eyeSee: A Multimodal Lookup Tool

Jan Pilzer
University of British Columbia
pilzer@cs.ubc.ca

Hayley Guillou
University of British Columbia
guillouh@cs.ubc.ca

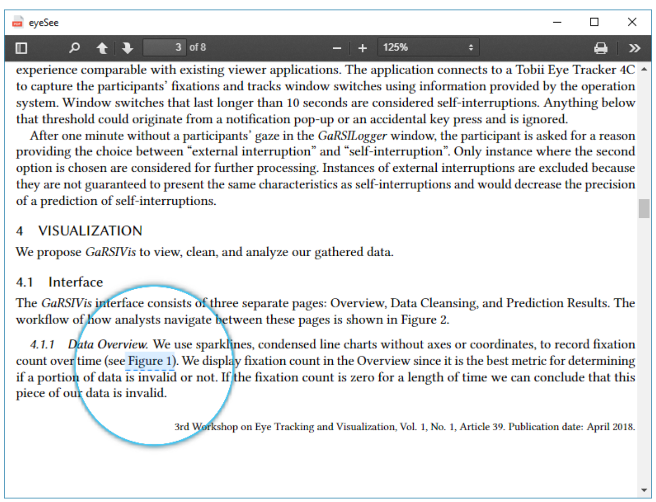


Figure 1: Highlighting a reference to a figure on gaze in *eyeSee*.

Abstract

Digital documents are a great way to read paper without the environmental impact of using paper; however, scrolling through a long document can be much more difficult than simply flipping a page. We introduce a novel technique for jumping to referenced content within a paper while preserving the shape of the paper and reducing the resumption lag for the reader. While our system was originally designed for jumping between text and citations, an analysis of the reading habits of potential users prompted a redesign to focus on jumping between text and information-rich tables and figures. Our evaluation found that users enjoyed the tool and most thought that they would use it in real life.

Introduction

Reading PDF-format research papers is an everyday activity for graduate students and professionals working in research. This type of reading requires continuous visual attention and it is easy for the reader to lose their place if interrupted from reading text. The problem with academic research papers is that they contain referenced content within the text, including figures, tables, and citations. This means that in order to fully understand and comprehend the content of the paper, the reader must leave their current place in the text to explore another location within the digital document. This change in location causes an interruption and thus introduces resumption lag (the time taken to restart reading after an interruption) into the reading experience.

As a solution, we present *eyeSee*, a PDF viewer capable of utilizing the user's gaze to enable quick navigation to referenced content. The key ideas of *eyeSee* are to highlight linked content, to allow the user to see the content by holding a keyboard key, and to highlight the last reading position when the user releases the key. We conduct an iterative design process and adjust the design to best suit the natural reading patterns of our users. Through evaluation, we find *eyeSee* does not show improvement in reducing cognitive load; however, this is likely due to flaws and oversights in the experimental design. We address these issues and suggest future evaluations to overcome them. More importantly, we receive positive feedback regarding the comprehensiveness and usefulness of the tool.

Related work

Lee's paper [4] seeks to improve a user's performance in skim reading, by blurring the majority of the text and elements and spotlighting certain important figures and elements while scrolling quickly. This ensures minimal processing time per object and avoids motion blur by keeping these limited objects static. We also seek to feature figures and important elements; however our focus is on natural reading tasks and not skim reading.

EyeBookmark [3], a gaze-based bookmarking systems seeks to reduce the resumption lag of an interruption during reading. It provides a visual cue to help the reader remember where they should resume reading. This project is quite similar to ours, but we hope to assist the user in finding the content that causes the interruption in addition to improving their resumption lag.

Xiaohan's iReference tool [7] is a gaze-based reading aid that displays the citation details nearby whenever the reader is interested in them. iReference's goal is to help readers access the citation with natural eye movements in a convenient and natural manner. iReference is the project with the most similarities to ours, but our projects have different goals. iReference seeks to use natural eye movements for accessing citations, while our goal is to aid the user's natural reading experience, including all of the inevitable interruptions.

Design rationale

In essence, *eyeSee* intends to bridge the gap between a reference and the referenced context, for example a citation and its source, when the reader requires it. Research papers contain several types of content: The text, floating figures and tables, and a bibliography. To tie them together, figures and related works are referenced in the text. The details of a citation or the actual image of a figure reference are not positioned in and often not even next to the paragraph of text mentioning them. This positioning has aesthetic as well as practical reasons. Floating images and the source of a citation can break the reading flow when position in the text and are helpful for understanding, but usually not necessarily required. With their current usual positioning, an uninterested reader can skip over figures in the text and stop reading when they reach the bibliography.

A tool like *eyeSee* therefore should not restructure the base document, but provide some optional interaction that enables reader to access referenced content. We see two main options of either showing the content, for example a figure, next to the reference in any form of overlay or navigating to the figure from the reference. We decided on the latter as we believe there are several advantages to this approach. A navigation-interaction preserves the structure of the document which allows readers to keep a mental model of the paper for easer skimming and potential switches of medium, another device or a print out, later on. Additionally, it does have the issue of occlusion that is a major difficulty for any overlay-based interaction. By using highlighting for the reference target, we ensure that the user can quickly focus on the right content.

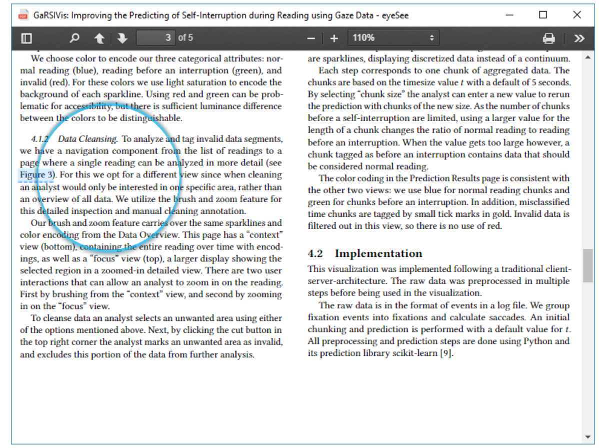
Figure 2 shows the usage of *eyeSee*. The user is reading a research paper that contains figures and references to those in the text. On fixating on one of those figure references, it becomes highlighting indicating the possible interaction. When pressing and holding down the Shift-key, *eyeSee* navigates to the referenced figure and highlights it using a bright border. The user can inspect the figure while holding down the key. On release, *eyeSee* navigates back to the reference, which is still highlighted to aid resumption of the reading. Until the user fixates another reference they can repeat the navigation to the same figure while continuing their reading.

Implementation

In order to evaluate our idea with actual papers, we built *eyeSee* as PDF viewer application using Electron [1] and PDF.js [5]. This setup allows us to create a desktop app providing all of the features users expect from a standard PDF viewer leading to an optimized user experience for our participants.

The linking of references to their targets takes advantage of existing semantic markup included in the PDF documents. When created with the proper tools, citations and references to figures, tables, and sections are recognized as links by PDF.js. On fixation, our implementation adds the highlighting to the link and handles the user's jump command.

The connection to the Tobii 4c eye tracker [6] is established by using the Tobii C# SDK. The fixation events provided by


GaRSIVis: Improving the Predicting of Self-Interruption during Reading using Gaze Data - eyeSee

3 of 5 110%

We choose color to encode our three categorical attributes: normal reading (blue), reading before an interruption (green), and invalid (red). For these colors we use light saturation to encode the background of each sparkline. Using red and green can be problematic for accessibility, but there is sufficient luminance difference between the colors to be distinguishable.

4.1.2 Data Cleansing. To analyze and tag invalid data segments, we have a navigation component from the list of readings to a page where a single reading can be analyzed in more detail (see Figure 3). For this we opt for a different view since when cleaning an analyst would only be interested in one specific area, rather than an overview of all data. We utilize the brush and zoom feature for this detailed inspection and manual cleaning annotation.

Our brush and zoom feature carries over the same sparklines and color encoding from the Data Overview. This page has a "context" view (bottom), containing the entire reading over time with encodings, as well as a "focus" view (top), a larger display showing the selected region in a zoomed-in detailed view. There are two user interactions that can allow an analyst to zoom in on the reading. First by brushing from the "context" view, and second by zooming in on the "focus" view.

To cleanse data an analyst selects an unwanted area using either of the options mentioned above. Next, by clicking the cut button in the top right corner the analyst marks an unwanted area as invalid, and excludes this portion of the data from further analysis.

are sparklines, displaying discretized data instead of a continuum.

Each step corresponds to one chunk of aggregated data. The chunks are based on the timesize value t with a default of 5 seconds. By selecting "chunk size" the analyst can enter a new value to rerun the prediction with chunks of the new size. As the number of chunks before a self-interruption are limited, using a larger value for the length of a chunk changes the ratio of normal reading to reading before an interruption. When the value gets too large however, a chunk tagged as before an interruption contains data that should be considered normal reading.

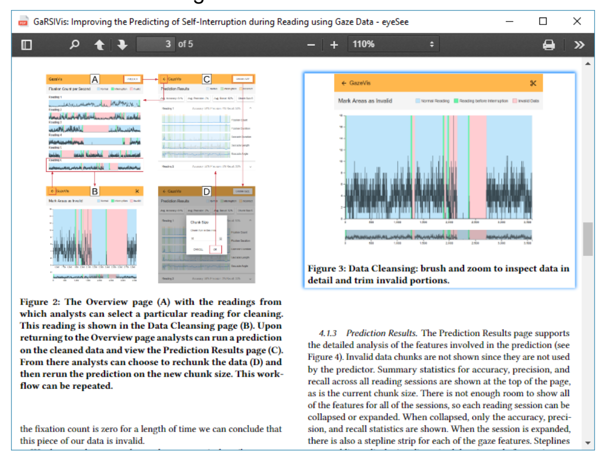
The color coding in the Prediction Results page is consistent with the other two views: we use blue for normal reading chunks and green for chunks before an interruption. In addition, misclassified time chunks are tagged by small tick marks in gold. Invalid data is filtered out in this view, so there is no use of red.

4.2 Implementation

This visualization was implemented following a traditional client-server-architecture. The raw data was preprocessed in multiple steps before being used in the visualization.

The raw data is in the format of events in a log file. We group fixation events into fixations and calculate saccades. An initial chunking and prediction is performed with a default value for t. All preprocessing and prediction steps are done using Python and its prediction library scikit-learn [9].

(a) Highlighting the reference to a figure on gaze. The circle indicates the reader's gaze and is not visible to them.


GaRSIVis: Improving the Predicting of Self-Interruption during Reading using Gaze Data - eyeSee

3 of 5 110%

Figure 2: The Overview page (A) with the readings from which analysts can select a particular reading for cleaning. This reading is shown in the Data Cleansing page (B). Upon returning to the Overview page analysts can run a prediction on the cleaned data and view the Prediction Results page (C). From there analysts can choose to rechunk the data (D) and then rerun the prediction on the new chunk size. This workflow can be repeated.

the fixation count is zero for a length of time we can conclude that this piece of our data is invalid.

Figure 3: Data Cleansing: brush and zoom to inspect data in detail and trim invalid portions.

4.1.3 Prediction Results. The Prediction Results page supports the detailed analysis of the features involved in the prediction (see Figure 4). Invalid data chunks are not shown since they are not used by the predictor. Summary statistics for accuracy, precision, and recall across all reading sessions are shown at the top of the page, as is the current chunk size. There is not enough room to show all of the features for all of the sessions, so each reading session can be collapsed or expanded. When collapsed, only the accuracy, precision, and recall statistics are shown. When the session is expanded, there is also a stepline strip for each of the gaze features. Steplines

(b) Highlighting the figure after navigation

Figure 2: Usage of *eyeSee*: Gaze at reference and navigation to referenced figure.

the API are sent via WebSocket connection to the viewer application where the screen coordinates are translated to locations in the PDF document.

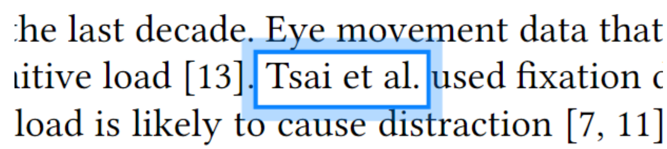
Pilot Study

To better understand the situational benefits of our design, we conducted a pilot study in which we interviewed participants about their reading habits. For each participant, we went over the last 3 papers they read, noting the purpose, their reading workflow, and their use of figures, tables, and citations while reading. We also demonstrated our preliminary prototype and asked for feedback on the system.

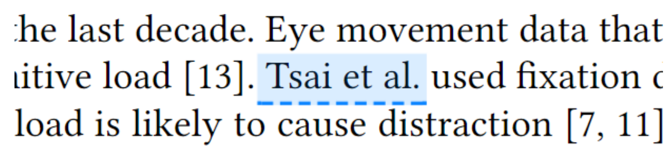
Implications of the Pilot Study

We concluded from the feedback of our pilot study participants that citations are visited less often than we assumed. During regular reading tasks aimed at understanding the paper for an upcoming group discussion or similar activity our participants reported barely looking at any citation at all. In contrast, our participants recalled performing literary research where they almost exclusively focused on related works and citations in a paper. This mode of reading is only employed infrequently and is not the ideal target for *eyeSee*.

Based on their reported reading habits, we concluded however that figures play an important role in the effort to understand a paper. They are usually referenced in at least one section, but rarely positioned next to that section. In many cases the figure can be found on a different page entirely, requiring the reader to scroll back and forth to keep the text and the image in context. Given their important role we decided that references to figures and tables are a more worthwhile target for *eyeSee* and we subsequently shifted our focus from citations to primarily supporting figures. References to figures are now highlighted on gaze



(a) Initial highlighting.



(b) Improved highlighting.

Figure 3: Changes in the highlighting based on feedback from the pilot study.

and the referenced figure is emphasized on navigation, a feature especially important in cases where multiple figures appear at the top of a page.

Furthermore, we decided to improve the highlighting of *eyeSee* based on the feedback. We changed the focused text to have a lightly highlighted background instead of a bright border while keeping the same color scheme. Figure 3 shows the highlighting in detail before and after the design iteration. Removing the border and especially the vertical lines improves the reading flow in cases where the user does not want to trigger the navigation. At the same time, the underline visualization is typical for links and intuitively symbolizes the possible interaction.

Initially, a highlighting would remain for one second after a reference was fixated or until the next reference was highlighted. This was intended to keep the interaction localized to the current reading location and prevent unintended navigation. With our shift to figures however, participants mentioned the importance of a figure not just to one sentence but the entire paragraph or section. Therefore, as another improvement, we changed highlighting to remain until another reference is fixated. This changes allows our users to navigate to a figure referenced earlier in a section without requiring them to find and look at the reference again.

Evaluation

We had 6 participants in total (all male grad students, average age of 26). We originally planned on having 8-12 participants; however, after a clear pattern of results that were not useful in our task comparison, we decided that 6 participants were enough for subjective feedback and we would spend the extra time working on a plan for redesigning a better evaluation.

Each participant completed two reading tasks using the same setup containing an eye tracker (see Figure 4). The first task consisted of reading one section of a software engineering paper, introducing the feasibility of functional magnetic resonance imaging to analyze code comprehension, in a conventional PDF viewer. The eye tracker was installed and visible to participants, but not actively tracking during this task. The second task had the participants read the majority of a short paper written by one of the authors, presenting a visualization tool used to analyze and interactively clean gaze data. This paper was chosen as access to the LaTeX-source allowed us to ensure proper semantic linking in the document markup.

For each reading task, the participant was instructed to read the paper how they would to prepare for a reading group. They could spend as much time as they wanted on reading and were told in advance which sections would be up for discussion afterwards. When the participant indicated

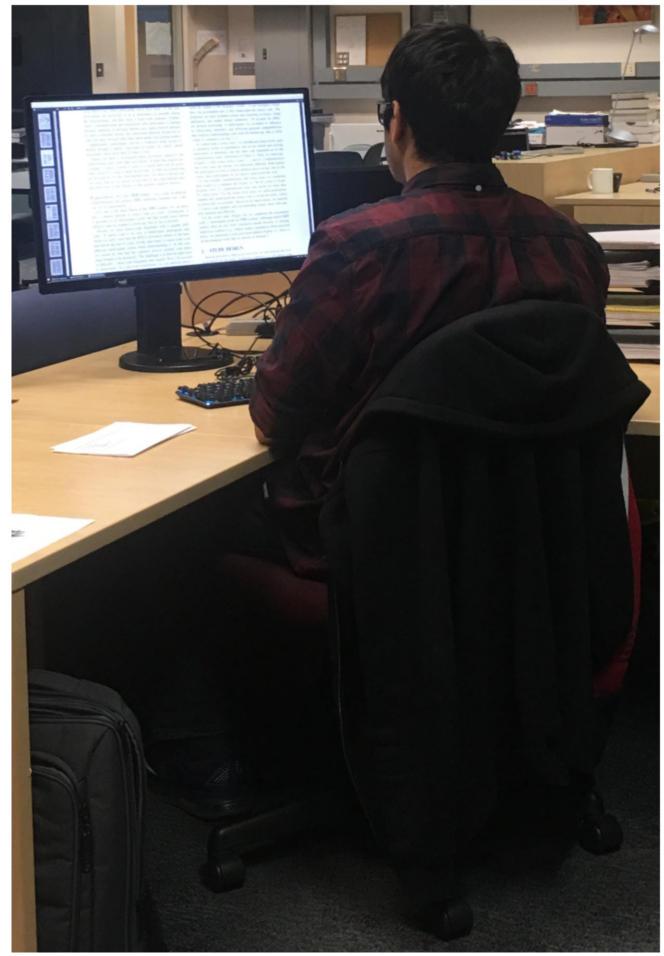


Figure 4: Setup during our evaluation. A participant reading the first document.

they were satisfied with their understanding of the paper, the experimenters would engage in a 4 minute discussion simulating a paper reading group with the participant. After the discussion, the participant would complete a NASA-TLX [2] (task load index) survey to measure the cognitive load of reading the paper and subsequently discussing it.

After both reading tasks were finished, the participant filled out a survey evaluating their impression and subjective opinion of *eyeSee*.

Results

Unfortunately, the cognitive load was not reduced for our participants while using *eyeSee* compared to the standard PDF reader. We realized the flaws in our experimental design too late in the process for this particular project and thus ended the evaluation after 6 participants instead of the intended 8-12 participants. We will revisit our mistakes and potential solutions to avoid these mistakes in the discussion section.

While our objective measurements of success were underwhelming, we received great feedback from participants regarding their subjective experience using the tool. According to a survey conducted at the end of the evaluation (see Table 1), all of our participants clearly understood the intention of the tool and most participants thought the tool would be useful to them while reading PDF-format research papers.

Discussion

Variance between the papers made the NASA TLX significantly less effective as an A-B comparison tool, something we underestimated when originally coming up with our evaluation's design. It is very difficult to determine if two papers are of similar difficulty and this level of difficulty varies for each individual participant as well.

eyeSee is effective for long papers and for maintaining context while reading over a long period of time; however, we could not capture this with our controlled lab study. Our goal for this tool has been to improve natural reading, so we wanted to conduct an evaluation with as much external validity as possible. Unfortunately, we were not able to achieve this validity for a number of reasons. Our participants were not personally invested in reading the paper, so it is difficult to assess whether they read the paper with the same level of care and attention to detail that they would otherwise. It was also difficult to assess whether our tool worked well for reading over a long period of time. Our participants read for about 10 minutes while an actual reading session might last over an hour.

A more effective evaluation may have been possible if we let users use the tool for their own reading over a period of a week. We could have an existing reading group volunteer where half the group uses *eyeSee* and then evaluate the unders This was not possible due to the time constraints of the project and a limited number of eye trackers available.

Participants who had not looked at figures in the first paper, did so in the second paper. One participant in particular said that the only reason he jumped to the figure is because it was easy to get back to his position in the text.

Limitations

As with many new interaction techniques, *eyeSee* requires external hardware not commonly found on laptops or desk computers: the eye tracker. Existing integration of Tobii eye trackers in laptops and further improvements in the technology could lead to a future where the existence of eye tracking devices in everyday devices is as natural as the existence of a webcam. While this vision is promising, the

	P1	P2	P3	P4	P5	P6	avg
I understand the intention of this tool.	5	4	4	5	5	4	4.5
I think this tool is useful.	4	3.5	3	5	5	4	4.1
I would use this tool.	4	3	3	5	5	4	4.0
I think the design was clear.	4	4	3	5	4	5	4.2

Table 1: Survey Results per participant using a 5 point Likert scale.

current state of distribution of eye tracking hardware prevents an easy and large-scale distribution of *eyeSee*.

Our project used the Tobii 4C eye tracker as it is a commercial product intended for a consumer audience and could be realistically distributed to user's in a larger study. Our decision for this device comes at the cost of accuracy. Professional eye trackers can deliver significantly more accurate results, albeit require an even more unnatural setup. Participants mentioned the minor issues with the accuracy of our tool. We are unable to determine if the source is the hardware, the calibration, the general setup, or most likely a combination of those factors.

Utilizing the eye gaze for interaction imposes a decision on users in certain social settings. We evaluated *eyeSee* simulation a reading group discussion. In such a situation, users have to choose between navigation using their gaze and looking at the other attendants of the discussion. Having a look at the document trigger an interaction, even as simple as a highlighting, could be distracting during a discussion.

Future work

To continue the evaluation started with our lab study, future work could retry the setup with more comparable material for the tasks. Options are using two sections of the same paper or using two introduction sections of similar papers for comparable structure of the used sections. Additionally, better randomization of order and assignment to control and treatment are required.

Additionally, we feel that a larger field study would be beneficial to evaluate *eyeSee* as we observed a noticeable feeling of novelty among our participants. This was expressed by participants using the jump functionality to view figures because it existed, not because it was required for their understanding. A longer study period where the understanding of the material is required for intrinsic reasons would reduce this factor. A larger control group could be used to determine if participants rate the understanding of a paper significantly higher when they are using *eyeSee*.

As discussed in section , We decided on navigating to referenced content instead of bringing the content to the reference to preserve the document structure. A study comparing our approach to iReference [7] is required to determine the preferred way of making referenced content easily accessible.

Conclusion

We presented *eyeSee*, a PDF viewer capable of utilizing the user's gaze to enable quick navigation to referenced content. During our iterative design phase, we shifted our focus from exclusively supporting citation to primarily figures and tables, while still including support for references to citations and sections. We also revised the highlighting of references based on feedback from participants of our pilot study to be less distracting during the reading flow while still being noticeable enough to not be missed. The results of the NASA-TLX of our final evaluation did not show any significant improvement. We assume the score to be not entirely representative as the participants reported large differences in the difficulty of the reading task material. Their survey responses and individual feedback are very positive, with multiple participants stating they would happily use *eyeSee* even in its current state.

REFERENCES

1. GitHub Inc. 2018. Electron. (2018). `https://electronjs.org`
2. Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. North-Holland, 139–183. DOI: `https://doi.org/10.1016/S0166-4115(08)62386-9`
3. Jaemin Jo, Bohyoung Kim, and Jinwook Seo. 2015. EyeBookmark: Assisting Recovery from Interruption During Reading. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 2963–2966. DOI: `https://doi.org/10.1145/2702123.2702340`
4. Byungjoo Lee, Olli Savisaari, and Antti Oulasvirta. 2016. Spotlights: Attention-Optimized Highlights for Skim Reading. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 5203–5214. DOI: `https://doi.org/10.1145/2858036.2858299`
5. Mozilla. 2018. PDF.js. (2018). `https://mozilla.github.io/pdf.js`
6. Tobii. 2017. Tobii Eye Tracker 4C. (2017). `https://tobiigaming.com/eye-tracker-4c`
7. Song Xiaohan, Feng Guihuan, and Luo Bin. 2016. iReference: Make the Reference Details Under Your Nose. In *Proceedings of the Fourth International Symposium on Chinese CHI (ChineseCHI2016)*. ACM, New York, NY, USA, Article 4, 5 pages. DOI: `https://doi.org/10.1145/2948708.2948714`