Reading-based Screenshot Summaries for Supporting Awareness of Desktop Activities

Tilman Dingler[1], Passant El Agroudy[1], Gerd Matheis[2], Albrecht Schmidt[1]
Institute for Visualization and Interactive Systems, University of Stuttgart
[1]{tilman.dingler, passant.el.agroudy, albrecht.schmidt}@vis.uni-stuttgart.de,
[2]matheigd@studi.informatik.uni-stuttgart.de

ABSTRACT

Lifelogging augments people's ability to keep track of their daily activities and helps them create rich archives and foster memory. Information workers perform a lot of their key activities throughout the day on their desktop computers. We argue that activity summaries can be informed by eye-tracking data. Therefore we investigate 3 heuristics to create such summaries based on screenshots to help reconstruct people's work day: a fixed time interval, people's focus of attention as indicated by their eye gaze, and a reading detection algorithm. In a field study with 12 participants who logged their desktop activities for 3 consecutive days we evaluated the usefulness of screenshot summaries based on these heuristics. Our results show the utility of eye tracking data, and more specifically of using reading detection to determine key activities throughout the day to inform the creation of activity summaries that are more relevant and require less time to review.

Keywords

smart summaries; desktop activities; lifelogging; productivity; memory augmentation; recall

Categories and Subject Descriptors

H.5.m. [**Information Interfaces and Presentation (e.g. HCI)**]: Miscellaneous

1. INTRODUCTION

Lifelogging is getting more prevalent and allows people to archive their daily activities: from counting steps to recording images throughout the day, the quantified self movement has spurred a series of research that has shown the benefits the review of such lifelogs has on memory and well-being [1, 14]. Especially information workers leave a rich digital trail that can be used to reconstruct people's work day, which allows for performance assessments, conveying a feeling of productivity, and providing opportunities for reflection. Desktop activities are characterized by the continual switching of attention throughout the day, such as attending to emails or checking news and social networks [17]. Such back-and-forth behavior makes it hard to recapitulate the achievements of the day. Browser histories, for example, help us to keep track of sites visited and articles read. To create activity summaries based on the browser's history alone may not be sufficient though. Taking screenshots in regular intervals gives a more holistic image of people's daily activities [9, 8]. However, this may result in a huge set of images to sift through. OCR (optical character recognition) can be used to extract text and make such archives searchable, but for creating activity summaries based on screenshots we still lack a set of heuristics to determine the key moments for triggering such screenshots.

Using eye tracking data has been found useful for determining people's focus of attention, as well as giving insights into people's cognitive activities [20]. Low-cost eye trackers are becoming increasingly popular and allow us to track such activities more permanently. Furthermore, activities, such as reading, can be derived from looking at eye movements [4]. To augment workers' memory capabilities by creating activity summaries based on a sequence of screenshots throughout the day, we set out to assess heuristics to break down the sheer amount of images captured and increase their relevance by using eye tracking data. Therefore, we augmented people's screens with eye trackers to determine which application windows people focus on, and also to perform reading detection rooted in the assumption that detailed reading hints to key activities and information. With the hypothesis that items read are more relevant for screenshot summaries than random active windows, the contribution of this paper is *twofold*:

1. We explore the utility of using reading detection to capture desktop activities to create screenshot summaries for memory support.

2. We show that reading detection is a useful heuristic effectively reducing image volume and increasing the relevance of reviews.

2. RELATED WORK

We base our work on research from the field of lifelogging, eye tracking, and reading detection. The idea of using personal cameras as 'visual memory prosthetics' [16] started a movement towards capturing and archiving real-life expe-

AH 2016, February 25-27, 2016, Geneva, Switzerland

DOI: http://dx.doi.org/10.1145/2875194.2875224

riences which gave rise to the era of lifelogging. Research around *Microsoft's SenseCam* has demonstrated the application of lifelogging for self and social refection [7, 12], for monitoring health behaviors [18], and for supporting memory [1, 13]. Reviewing lifelogging data has been shown to boost episodic, working and autobiographical memory [21]. Gemmell *et al.* [8], inspired by Bush's idea of creating a system to augment human memory through external tools [3], developed *MyLifeBits* - a lifelogging platform to collect a wide range of data including images, email, browser history, and also analogue content [10]. Such vast activity archives can be browsed, searched, and used for engaging in reflective activities. Isaacs *et al.* [14] showed that using electronic tools to reflect on past memories has a positive effects on memory and increases overall well-being. Le *et al.* [22] created a set of design guidelines for reflective video summaries based on their impact on episodic memory. But the sheer amount of such digital records poses a challenge with regard to how to present the resulting data sets. Prior work has addressed this challenge by collecting additional context data during image capture to derive an image's significance [11] and detect and recognize activities [6, 5, 19].

Buscher *et al.* [2] proposed eye tracking for detecting desktop activities. Eye movements can give insights into people's cognitive processes and attentional focus [20], but especially reading activities can be detected by looking at certain eye movement pattern [4]. Kunze *et al.* [15] were even able to distinguish different document types by classifying distinct eye movement patterns. To enable information workers to archive their work and invite them to reflect on their daily desktop activities, we propose a system that is inspired by related lifelogging research and uses the tracking of eye movements - more specifically eye focus and reading detection - to filter relevant activities throughout the day.

3. USER STUDY

We conducted a user study to evaluate *three* modes for capturing desktop activities:

M1: at fixed **time-intervals**.

M2: eye-focus : focusing on a window with eye gaze.

M3: reading-detection : registering a reading activity.

M1 took a screenshot of the active application window every 2 minutes . For **M2** a screenshot was taken every time the user's eye gaze switched from one application window to another or stayed in one window for 2 minutes, and **M3** took screenshots whenever a reading activity was detected as well as after 2 minutes of continued reading in the same application window.

Two *hypotheses* build the foundation of this study: *(H1)* we can draw a decently accurate picture of people's desktop by recording it through screenshots and *(H2)* we can use eye-tracking, and more precisely reading activities as a heuristic to support the selection of useful and memorable desktop activities.

3.1 Methodology

The study had *three* conditions corresponding to the aforementioned capture modes. In a *repeated-measure design*, each participant spent one day in each condition. The conditions were presented across three consecutive days. At the end of each day, we presented an image summary of their desktop activity for that day, after which participants filled in a questionnaire assessing the review's utility, accuracy and effect on recall.

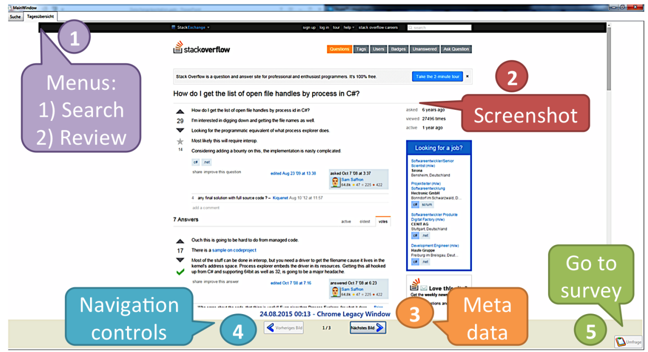

Figure 1: Review screen showing a screenshot including information about time taken and active application window.

3.2 Participants

The study was conducted with 12 participants with a mean age of 28 ($SD = 11.6$) years. 8 males and 4 females participated in the study, all of which were IT professionals or university students except for one retired person. 7 participants indicated to regularly wear glasses or contact lenses, which was taken into account when calibrating the eye-tracker. 9 participants were using external monitors at their work station.

3.3 Apparatus

We augmented the laptops of the participants with *Tobii EyeX* commercial eye trackers and our software developed using the *.NET* framework. The software took screenshots according to the assigned trigger and displayed the sequence of screenshots upon request. It provided a review mode, in which it enabled users to navigate between the screenshots and to search using date, time and custom tags.

We adapted the reading detection algorithm described by [4]. Therefore, we cleaned the data stream provided by the *Tobii EyeX* trackers by clustering incoming data points to eliminate outliers caused by measurement errors. Using these clusters, the algorithm calculates a regression line, which is roughly horizontal for eye movements along a line. This works especially well for detecting regressions, *i.e.* eye movements against the reading direction as it is the case for line breaks. In a pilot study we fine-tuned the algorithm by asking participants to read, skim, and search short texts for certain keywords.

3.4 Procedure

Participants were invited to our lab to initially set up the software on their laptops. We explained the purpose of the study and walked them through the calibration of the eye tracker and the basic software functionality (see Fig.1). Participants could quit the study at any time and pointed out the possibility to temporarily disable the tracking software. After participants had signed the consent form, they were asked to fill in a pre-questionnaire to collect demographics and other information, such as average usage duration of

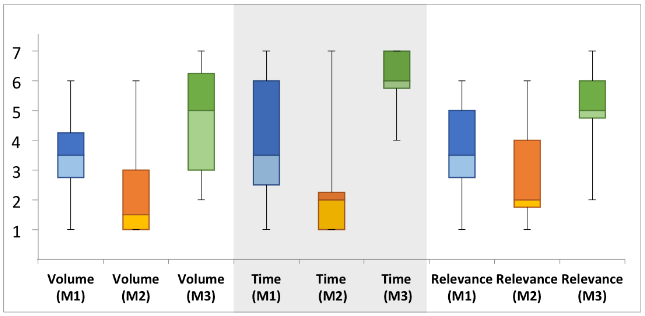

Figure 2: Subjective feedback results on perceived *volume* of review, appropriateness of *time* required, and *relevance* of screenshots for each mode. (X-axis: question per mode, Y-axis: 7-point likert-style rating. 1=Strongly disagree and 7=Strongly agree.)

computers per day and usage of eye glasses. After the setup, we asked participants to perform three reading tasks in order to determine the accuracy of the reading detection algorithm, namely to 1) read one paragraph in detail, 2) skim a document, and 3) search for certain keywords in another document. We asked participants to set up the eye tracker and software once they got home in order to start the experiment the following day. The sequence of conditions was counter-balanced via latin-square. At the end of each day, participants were asked to review the recorded screenshotss and answer the daily questionnaire containing 16 likert-style questions as depicted in table 2. The software triggered an hourly reminder to review the daily activities starting at 5pm of each day to make sure participants completed the review. When participants returned the eye tracker at the end of the three days, we helped them uninstall the software and administered a final questionnaire.

4. RESULTS

Software usage and accuracy.

In the pre-questionnaire participants reported to spend on average 6.3 ($SD = 2.2$) hours per day with their computer. We further logged interactions with our software: on average, participants used the tracking software for 5.6 ($SD = 3$) hours per day. The initial reading tasks were designed to give insights into the tracking accuracy, especially regarding the reading detection. As for reading a paragraph in detail, the algorithm correctly caught 75% of the reading events. When skimming text, reading was detected in 33% of the cases and when searching for certain keywords in 8% (skimming is characterized by fewer and shorter fixations and can involve back and forth eye movements, hence reading is harder to detect).

Utility of collected archives.

Table 1 summarizes the quantitative measurements for the presented reviews per day. The two measurements are: *number of captured images* and *time taken for daily review.*

Number of captured images. The results show that the number of pictures produced was significantly affected by the trigger condition, $F(1.12, 12.28) = 20.63$, $p < 0.001$, $r = 0.8$. *Reading detection based capture* resulted in the smallest number of images. Bonferroni post hoc tests revealed a significant difference in the number of pictures produced between all conditions, namely between time-interval based and eye-focus, $CI_{.95} = -1901.59$ (lower) -311.24 (upper), $p = 0.007$, between time-interval based and reading detection, $CI_{.95} = 47.83$ (lower) 488.51 (upper), $p = 0.017$, and between eye-focus and reading detection, $CI_{.95} = 635.027$ (lower) 2114.14 (upper), $p = 0.001$.

Time taken for daily review. There was no detected statistical difference between the modes, $F(1.20, 13.18) = 0.29$, $p = 0.64$.

Item	Mean value (SD) M1	M2	M3	Diff. amid modes?
Number of captured images	1447.58 (938.5)	341.17 (271.04)	**73.0** (57.57)	**Yes**
Time taken for daily review (*in minutes*)	10.6 (7.4)	13.5 (8.9)	12.0 (16.3)	No

Table 1: Summary of daily review quantitative measures and statistical significance of difference between modes

Table 2 summarizes the qualitative survey data. The median values for each survey question per condition are reported. The question ID reflects its order in the survey. We were able to detect statistical significance indicating the difference between the screen capture modes in **Q2**, **Q3** and **Q8** survey data. For the remaining questions, they revealed certain tendencies. The three questions were further investigated using post hoc analysis with Wilcoxon signed-rank tests after applying Bonferroni corrections. Hence, results were *significant* at $p < 0.017$. *Figure 2* summarizes the survey results for **Q2**, **Q3** and **Q8**. In the following, we report on these questions in detail.

Q2*: I perceived the volume/extent of the review as appropriate.* There was a statistically significant difference in perceived volume appropriateness between modes ($\chi^2(2) = 13.911, p = 0.001$). Median perceived volume levels for the conditions time-interval, eye-focus, and reading-detection were 3.5 (2.25 to 4.75), 1.5 (1 to 3) and 5 (3 to 6.75), respectively. There was a statistically significant reduction in perceived volume in the reading detection vs. eye-focus mode ($Z = -3.078$, $p = 0.002$). The review volume triggered by *reading detection* was perceived as more appropriate.

Q3*: The time taken for the review was appropriate.* There was statistically significant difference in the perceived time appropriateness between the modes ($\chi^2(2) = 15.350, p < 0.001$). Median perceived time taken for the conditions time-interval, eye-focus, and reading-detection were 3.5 (1.5 to 6), 2 (1 to 2.75) and 6 (5.25 to 7), respectively. There was a statistically significant reduction in perceived time taken in the reading detection vs. eye-focus mode ($Z = -2.82$, $p = 0.005$). The review time triggered by *reading detection* was perceived as more appropriate.

Q8*: I found the screenshots presented relevant.* There was statistically significant difference in the perceived relevance

Survey question	Median rating (SD)		
	M1	M2	M3
1) Perceived review time and volume			
Q2. I perceived volume/extent of review as appropriate	**3.5** (1.56)	**1.5** (1.5)	**5** (2.02)
Q3. The time taken for the review was appropriate	**3.5** (2.25)	**2** (2.2)	**6** (1.13)
2) Capture accuracy and content selection			
Q5. The review reflected well my activities today	**5.5** (1.3)	**5.5** (1.78)	**6.5** (2.26)
Q6. The review contained today's important activities	**5.5** (1.82)	**5** (1.85)	**6.5** (2.44)
Q8. I found the screenshots presented relevant	**3.5** (1.61)	**2** (1.56)	**5** (1.38)
Q11. I would have liked previous days' data in the review	**2** (2.08)	**2** (2.1)	**5** (2)
3) Perceived impact on daily routines			
Q1. Today, I have worked very efficiently	**5** (1.5)	**4.5** (1.33)	**5** (1.49)
Q4. I found the review to be useful	**4** (1.4)	**3.5** (2.2)	**5** (2.07)
Q7. The review helped me reconstruct my activities today	**5** (1.56)	**7** (1.85)	**6** (1.83)
Q10. The review helped me recall my activities today	**5** (1.7)	**6** (2.02)	**6** (1.82)
4) Data archiving and privacy concerns			
Q9. I would like to archive this review for future reference	**2.5** (2.28)	**2** (1.7)	**4** (2.3)
Q12. The review contained sensitive content	**2.5** (2.53)	**2** (2.71)	**3** (2.34)
Q13. The review contained data I would rather not archive	**5.5** (2.68)	**6.5** (2.84)	**5** (2.83)
5) Search and retrieval of reviews			
Q14. I made extensive use of the search & filter function	**1** (1.88)	**1** (2)	**1** (2.15)
Q15. I found the search function to be useful	**3.5** (1.93)	**4** (1.94)	**3.5** (2)
Q16. The search function yielded useful results	**4** (2.02)	**4** (2.02)	**3.5** (1.93)

Table 2: Medians of user ratings in 7-point likert-style for each condition. 1=Strongly disagree and 7=Strongly agree. Questions with statistical significance in mode difference are highlighted in green. The highest-rated mode is highlighted in yellow.

of the screenshots based on the mode ($\chi^2(2) = 8.227$, $p = 0.016$). Median perceived screenshot relevance for the triggers time-interval, eye-focus, and reading-detection were 3.5 (2.25 to 5), 2 (1.25 to 4) and 5 (4.25 to 6), respectively. There was a statistically significant reduction in perceived relevance of screenshots in the reading detection vs. eye-focus mode ($Z = -2.862$, $p = 0.004$). The screenshots triggered by reading detection were perceived as more relevant.

5. DISCUSSION

We focused our analysis on time, volume and content relevance, since these are the most critical aspects for getting users to engage in daily reviews. Here we found significant differences in the volume produced by using the proposed heuristics, but also in the subjective ratings of the resulting screenshot summaries. We further grouped the exploration of these heuristics into *five* categories: 1) comparison between quantitative and perceived review time and volume measurements, 2) capture accuracy and perceived effectiveness of content selection in the review, 3) perceived impact of reviewing process on daily routines, 4) data archiving and privacy concerns, and 5) search and retrieval of reviews

Perceived vs. quantitative review time and volume.
Volume and time required to perform reviews were perceived to be more appropriate when triggered by the reading-based screen capturing. This is in line with the quantitative measurements as reading-based capture produced only 5% of the number of images compared to fixed time-interval screenshots. Hence, the time required for a review was perceived to be more appropriate for reading-based collections. However, the time measurement showed no significant difference between the three modes. People tend to spend more time with relevant images rather than simply fast-forwarding irrelevant screenshots. Hence, it can be argued that reading-based screenshots are an effective means for capturing key events bearing more relevance. Lingering on such records allows users to engage in reflection, which has been shown to support long-term memory [14].

Capture accuracy and content selection. Screenshots were rated as most relevant when reading detection was the trigger as compared to time-interval and eye-focus based capture. Users tended to agree that the reviews reflected their daily activities regardless the capture mode. Some participants stated they would like data from previous days to be included in their reviews.

Perceived impact on daily routines. Participants found the reviews generally useful, particularly the reading-detection based reviews. Reviews helped the participants to recall daily activities with a tendency towards preferring eye-focus based captured reviews.

Data archiving and privacy concerns. Participants did not perceive the captured content as overly sensitive. However, participants were consistently agreeing that reviews contained data they would not want archived.

Search and retrieval of reviews. Participants did not make extensive use of search or filter functions. This could be due to the rather short duration of the experiment, so there was no need to search for items from the past. As the archive grows, such functions could prove useful as participants found the functionality and the results useful within their limited usage.

6. CONCLUSION

We set out to explore three different heuristics for triggering desktop screenshots to create desktop activity summaries. Overall, the reading detection trigger produced the smallest number of images while maintaining the highest satisfaction ratings by participants in terms of assigned relevance, perceived volume and time required for the review. Eye-focus alone can already be used to help filter out a significant portion of captured screenshots, which confirms our hypothesis (H2) of using eye-tracking data to detect key activities. With regard to the accuracy of reflecting people's activities based on screenshots (H1) we found that screenshot summaries were well received, although their usage as archives needs further investigation due to privacy implications and the limited duration of the study. Our investigation shows the feasibility of using eye-tracking to augment information workers' capabilities: by identifying and recording key activities, such records can be used to create activity summaries that foster people's memory and sense of achievement. Such tracking will not be limited to pure desktop activities, but can expand to mobile devices as well to create comprehensive activity logs of the day.

7. ACKNOWLEDGMENTS

We acknowledge the financial support of the Future and Emerging Technologies (FET) programme within the 7th Framework Programme for Research of the European Commission, under FET grant number: 612933 (RECALL).

8. REFERENCES

[1] G. Browne, E. Berry, N. Kapur, S. Hodges, G. Smyth, P. Watson, and K. Wood. Sensecam improves memory for recent events and quality of life in a patient with memory retrieval difficulties. *Memory*, 19(7):713–722, 2011.

[2] G. Buscher, A. Dengel, and L. van Elst. Eye movements as implicit relevance feedback. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '08, pages 2991–2996, New York, NY, USA, 2008. ACM.

[3] V. Bush and A. W. M. Think. The atlantic monthly. *As we may think*, 176(1):101–108, 1945.

[4] C. S. Campbell and P. P. Maglio. A robust algorithm for reading detection. In *Proceedings of the 2001 Workshop on Perceptive User Interfaces*, PUI '01, pages 1–7, New York, NY, USA, 2001. ACM.

[5] A. Fathi, A. Farhadi, and J. M. Rehg. Understanding egocentric activities. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 407–414. IEEE, 2011.

[6] A. Fathi, X. Ren, and J. M. Rehg. Learning to recognize objects in egocentric activities. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference On*, pages 3281–3288. IEEE, 2011.

[7] R. Fleck and G. Fitzpatrick. Teachers' and tutors' social reflection around sensecam images. *Int. J. Hum.-Comput. Stud.*, 67(12):1024–1036, Dec. 2009.

[8] J. Gemmell, G. Bell, and R. Lueder. Mylifebits: a personal database for everything. *Communications of the ACM*, 49(1):88–95, 2006.

[9] J. Gemmell, G. Bell, R. Lueder, S. Drucker, and C. Wong. Mylifebits: fulfilling the memex vision. In *Proceedings of the tenth ACM international conference on Multimedia*, pages 235–238. ACM, 2002.

[10] J. Gemmell, L. Williams, K. Wood, R. Lueder, and G. Bell. Passive capture and ensuing issues for a personal lifetime store. In *Proceedings of the the 1st ACM Workshop on Continuous Archival and Retrieval of Personal Experiences*, CARPE'04, pages 48–55, New York, NY, USA, 2004. ACM.

[11] C. Gurrin, G. J. F. Jones, H. Lee, N. O'Hare, A. F. Smeaton, and N. Murphy. Mobile access to personal digital photograph archives. In *Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices & Services*, MobileHCI '05, pages 311–314, New York, NY, USA, 2005. ACM.

[12] R. Harper, D. Randall, N. Smyth, C. Evans, L. Heledd, and R. Moore. The past is a different place: They do things differently there. In *Proceedings of the 7th ACM Conference on Designing Interactive Systems*, DIS '08, pages 271–280, New York, NY, USA, 2008. ACM.

[13] S. Hodges, L. Williams, E. Berry, S. Izadi, J. Srinivasan, A. Butler, G. Smyth, N. Kapur, and K. Wood. Sensecam: A retrospective memory aid. In *UbiComp 2006: Ubiquitous Computing*, pages 177–193. Springer, 2006.

[14] E. Isaacs, A. Konrad, A. Walendowski, T. Lennig, V. Hollis, and S. Whittaker. Echoes from the past: How technology mediated reflection improves well-being. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 1071–1080, New York, NY, USA, 2013. ACM.

[15] K. Kunze, Y. Utsumi, Y. Shiga, K. Kise, and A. Bulling. I know what you are reading: recognition of document types using mobile eye tracking. In *Proceedings of the 17th annual international symposium on International symposium on wearable computers*, pages 113–116. ACM, 2013.

[16] S. Mann. Wearable computing: A first step toward personal imaging. *Computer*, 30(2):25–32, 1997.

[17] G. Mark. Multitasking in the digital age. *Synthesis Lectures On Human-Centered Informatics*, 8(3):1–113, 2015.

[18] G. O'Loughlin, S. J. Cullen, A. McGoldrick, S. O'Connor, R. Blain, S. O'Malley, and G. D. Warrington. Using a wearable camera to increase the accuracy of dietary analysis. *American journal of preventive medicine*, 44(3):297–301, 2013.

[19] H. Pirsiavash and D. Ramanan. Detecting activities of daily living in first-person camera views. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2847–2854. IEEE, 2012.

[20] K. Rayner. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124(3):372, 1998.

[21] A. R. Silva, S. Pinho, L. M. Macedo, and C. J. Moulin. Benefits of sensecam review on neuropsychological test performance. *American Journal of Preventive Medicine*, 44(3):302–307, 2013.

[22] H. Viet Le, S. Clinch, C. Sas, T. Dingler, N. Henze, and N. Davies. Impact of video summary viewing on episodic memory recall - design guidelines for video summarizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '16, New York, NY, USA, 2016. ACM.