I Know What You Want: Using Gaze Metrics to Predict Personal Interest

Jakob Karolus
LMU Munich
Munich, Germany
jakob.karolus@ifi.lmu.de

Patrick Dabbert
University of Stuttgart
Stuttgart, Germany
patrick.dabbert@gmail.com

Paweł W. Woźniak
University of Stuttgart
Stuttgart, Germany
pawel.wozniak@vis.uni-stuttgart.de

ABSTRACT
In daily communications, we often use interpersonal cues – telltale facial expressions and body language – to moderate responses to our conversation partners. While we are able to interpret gaze as a sign of interest or reluctance, conventional user interfaces do not yet possess this possible benefit. In our work, we evaluate to what degree fixation-based gaze metrics can be used to infer a user's personal interest in the displayed content. We report on a study (N=18) where participants were presented with a grid array of different images, whilst being recorded for gaze behavior. Our system calculated a ranking for shown images based on gaze metrics. We found that all metrics are effective indicators of the participants' interest by analyzing their agreement with regard to the system's ranking. In an evaluation in a museum, we found that this translates to in-the-wild scenarios despite environmental constraints, such as limited data accuracy.

CCS Concepts
•Human-centered computing → Empirical studies in ubiquitous and mobile computing;

Author Keywords
Eye-tracking; gaze metrics; personal interest.

INTRODUCTION
In Human-human interaction, gaze plays a vital part when communicating with others [4, 5]. While humans can interpret gaze as a sign of interest or reluctance to a high degree, computer systems struggle in this regard. Hence, these interfaces tend to be cumbersome, adding to the frustration of the user and diminishing user experience. Isolating interesting and worthwhile content is key to engaging the user. Yet, the exact relationship between common eye gaze features (e.g. fixations) and user interest is still unclear.

Recent research has shown that gaze patterns can reveal points of interest for the user, aiding in search queries [1] or photo selection [12]. Most commonly, gaze position is used to augment information within the focal point of the user as depicted by Duchowski et al. [2]. Examples of this idea include higher compression rates of images [7], automatic cropping [11] and selection [12] of pictures; Generating a tourist route with iTourist [8] is possible.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
UIST '18 Adjunct October 14–17, 2018, Berlin, Germany
© 2018 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-5949-8/18/10.
DOI: https://doi.org/10.1145/3266037.3266116

Figure 1. Deployment in a museum exhibition. Text (in German) reads: Please step closer.

In our work, we employ a more generic approach by using a diverse set of images[1] and evaluate different fixation-based gaze metrics based on related work, while also considering saliency-normalized versions of these metrics. Additionally, we not only show each image individually, but in a grid of up to six images. This forces participants to decide on a subset of interesting images.

In a study with 18 participants, we evaluated five different gaze metrics with regard to their correctness of predicting the participants' interest towards shown images and found that all were significantly better than a baseline, which picked images at random. Yet, more computationally complex metrics did not provide better estimations. Additionally, we deployed the system in a museum exhibition (see Fig. 1) and found that simple metrics performed satisfactorily even in a real-world scenario.

METHOD
Based on previous work, we adapted five different metrics based on eye gaze features: fixation count and dwell time, both in an unnormalized and in a saliency-normalized[2] version.

[1] From the PASCAL Visual Object Classes Challenge [3].
[2] Saliency of an image was approximated using Harris corner detection.

Furthermore, we employed a modified version of the iScore [9] metric, which incorporates how many times a users has switched to and from a certain image.

In a within-subject design, we showed a 2x3 grid[3] of images from the *The PASCAL Visual Object Classes Challenge [3]* on a 28 inch LCD display, whilst recording the eye movements of the participants using a Tobii Eye Tracker 4C. Using an I-VT algorithm [10], we extracted fixations and calculated our metrics. The study consisted of four runs, each run cycling through all metrics[4]. Each grid was displayed for a total of 30 seconds, followed by the two images for which the system calculated the highest metric score. To collect participant agreement with regard to the utilized metric, we asked them to rate the system's estimation of their personal point of interest on a five item Likert scale[5].

Each participant saw a total of 144 pictures from the dataset. The image set consisted of multiple categories that were chosen randomly during image selection. No single image was selected twice. Participants were recruited through mailing lists from the University of Stuttgart. A total of 20 ($\bar{x} = 23.3\,y, \sigma = 2.43\,y$) users took part in the study. For analysis, we submitted the data from 18 (15 male) participants.

RESULTS AND DISCUSSION

To evaluate the predictive power of our gaze metrics, we conducted a Friedman test. Participants' agreement ratings were significantly effected by the used gaze metric, $\chi^2(5) = 69.71, p < 0.001$. To test individual metrics, we applied pairwise Wilcoxon tests after Bonferroni correction. All calculated metrics were significantly better than the random baseline condition. Furthermore, there was no significant difference between our five metrics. A visualization of these results can be seen in Figure 2.

Since we used different image categories and positions, those factors might have influenced metric values. Therefore, we additionally tested for effects on the participants agreements with regard to the different image categories and image positions (potential center bias). Here, we found that both factors did not influence metric values.

Our results showcase that all presented metrics perform better than chance at predicting a user's main interest from a set of images. However, we found that computationally more expensive metrics, such as those including saliency normalization and the iScore metric, do not pose any merit. We, therefore, suggest to use a simple metric, e.g. fixation count or dwell time.

While the utilized metrics were suitable to predict the user's interest, it is unclear how such a metric performs when the user is not interested in any of the shown pictures. While the image dataset covers a range of different images and categories, it cannot account for all possibilities.

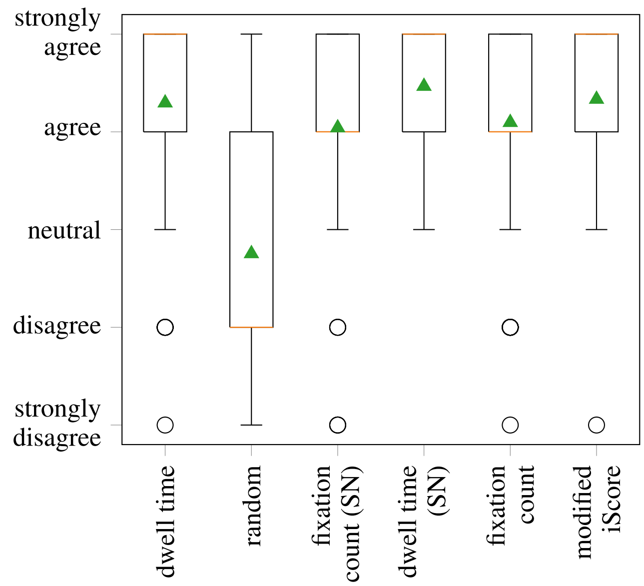

Figure 2. Likert ratings (user agreement with the system's prediction) for the different gaze metrics (SN = saliency normalized). Medians are depicted in orange; means by green triangles. Boxes cover the inter quartile range; whiskers mark data within 1.5 times this range; circles are outliers.

DEPLOYMENT IN A MUSEUM

Genuine lack of interest might negatively influences the presented gaze metrics, due to lack of incentives [6]. To address this issue, we decided to deploy a prototype in a museum exhibit[6] to increase the ecological validity of our study. Here, we used pictures of artefacts presented throughout the exhibit. A visitor would approach a screen mounted with an eye tracker and could interact with the system using their gaze. After presenting a set of artefact images, we showed the user the system's prediction (using dwell time) and asked them to rate whether they agree or disagree with the system.

Our results show that 68% (775 out of 1141) of the visitors agreed with the system's suggestions. Statistical analysis showed that people spent significantly more time looking at the artefact image that the system predicted as being most interesting to the visitor.

CONCLUSION

In our work, we evaluate the feasibility of using gaze metrics to infer a user's interest among multiple images. To that end, we conducted a lab study with a diverse set of images and found that simple metrics, such as fixation count and dwell time of a user's gaze, were effective indicators. Furthermore, we deployed the system in a museum exhibition simulating a real-word use case. Despite the introduction of environmental factors, the collected data still showed that users agreed with the system's prediction.

We envision that the proliferation of ubiquitous sensing techniques, for example eye trackers in public displays or information screens, will enable these systems to tailor their content

[3]This allowed for the images to be shown in the original aspect ratio.
[4]The metric *random* was added as a baseline.
[5]How good is the system's estimation: strongly agree, agree, neutral, disagree, strongly disagree (translated from German).

[6]https://www.mpk.de/archiv-details/events/ohne-schluessel-und-schloss.html

to the user's interest and needs. Since content is diverse, we believe it is necessary to research the effects of influencing factors, such as visual representations and interaction modalities.

REFERENCES

1. Antti Ajanki, David R Hardoon, Samuel Kaski, Kai Puolamäki, and John Shawe-Taylor. 2009. Can eyes reveal interest? Implicit queries from gaze patterns. *User Modeling and User-Adapted Interaction* 19, 4 (2009), 307–339.
2. Andrew T. Duchowski, Nathan Cournia, and Hunter Murphy. 2004. Gaze-Contingent Displays: A Review. *Cyberpsychology & Behavior: The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society* 7, 6 (Dec. 2004), 621–634. DOI: http://dx.doi.org/10.1089/cpb.2004.7.621
3. M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results. (????). http://host.robots.ox.ac.uk/pascal/VOC/voc2007/workshop/index.html
4. Tom Foulsham, Joey T. Cheng, Jessica L. Tracy, Joseph Henrich, and Alan Kingstone. 2010. Gaze Allocation in a Dynamic Situation: Effects of Social Status and Speaking. *Cognition* 117, 3 (Dec. 2010), 319–331. DOI: http://dx.doi.org/10.1016/j.cognition.2010.09.003
5. Lotta Hirvenkari, Johanna Ruusuvuori, Veli-Matti Saarinen, Maari Kivioja, Anssi Peräkylä, and Riitta Hari. 2013. Influence of Turn-Taking in a Two-Person Conversation on the Gaze of a Viewer. *PLoS ONE* 8, 8 (Aug. 2013). DOI: http://dx.doi.org/10.1371/journal.pone.0071569
6. Kenneth Holmqvist, Marcus Nystrom, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost van de Weijer. 2011. *Eye Tracking. A comprehensive guide to methods and measures*. Oxford University Press.
7. Anthony Nguyen, Vinod Chandran, and Sridha Sridharan. 2006. Gaze tracking for region of interest coding in JPEG 2000. *Signal Processing: Image Communication* 21, 5 (2006), 359–377.
8. Pernilla Qvarfordt and Shumin Zhai. 2005a. Conversing with the User Based on Eye-gaze Patterns. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*. ACM, New York, NY, USA, 221–230. DOI: http://dx.doi.org/10.1145/1054972.1055004
9. Pernilla Qvarfordt and Shumin Zhai. 2005b. Conversing with the User Based on Eye-Gaze Patterns. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*. ACM, New York, NY, USA, 221–230. DOI: http://dx.doi.org/10.1145/1054972.1055004
10. Dario D. Salvucci and Joseph H. Goldberg. 2000. Identifying Fixations and Saccades in Eye-Tracking Protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*. ACM, 71–78. DOI: http://dx.doi.org/10.1145/355017.355028
11. Anthony Santella, Maneesh Agrawala, Doug DeCarlo, David Salesin, and Michael Cohen. 2006. Gaze-based Interaction for Semi-automatic Photo Cropping. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. ACM, New York, NY, USA, 771–780. DOI: http://dx.doi.org/10.1145/1124772.1124886
12. Tina Caroline Walber, Ansgar Scherp, and Steffen Staab. 2014. Smart Photo Selection: Interpret Gaze As Personal Interest. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 2065–2074. DOI: http://dx.doi.org/10.1145/2556288.2557025