Effects of Ad Quality: Figure 7 shows the mean fixation impact for AOIs, separated based on ad quality (good or bad). Buscher et al. did not find any statistical difference between the time spent in SERPs with good and bad ads, but showed that participants devoted about twice as much visual attention to top ads when the ads were of good quality. In contrast, participants paid less attention to the organic results when good quality ads were displayed, as shown in Figure 7a. Our findings in Figure 7b indicate in many cases a totally different picture, revealing that webcam eye tracking can miss such subtle differences. The fact that our study included 12 instead of 32 tasks might have also reduced the effect that ads normally have. After normalization, the average difference between the mean fixation impact is 23.63% for pages with good ads and 25.48% for bad ads.

7. RESTRICTED FOCAL VIEW RESULT EXAMINATION STUDY

Lagun and Agichtein [21] created ViewSer, a tool that automatically modifies the appearance of a SERP to show one result at a time, while blurring the rest of the interface using a restricted focal view. The participant can uncover only one result using their cursor, thus the search engine knows which result a user is examining. Although ViewSer is an interface and not an eye tracker, it allows researchers to infer web search behavior remotely without the need of additional equipment. Our work with SearchGazer builds on ViewSer's idea of capturing examined regions of the search page at scale, and their assessing the feasibility of the work through crowd-workers. The authors validated the utility of ViewSer by running a remote user study with 106 crowd-workers. Each worker went through a list of 25 benchmark search tasks from the Web Track of the TREC 2009 competition. The results were compared to an in-lab study that was performed with 10 participants using a Tobii x60 eye tracker. Clickthrough and viewing rates were comparable between participants using ViewSer and those tracked using the physical eye tracker.

7.1 Experimental Design and Procedure

We did not have access to the original SERPs, so instead replicated this study using Google and focusing on 12 queries. We downloaded the 12 Google SERPs and added SearchGazer on each. As with all 3 replication studies crowd-workers were allowed to only click within the first page of returned results and could not manipulate the query. The rest of the protocol was the same as the Result Examination Behavior Study.

7.2 Participants

Forty-seven crowd-workers performed this study. Eleven were excluded due to incomplete and abandoned tasks, resulting to 36 participants (12 female, 24 male). Their ages ranged from 21 to 63 years (M=31.97, SD=10.42). Twenty-five had normal vision, 10 wore glasses and 1 wore contacts. The study lasted on average 10.36 minutes (SD=6.83). Across all participants there were 630 clicks and 76,602 gaze predictions.

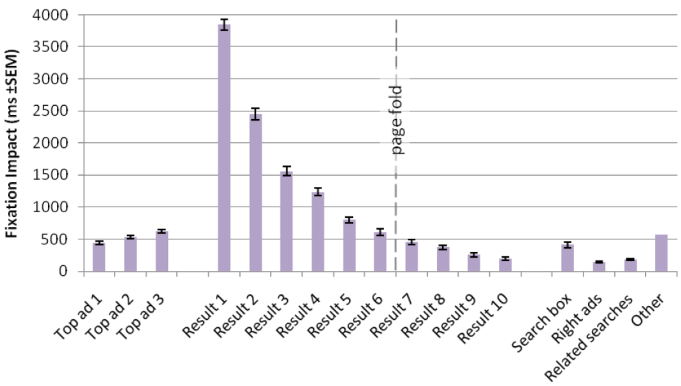
7.3 Results

Gaze Distribution: Figure 8 shows an example heatmap of the relative viewing time spent on the SERP that corresponds to the query "toilet". For [21] this heatmap can be created only with data collected from the in-lab eye tracking study, as shown in Figure 8a. Data gathered from ViewSer can be visualized with vertical colorbars as shown in Figure 8a, as colorbar (b). The (a) colorbar corresponds to data gathered from the in-lab eye tracking study. SearchGazer which predicts in real-time the gaze locations as screen coordinates can lead to both types of visualizations, allowing for richer information. Figure 8b shows the corresponding heatmap created with our data. As the organic results in the two SERPs are not identical, it is hard to compare them directly. It is worth noting though, that as the task is informational ("Find information on buying, installing, and repairing toilets"), participants tend to spend more time on the SERP, examining even lower-ranked results.

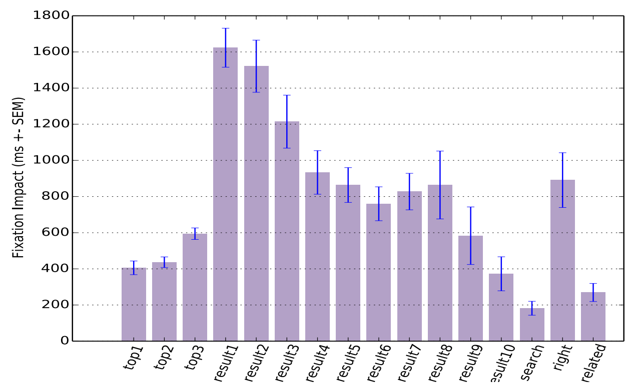
(a) Ad Examination Behavior Study

(b) SearchGazer

Figure 4: Gaze heatmap aggregated across all participants and queries and projected on a sample SERP.

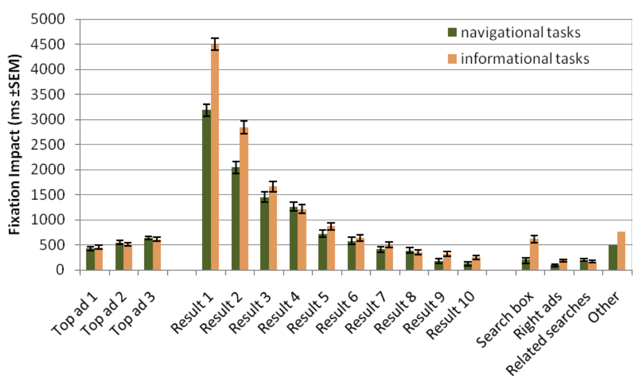


(a) Ad Examination Behavior Study

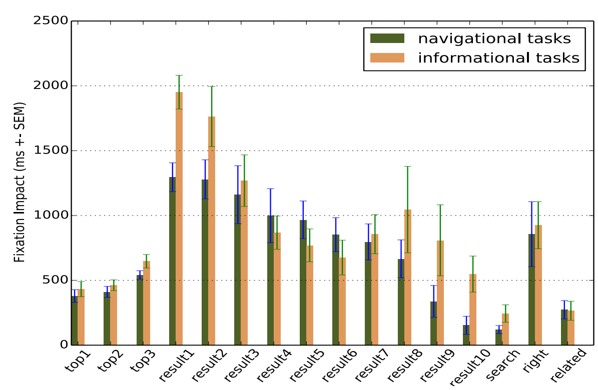


(b) SearchGazer

Figure 5: Mean fixation impact on SERP elements in milliseconds.



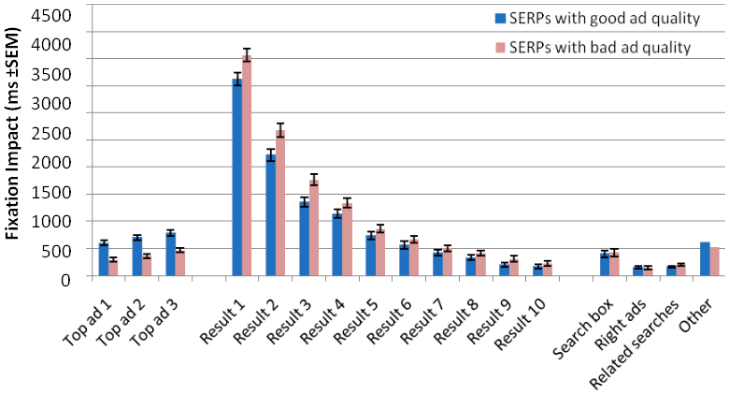
(a) Ad Examination Behavior Study



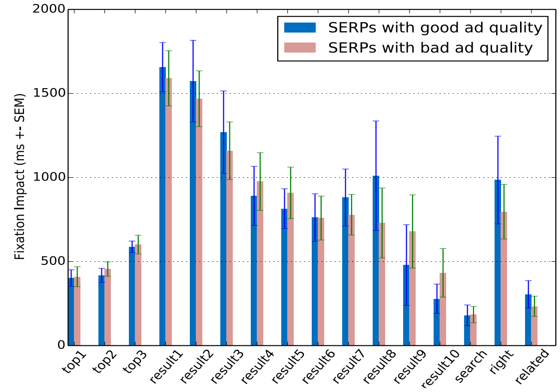
(b) SearchGazer

Figure 6: Comparison of mean fixation impact on SERP elements for navigational and informational tasks.

SERP Examination and Clickthrough: Figure 9 depicts the viewing and clickthrough rates across all queries and participants. As shown in Figure 9a, the data gathered from the ViewSer group demonstrate that both viewing and clickthrough rates decay in a linear fashion, with lower ranked results attracting less attention. Figure 9b shows the corresponding data gathered from the predictions made by SearchGazer. It is worth noting that even though the same linear trend is observed the rates are lower. After normalization, the average difference between the viewing rate is

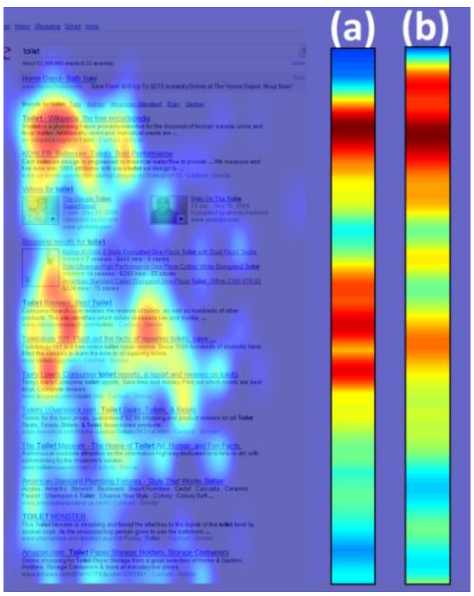


(a) Ad Examination Behavior Study

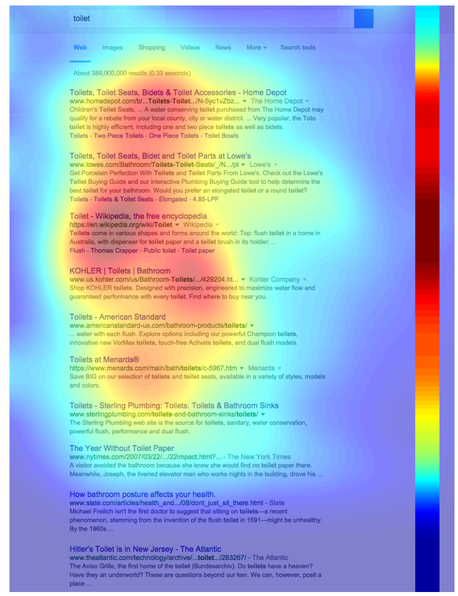


(b) SearchGazer

Figure 7: Comparison of mean fixation impact on AOIs for SERPs displaying good or bad ads.



(a) Result Examination Behavior Study via Restricted Focal View. Colorbar (a): eye tracking group. Colorbar (b): ViewSer group.



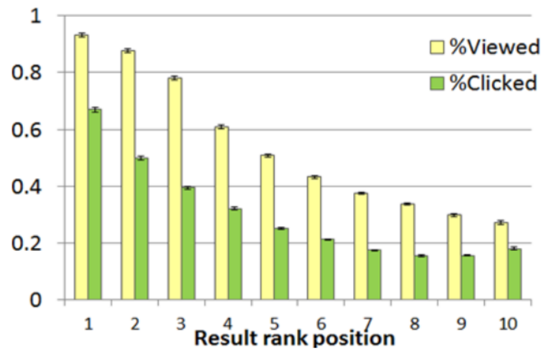
(b) SearchGazer. SearchGazer allows the creation of gaze heatmaps, unlike ViewSer which is restricted to density colorbars.

Figure 8: Attention heatmap over a SERP for the query "toilet" and its corresponding colorbar showing the heatmap density.

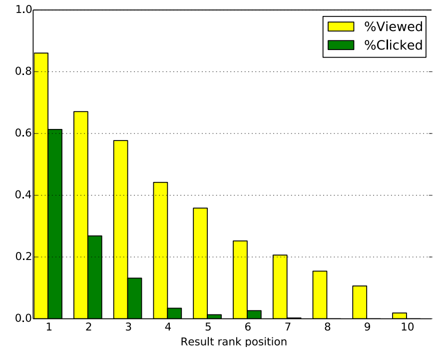
14.11% and 28.54% for the clickthrough rates. This could be a result of the differences in the user experience across the two studies. Restricted focus viewing can lead users to carefully examine more results instead of just scanning them, as they now have to move their cursor to reveal one result at a time. At the same time, ViewSer can lead to a closer examination of results that would otherwise be overlooked, leading to higher clickthrough rates in lower ranked results. In our study, the bottom results attracted less attention and even fewer clicks. As many of the informational tasks were vague and acceptable target information existed in many results, it is not unlikely that our crowd-workers ended up clicking on the first few results, trusting the search engine.

8. DISCUSSION

Replicating these three studies revealed both the potential and limitations of performing webcam eye tracking in place of specialized equipment and interfaces. Many of our findings, such as Figure 3, achieved similar conclusions compared with the original studies, showcasing that SearchGazer can be successfully used in experiments where the goal is to measure the distribution of gaze locations. SearchGazer was able to recreate general trends and even highlight differences in viewing times across individual organic results. In comparison to a restricted focus viewer [21], SearchGazer does not disturb the user experience by blurring the SERP. Once users consent to giving access to their webcam, they can continue navigating the web page as they would normally do, while SearchGazer collects interactions and predicts gaze activity in the background. Overall, using SearchGazer we were able to reproduce three studies with numerous charts and heatmaps at a fraction of cost, effort, and time it would normally take if those studies were conducted in lab. Multiple crowd-workers performed the study simultaneously and without the need of



(a) Result Examination Behavior Study via Restricted Focal View



(b) SearchGazer

Figure 9: Viewing and clickthrough rates for each rank, aggregated across all queries and participants.

monitoring, allowing us to test SearchGazer with far richer and more diverse computational and ambient environments.

On the other hand, there are certain limitations that we cannot ignore. Although [5] is the closest to what we could replicate most precisely, there were specific differences that were surprising and demonstrate the need to expand our understanding of SearchGazer's constraints. A possible explanation for those differences is our lack of an algorithm to identify fixations, relying instead on raw gaze data. Although the overall aggregated data were almost always very close to the original studies, we can hypothesize that removing saccades would lead to a clearer picture that would allow replication of more fine-detailed studies. Coming up with an algorithm custom-built for SearchGazer to identify fixations as is available in existing eye trackers is one future direction. Further, SearchGazer assumes that the location of clicks and cursor movements is equal to that of the gaze. Temporal and spatial differences in this relationship might bias our model. Since our studies were conducted remotely, we lack information about the nature of SearchGazer's errors. In the future, we are interested in pursuing in-lab studies and extensively compare SearchGazer with physical eye trackers to better understand its behavior, strengths, and limitations.

Our current implementation of SearchGazer is simply a prototype demonstrating webcam eye tracking. It is not intended to be the most optimal or accurate version of online webcam eye tracking. Part of why we are releasing the software online as open source is so other researchers and developers can refine its computer vision components. The software is completely modular and different eye detectors, regression models, or interaction-based learning algorithms can be completely swapped out. In fact, we have successfully tried 3 publicly available eye detectors, 2 regression models, and multiple ways of using user interactions to train the model. The version selected for the study replications was chosen as it performs well across many peoples' facial appearances and environmental conditions. For a more extensive evaluation of SearchGazer in generic sites see [27].

8.1 Privacy

Ultimately, the use of webcams in online applications poses privacy risks, but there can be significant benefits if used appropriately, allowing websites to improve their usability or conduct experiments to better understand human behavior. Hong et al. [14] state that users will accept the privacy risks only if there are clear benefits that outweigh them.

SearchGazer is opt-in, as browsers request access to the webcams and the gaze predictions are computed in real time on the user's computer so the video stream is not sent over the web (unlike webcam eye trackers like Sticky or GazeHawk). Therefore, users decide between the trade-off of transmitting their eye-gaze behavior, in exchange for some benefit from the search engine. We imagine scenarios where users are compensated or offered other incentives, like an advertising-free search interface or additional gaze-based search features.

9. CONCLUSION

We proposed SearchGazer, a real-time online eye tracker using only the common webcam as a way to determine users' examination behavior on search pages. Using SearchGazer, we revisit in today's search environments the key findings from: a search results page examination study from CHI 2007, a search advertisement examination study from SIGIR 2010, and study of a restricted focus viewer based on the cursor from SIGIR 2011. The findings from reproducing past web search studies showed that the approach of conducting remote eye tracking studies through webcams is not unreasonable.

This new approach can be transformative, as examination behavior can be understood at scale for diverse search scenarios: when users perform infrequent queries, when search interface designers seek to test new features or layouts. In fact, numerous information retrieval models seek to infer which search results a user has examined (e.g., [7, 29]); clearly, this signal is important to the web search community, even when not measured perfectly. Compared to lab studies, remote crowd-workers can perform tasks whenever and wherever they choose, without the need for any special equipment or software installation. Remote webcam eye tracking is therefore considerably cheaper than an in-person lab study required for typical eye trackers, saving time for both the participants and experimenters. Additionally, experimenters are able to release the tasks which can be performed by remote crowd-workers immediately and simultaneously, allowing for faster feedback to inform search engine design.

10. ACKNOWLEDGMENTS

This research is supported by NSF grants IIS-1464061, IIS-1552663, and the Brown University Salomon Award.

11. REFERENCES

[1] F. Alnajar, T. Gevers, R. Valenti, and S. Ghebreab. Calibration-free gaze estimation using human gaze patterns. In *Proc. ICCV*, pages 137–144, 2013.

[2] P. Boi, G. Fenu, L. D. Spano, and V. Vargiu. Reconstructing user's attention on the web through mouse movements and perception-based content identification. *ACM TAP*, 13(3):15:1–15:21, 2016.

[3] G. Buscher, E. Cutrell, and M. R. Morris. What do you see when you're surfing?: using eye tracking to predict salient regions of web pages. In *Proc. CHI*, pages 21–30, 2009.

[4] G. Buscher, A. Dengel, and L. van Elst. Eye movements as implicit relevance feedback. In *Proc. CHI Extended Abstracts*, pages 2991–2996, 2008.

[5] G. Buscher, S. Dumais, and E. Cutrell. The good, the bad, and the random: an eye-tracking study of ad quality in web search. In *Proc. SIGIR*, pages 42–49, 2010.

[6] G. Buscher, R. White, S. Dumais, and J. Huang. Large-scale analysis of individual and task differences in search result page examination strategies. In *Proc. WSDM*, pages 373–382, 2012.

[7] O. Chapelle and Y. Zhang. A dynamic bayesian network click model for web search ranking. In *Proc. WWW*, pages 1–10, 2009.

[8] E. Cutrell and Z. Guan. What are you looking for?: an eye-tracking study of information usage in web search. In *Proc. CHI*, pages 407–416, 2007.

[9] S. Dumais, G. Buscher, and E. Cutrell. Individual differences in gaze patterns for web search. In *Proc. IIiX*, pages 185–194, 2010.

[10] C. L. Folk, R. W. Remington, and J. H. Wright. The structure of attentional control: contingent attentional capture by apparent motion, abrupt onset, and color. *Journal of Experimental Psychology: Human perception and performance*, 20(2):317, 1994.

[11] Q. Guo and E. Agichtein. Towards predicting web searcher gaze position from mouse movements. In *Proc. CHI Extended Abstracts*, pages 3601–3606, 2010.

[12] D. W. Hansen and Q. Ji. In the eye of the beholder: A survey of models for eyes and gaze. *IEEE TPAMI*, 32(3):478–500, 2010.

[13] D. Hauger, A. Paramythis, and S. Weibelzahl. Using browser interaction data to determine page reading behavior. In *User Modeling, Adaption and Personalization*, pages 147–158. Springer, 2011.

[14] J. Hong, J. Ng, S. Lederer, and J. Landay. Privacy risk models for designing privacy-sensitive ubiquitous computing systems. In *Proc. DIS*, pages 91–100, 2004.

[15] J. Huang, R. W. White, and G. Buscher. User see, user point: Gaze and cursor alignment in web search. In *Proc. CHI*, pages 1341–1350, 2012.

[16] J. Huang, R. W. White, and S. Dumais. No clicks, no problem: using cursor movements to understand and improve search. In *Proc. CHI*, pages 1225–1234, 2011.

[17] M. X. Huang, T. C. Kwok, G. Ngai, S. C. Chan, and H. V. Leong. Building a personalized, auto-calibrating eye tracker from user interactions. In *Proc. CHI*, pages 5169–5179, 2016.

[18] T. Joachims, L. Granka, B. Pan, H. Hembrooke, and G. Gay. Accurately interpreting clickthrough data as implicit feedback. In *Proc. SIGIR*, pages 154–161, 2005.

[19] T. Joachims, L. Granka, B. Pan, H. Hembrooke, F. Radlinski, and G. Gay. Evaluating the accuracy of implicit feedback from clicks and query reformulations in web search. *ACM TOIS*, 25(2):7, 2007.

[20] B. Kules, R. Capra, M. Banta, and T. Sierra. What do exploratory searchers look at in a faceted search interface? In *Proc. JCDL*, pages 313–322, 2009.

[21] D. Lagun and E. Agichtein. Viewser: Enabling large-scale remote user studies of web search examination and interaction. In *Proc. SIGIR*, pages 365–374, 2011.

[22] D. Lagun and E. Agichtein. Inferring searcher attention by jointly modeling user interactions and content salience. In *Proc. SIGIR*, pages 483–492, 2015.

[23] Y. Liu, Z. Liu, K. Zhou, M. Wang, H. Luan, C. Wang, M. Zhang, and S. Ma. Predicting search user examination with visual saliency. In *Proc. SIGIR*, pages 619–628, 2016.

[24] Y. Liu, C. Wang, K. Zhou, J. Nie, M. Zhang, and S. Ma. From skimming to reading: A two-stage examination model for web search. In *Proc. CIKM*, pages 849–858, 2014.

[25] A. Mathias. clmtrackr: Javascript library for precise tracking of facial features via Constrained Local Models. https://github.com/auduno/clmtrackr, 2014. [Online; accessed 2016-05-08].

[26] V. Navalpakkam, L. Jentzsch, R. Sayres, S. Ravi, A. Ahmed, and A. Smola. Measurement and modeling of eye-mouse behavior in the presence of nonlinear page layouts. In *Proc. WWW*, pages 953–964, 2013.

[27] A. Papoutsaki, P. Sangkloy, J. Laskey, N. Daskalova, J. Huang, and J. Hays. Webgazer: Scalable webcam eye tracking using user interactions. In *Proc. IJCAI*, pages 3839–3845, 2016.

[28] K. Rodden, X. Fu, A. Aula, and I. Spiro. Eye-mouse coordination patterns on web search results pages. In *Proc. CHI Extended Abstracts*, pages 2997–3002, 2008.

[29] R. Srikant, S. Basu, N. Wang, and D. Pregibon. User browsing models: relevance versus examination. In *Proc. SIGKDD*, pages 223–232, 2010.

[30] Y. Sugano, Y. Matsushita, and Y. Sato. Calibration-free gaze sensing using saliency maps. In *Proc. CVPR*, pages 2667–2674, 2010.

[31] A. Wallar, A. Sazonovs, C. Poellabauer, and P. Flynn. Camgaze.js: Browser-based Eye Tracking and Gaze Prediction using JavaScript. http://aw204.host.cs.st-andrews.ac.uk/camgaze.js, 2014. [Online; accessed 2015-03-12].

[32] P. Xu, K. A. Ehinger, Y. Zhang, A. Finkelstein, S. R. Kulkarni, and J. Xiao. Turkergaze: Crowdsourcing saliency with webcam based eye tracking. *arXiv preprint arXiv:1504.06755*, 2015.

[33] P. Xu, Y. Sugano, and A. Bulling. Spatio-temporal modeling and prediction of visual attention in graphical user interfaces. In *Proc. CHI*, pages 3299–3310, 2016.

[34] P. Zielinski. Opengazer: open-source gaze tracker for ordinary webcams. http://www.inference.phy.cam.ac.uk/opengazer, 2007. [Online; accessed 2014-08-24].