A Crowdsourced Alternative to Eye-tracking for Visualization Understanding

Nam Wook Kim
Harvard SEAS
33 Oxford St.
Cambridge, MA 02138
namwkim@seas.harvard.edu

Aude Oliva
MIT CSAIL
32 Vassar St.
Cambridge, MA 02139
oliva@csail.mit.edu

Zoya Bylinskii
MIT CSAIL
32 Vassar St.
Cambridge, MA 02139
zoya@mit.edu

Krzysztof Z. Gajos
Harvard SEAS
33 Oxford St.
Cambridge, MA 02138
kgajos@seas.harvard.edu

Michelle A. Borkin
Univ. of British Columbia
201-2366 Main Mall
Vancouver, BC, V6T 1Z4,
Canada
borkin@cs.ubc.ca

Hanspeter Pfister
Harvard SEAS
33 Oxford St.
Cambridge, MA 02138
pfister@seas.harvard.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/10.1145/2702613.2732934

Abstract
In this study we investigate the utility of using mouse clicks as an alternative for eye fixations in the context of understanding data visualizations. We developed a crowdsourced study online in which participants were presented with a series of images containing graphs and diagrams and asked to describe them. Each image was blurred so that the participant needed to click to reveal bubbles - small, circular areas of the image at normal resolution. This is similar to having a confined area of focus like the human eye fovea. We compared the bubble click data with the fixation data from a complementary eye-tracking experiment by calculating the similarity between the resulting heatmaps. A high similarity score suggests that our methodology may be a viable crowdsourced alternative to eye-tracking experiments, especially when little to no eye-tracking data is available. This methodology can also be used to complement eye-tracking studies with an additional behavioral measurement, since it is specifically designed to measure which information people consciously choose to examine for understanding visualizations.

Author Keywords
Visualization;eye tracking;visual attention;comprehension;crowdsourcing

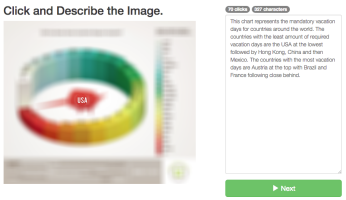

Figure 1: Bubbles experiment user interface.

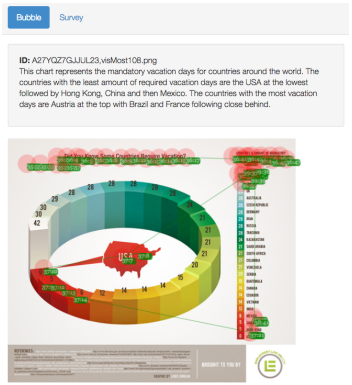

Figure 2: The evaluation user interface displays bubbles, mouse movements, and timestamps over an image stimulus.

ACM Classification Keywords

H.5.1 [Information interfaces and presentation]: Multimedia Information Systems.

Introduction

Eye-tracking is a technique to measure an individual's visual attention, focus, and eye movements. This experimental methodology has proven useful both for human-computer interaction research and for studying the cognitive processes involved in visual information processing, including which visual elements people look at first and spend the most time on [6]. However, collecting accurate eye-tracking data is often expensive and tedious, as quality eye-tracking equipment is costly and requires sophisticated calibrations.

Previous research has investigated how to develop cheaper alternatives for tracking visual attention. Johansen and Hansen compared predicted eye movements based on users' recall as well as designers' guesses to actual eye movements [8]. Masciocchi and Still used a computational saliency model to predict eye fixations on web interfaces [10]. Another popular methodology is the use of a viewing window to track visual attention and visual information acquisition. The Restricted Focus Viewer (RFV) uses this methodology by displaying visual stimuli in a blurred form and revealing a clear focus area ("window") that can be moved around the image in a continuous manner using a mouse [7]. Researchers have employed the RFV to investigate cognitive behaviors of users in diverse contexts such as diagrammatic reasoning and program debugging, and to study the usability of web sites [7, 1, 11]. Similarly, other researchers have explored the relationship between users' mouse movements and eye movements on web pages [3, 9]. Our approach is different from the viewing window approach in that we explicitly collect discretized click data, as each click represents a conscious choice made by the user to reveal a portion of the image. Since the clicks correspond to individual locations of attention, we can directly compare them to eye fixations.

Our approach was specifically inspired by the work of Deng et al. [4] in the computer vision community, in which the bubble paradigm of [5] was used to discover the object/image regions people explicitly choose to use when performing fine-grained object recognition.

This work-in-progress addresses the question of whether we can use the bubble paradigm to discover the regions of a visualization that people examine while trying to gain understanding. In other words, can mouse clicks approximate human fixations in the context of data visualization understanding? We developed a crowdsourced study whereby a user visually scans a blurred visualization, and clicks to reveal small circular areas of the image ("bubbles") in order to describe the visualization in sufficient detail. In order to compare the mouse clicks to the fixation data from a complementary eye-tracking experiment, we calculated the similarity between their respective heatmaps. We find that our methodology allows click data to coarsely approximate fixation data, while our task set-up is designed to explicitly measure which information a user requires for understanding a visualization.

Eye-tracking Experiment

We subsampled 51 visualizations from [2] drawing evenly from infographic, news media, and government publication sources. The visualizations are also all distributed across data encoding type (e.g., bar graph, pie chart, diagram, etc.), topic, and design aesthetic. The visualizations were

Figure 3: An example stimulus: (top) original image, (middle) eye fixations, (bottom) mouse clicks.

all resized to be 1000 pixels on one side, and were shown to participants for 10 seconds at a time, separated by a 0.5 second fixation cross. Participants were told to pay attention to the visualizations because they would later be asked to recall and describe them. Eye-tracking was performed using an SR Research EyeLink1000 with a chin-rest mount 22 inches from a 19 inch CRT monitor with a resolution of 1280x1024 pixels. An average of 16.8 (SD=2.4) participants viewed each of the visualizations. We preprocessed the resulting fixation data by removing the first 2 fixations on each visualization to reduce any viewing biases caused by the experimental set-up.

Bubble Experiment

We conducted our bubble experiments with the same visualizations as the eye-tracking experiments.

The experiments were deployed on Amazon's Mechanical Turk (AMT). Each worker was presented with a sequence of blurred visualization images, and had to describe them in turn (Fig. 1). We used Gaussian blur with a 40-pixel sigma, which we found to be enough to distort the text beyond legibility. The worker could, however, click to reveal full details of small, circular regions ("bubbles"). The images were scaled to have a maximum dimension of 500 pixels per side while maintaining aspect-ratios in order to consistently fit within the AMT task window. We chose the bubble size (16 pixels) to be equivalent to one degree of visual angle in the original eye-tracking experiments (32 pixels in images with twice the resolution). This is to mimic the amount of visual information available in the human fovea.

We posted 17 HITs, each consisting of 3 images randomly selected from the 51 images. To accept one of our HITs, a participant had to have an approval rate of over 95% and live in the United States. A participant was paid $0.05 for each successfully-completed HIT. We had 22 assignments for each HIT, resulting in 22 user data points per image. We also required the text description of an image to be at least 150 characters to ensure that participants completed the task with enough thoroughness.

Requiring workers to write text descriptions was specifically designed to determine if they could accurately report what a visualization depicted (i.e., the main data trends). A good text description indicated that the worker had performed the task correctly and had clicked on the areas of the image necessary for understanding the visualization. Thus we excluded any HITs with poor-quality text descriptions from further analysis. The quality of the text descriptions were evaluated manually by a visualization expert using the evaluation interface (Fig. 2). This procedure resulted in the exclusion of approximately 9% of the data points. We also filtered out HITs whose number of clicks were not within 3 interquartile ranges from the median, resulting in the exclusion of less than 2% of the data points.

Results

To compare the bubble experiment results to the eye-tracking results, we first visually compared heatmaps of bubble clicks and fixation data. Overall, the distributions of clicks overlapped substantially with the eye-movement data (e.g., Fig. 3). This phenomenon was consistent across all image stimuli. Similar to the eye-tracking experiment results, when a participant was asked to describe an image they tended to click on textual elements such as the title, caption, and legend. Workers rarely clicked on non-data elements such as images of humans or human-recognizable objects. Similarly, data elements such as bars, pie chart wedges, or lines did not

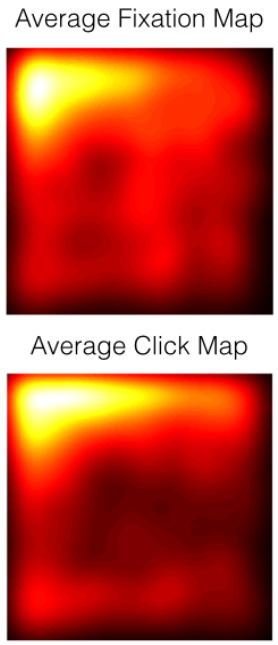

Figure 4: (above) An average taken over all fixation maps and all visualizations. (below) An average taken over all bubble click maps and all visualizations, resized to 500×500.

receive many clicks either. By looking at the textual descriptions, we observe that participants were able to see the trend of data in a blurred image without revealing many details. A worker often clicked on a visually distinctive element surrounded by non-distinctive backgrounds - which aligns with "pop-out" theory. The mouse movement patterns reflect reading patterns (i.e., clicks from top to bottom and left to right). The click patterns also demonstrate that some workers moved back and forth from an image to a text box, likely integrating and progressively gaining understanding of the visualization.

Next, we quantitatively compare the two methodologies. We observe a mean click count of 109.20 (SD=77.75), while eye fixations have a mean count of 40.83 (SD=5.57) across visualizations and participants. As observed in previous research [7], additional human motor control and the blurring of the image are likely the causes of more mouse clicks. Further studies are required to establish a mapping between clicks and fixations (i.e., how many clicks are equivalent to a single fixation?).

To quantify the similarity between the two methodologies, we first computed heatmaps from clicks and fixations using Gaussian blurring. We computed a separate fixation map for each participant in the eye-tracking experiments as well as a separate click map for each participant in the bubble experiments. This allows us to pairwise-compare the fixation and click maps across participants. We use a similarity function based on histogram intersection, yielding a score between 0 and 1. We observe an average pairwise-similarity of 0.58 between fixation maps. This is a measure of the consistency between participants in the eye-tracking experiments. We can compare this to the average similarity between fixation maps and click maps to see how well clicks of participants approximate fixations of other participants. The pairwise-similarity between fixation and click maps is 0.54. The difference between these two modalities (similarity between fixation maps versus similarity between fixation and click maps) is statistically significant ($t(50) = 7.72$, $p<0.01$), pointing to some small, but systematic differences. We also compute a chance baseline called a "permutation control" by using the fixation map from another image to predict the fixations on the current image (a stronger baseline than just random fixations, as it maintains statistical properties of human fixations). The similarity score between fixation maps and permutation controls is 0.33, demonstrating that our click methodology (with a similarity score of 0.54) is significantly above chance at predicting fixations.

Interestingly, if we consider the pairwise-similarity between click maps, we see that the score is 0.64, demonstrating significantly higher consistency between bubble participants in where they click, than between eye-tracking participants in where they look ($t(50) = 8.549$, $p<0.01$). In fact, for 48 of the 51 visualizations, similarity between click maps is higher than similarity between fixation maps, possibly due to click behavior being driven by slower, more conscious choices.

Next, we computed an average fixation map for each visualization, by aggregating and Gaussian blurring all of the eye-tracking participants' fixations on each visualization. We did the same for the click data of the bubble experiment participants to obtain an average click map. Similar to [11], we found high similarities between the average fixation and click maps (AVG=0.71, SD=0.06), indicating that although systematic differences might exist between participants in both modalities, overall, the major trends (averaged over many

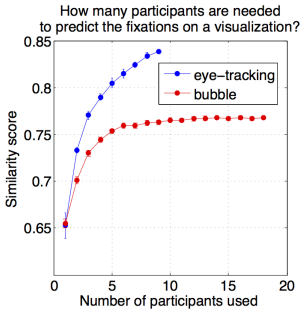

Figure 5: When there is little or no human eye-tracking data available, bubble clicks can help predict ground-truth fixations on visualizations (as compared to a chance baseline with a similarity score of 0.33, see text). However, we also observe systematic differences between the two modalities.

participants) of the fixation data are well captured by the click data. See Fig. 6 for some examples.

Consider the average overall fixation map across all visualizations in comparison to the average overall click map in Fig. 4. This allows us to visualize the systematic differences between the two modalities, e.g., the fixation maps contain a center bias, a natural tendency of observers when free-viewing images [12].

To more carefully examine where the similarities and differences may exist, we consider how much participant data (eye-tracking versus clicks) is required to closely model the fixation patterns of a fixed set of participants. We split the eye-tracking participants in half, and use one half to compute "ground-truth fixation maps" by aggregating the fixations of these participants on each visualization. Next we can measure how well different numbers of eye-tracking participants or bubble participants can approximate the ground-truth fixation maps. We aggregate the fixations (correspondingly clicks) of an increasing number of participants into fixation (click) maps and measure the similarity with the ground-truth fixation maps.

From Fig. 5, we see that one bubble participant does just as well at predicting ground-truth fixation maps as one eye-tracking participant, and that 3-4 bubble participants are as good predictors as 2 eye-tracking participants, after which there is a divergence in predictive power. This divergence is likely the result of the systematic differences between the bubble and eye-tracking modalities. Nevertheless, we see that if human fixation data is unavailable or lacking, then the bubble click data can be used as an alternative measure of attentional patterns.

Discussion

The preliminary analyses presented in this study demonstrate that when averaged over many participants, the clicks generated by our bubble experiments offer coarse approximations to fixation maps on visualizations. Additionally, when very little eye-tracking data is available, the click data can be used for predicting ground-truth fixation patterns on visualizations. All of this provides evidence that our bubble experiments can provide a feasible crowdsourced alternative to eye-tracking.

Nevertheless, eye movements and conscious clicks are behaviors that are driven by different factors, and thus further analyses and experiments are needed to quantify the systematic differences that we observe between the two methodologies. Interestingly, in the task where participants consciously choose what information to examine before clicking, they are more consistent. In the case of raw fixation data, bottom-up features and observer biases might exert a greater influence on what people pay attention to. The higher consistency among participants in the bubble modality is a positive experimental result, indicating that the behavior we measure is more predictable. The task given to participants is to understand visualizations, and we control for this by filtering participant responses. Thus, the click data we obtain highlights the main elements of visualizations that participants explicitly choose to pay attention to in order to understand the visualizations.

Conclusions

This paper provides promising results and insight to justify further research and future improvements of the bubble experiment methodology. Tracking clicks is cheap, non-intrusive, and affords larger-scale experimentation compared to eye-tracking laboratory experiments. Thus,

Figure 6: Two example visualizations: (above) with high consistency and (below) with low consistency between fixation data and click data.

in some cases, click data may serve as a replacement for eye-tracking data, and in others, it may provide a complementary measurement for human attentional patterns. The preliminary results presented here also point to the possibility that click data might measure a more predictable aspect of human behavior. As we observed, the conscious choices people make when clicking leads to greater consistency across participants. In the study presented here, we use these conscious clicks from the bubble experiments to measure what information people explicitly choose to attend to for understanding visualizations. Not only will this lead to a better understanding in the future of what features make a visualization more comprehensible, but will also pave the way for cheaper, easier alternatives to eye-tracking studies.

Acknowledgements

This work is supported by Google and Xerox awards to A.O. Z.B. and M.B. are supported by the Natural Sciences and Engineering Research Council of Canada, and N.K is supported by the Kwanjeong Educational Foundation. The authors would also like to thank Phillip Isola and Jia Deng for their helpful discussions.

References

[1] Bednarik, R., and Tukiainen, M. Effects of display blurring on the behavior of novices and experts during program debugging. In *CHI EA*, ACM (2005), 1204–1207.

[2] Borkin, M. A., Vo, A. A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A., and Pfister, H. What makes a visualization memorable? *TVCG 19*, 12 (2013), 2306–2315.

[3] Chen, M. C., Anderson, J. R., and Sohn, M. H. What can a mouse cursor tell us more?: correlation of eye/mouse movements on web browsing. In *CHI EA*, ACM (2001), 281–282.

[4] Deng, J., Krause, J., and Fei-Fei, L. Fine-grained crowdsourcing for fine-grained recognition. In *CVPR* (2013), 580–587.

[5] Gosselin, F., and Schyns, P. Bubbles: a technique to reveal the use of information in recognition tasks. *VR 41*, 15 (2001), 2261–2271.

[6] Jacob, R. J., and Karn, K. S. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *Mind 2*, 3 (2003), 4.

[7] Jansen, A. R., Blackwell, A. F., and Marriott, K. A tool for tracking visual attention: The restricted focus viewer. *BRM 35*, 1 (2003), 57–69.

[8] Johansen, S. A., and Hansen, J. P. Do we need eye trackers to tell where people look? In *CHI EA*, ACM (2006), 923–928.

[9] Rodden, K., Fu, X., Aula, A., and Spiro, I. Eye-mouse coordination patterns on web search results pages. In *CHI EA*, ACM (2008), 2997–3002.

[10] Still, J. D., and Masciocchi, C. M. A saliency model predicts fixations in web interfaces. In *MDDAUI* (2010), 25.

[11] Tarasewich, P., Pomplun, M., Fillion, S., and Broberg, D. The enhanced restricted focus viewer. *IJHCI 19*, 1 (2005), 35–54.

[12] Tatler, B. The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *JoV 7*, 14 (2007).