Evaluation of a Low-Cost Open-Source Gaze Tracker

Javier San Agustin*
Henrik Skovsgaard†
IT University of Copenhagen

Emilie Mollenbach‡
Loughborough University

Maria Barret§
Martin Tall¶
Dan Witzner Hansen‖
John Paulin Hansen**
IT University of Copenhagen

Abstract

This paper presents a low-cost gaze tracking system that is based on a webcam mounted close to the user's eye. The performance of the gaze tracker was evaluated in an eye-typing task using two different typing applications. Participants could type between 3.56 and 6.78 words per minute, depending on the typing system used. A pilot study to assess the usability of the system was also carried out in the home of a user with severe motor impairments. The user successfully typed on a wall-projected interface using his eye movements.

CR Categories: H.5.2 [Input devices and strategies]: Evaluation/Methodology

Keywords: low cost, gaze interaction, gaze typing, off-the-shelf components, augmentative and alternative communication

1 Introduction

The cost of commercial gaze tracking systems often exceeds \$10.000, preventing access to this technology by many people who need it for daily communication. Using low-cost components benefits users who cannot afford an expensive commercial system, and it gives them a chance to try gaze interaction for the first time without making a big investment. Furthermore, open source solutions might provide a starting point for enthusiasts, who can contribute to an active project and adapt the software to fit specific needs.

Commercial gaze tracking systems have been available for more than 15 years, but the technology is still a high priced niche. A few systems built using off-the-shelf components have been presented. Babcock and Pelz [2004] developed a head-mounted system that uses two cameras mounted on a pair of safety glasses, one to track the eye and the other to record the scene in front of the viewer. However, building this system requires an in-depth understanding of electronics and the software only runs on the Linux operating system, thereby preventing most potential users from assembling their own system. The Opengazer eye tracker uses a web camera to track the user's eyes and is rather simple to build [Zieliński 2007].

*e-mail: javier@itu.dk
†e-mail: hhje@itu.dk
‡e-mail: author@affiliation.com
§e-mail: @affiliation.com
¶e-mail: mtall@itu.dk
‖e-mail: witzner@itu.dk
**e-mail: paulin@itu.dk

Copyright © 2010 by the Association for Computing Machinery, Inc.
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 commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions Dept, ACM Inc., fax +1 (212) 869-0481 or e-mail permissions@acm.org.
ETRA 2010, Austin, TX, March 22 – 24, 2010.
© 2010 ACM 978-1-60558-994-7/10/0003 \$10.00

However, this project is in a very early stage of development, and has not been active since 2007.

In this paper we introduce the ITU Gaze Tracker (available at http://www.gazegroup.org/downloads/23-gazetracker), a low-cost system that was released as open source in Spring 2009 [San Agustin et al. 2009], and we present an evaluation of a head-mounted setup that employs a webcam with built-in infrared light. Such setup does not require extra light sources, and therefore the system is very easy to build. However, the system is not head-pose invariant. The evaluation was carried out using two different eye-typing applications designed to handle a noisy input, GazeTalk and StarGazer. A pilot study with a person in the latest stages of ALS was also conducted.

2 The ITU Gaze Tracker

In order to provide an accessible gaze tracking system, the ITU Gaze Tracker was developed. The design considerations and the hardware and software are presented in this section.

2.1 Design Considerations

The approach taken to develop the ITU Gaze Tracker aimed to provide a fully accessible system, with a set of design requirements:

1. The gaze tracker should be robust and accurate enough to work with at least one gaze-communication system.
2. Use of low-cost off-the-shelf components. The hardware employed should be available in any electronics store or at online shops to allow for easy acquisition and replacement. Furthermore, no hardware modifications should be needed.
3. The user should be given flexibility to place the different components (camera, infrared lights, computer display) at various locations to fit specific needs. For instance, mounting the display on a wheel chair, moving it to a table or having it accessible in bed should make no difference.
4. Open-source software. Developing an efficient gaze tracker from low-cost components is a huge endeavor. No single group of developers is likely to come up with *the* ultimate solution. Open source allows anybody to improve and modify the source code to fit specific needs.

2.2 Hardware

Standard webcams usually have a low resolution and a broad field of view, but by placing the camera close to the user's eye we can obtain images of sufficient quality. Figure 1 shows the gaze tracker in two setups, (a) mounted on a piece of inexpensive balsa wood that the user bites on in order to fix the webcam in front of the eye, and (b) mounted on a flexible arm (e.g., an office lamp).

The setup displayed in Figure 1 employs a Sandberg Nightvision 2 webcam, which costs around \$20. The camera is equipped with 6 built-in infrared LEDs that create a dark pupil effect. The resolution is 640×480 pixels and the frame rate is 30 fps. The weight of the

(a) (b)

Figure 1: *Mounting solutions of the ITU Gaze Tracker: (a) on a piece of inexpensive balsa wood; (b) on the arm of an office lamp.*

camera is 200g. The amount of light shone into the eyes is not considered to be harmful [Mulvey et al.].

2.3 Software

2.3.1 Software Architecture

The gaze tracking software is developed in C#, and employs OpenCV for image processing. There are three main components: (1) The gaze tracking library, which implements all the methods to control a tracker such as extracting eye features, run a calibration procedure, estimate the gaze coordinates and detect the type of eye movement (i.e., fixation or saccade); (2) The camera class, responsible for initializing a generic camera and grabbing images that are then processed by the gaze-tracking library; and (3) The user interface, which provides the communication with the gaze-tracking library to set up the different parameters of the system.

2.3.2 Image Processing

The ITU Gaze Tracker supports tracking the pupil center and one or two corneal reflections (i.e., glints). The software has been developed to work with low-cost webcams with built-in infrared light. However, initial support for remote setups that make use of more expensive video cameras is also included in the software.

The experiments presented in this paper make use of the webcam-based setup shown in Figure 1. In this configuration, the software employs the pupil center to estimate gaze. The pupil center is calculated by thresholding the image and extracting points in the contour between pupil and iris. These points are then fitted to an ellipse using a RANSAC procedure that eliminates possible outliers. Figure 2 shows a screenshot of the application with the setup window open. The center of the pupil is estimated and a cross-hair is drawn on the image. The picture box shows a processed image of the eye and gives an indication of the quality of the tracking.

Figure 2: *Interface of the ITU Gaze Tracker showing a real time processed picture of the user's eye and the setup window where gaze tracking parameters can be set.*

2.3.3 Gaze Estimation

Our flexibility requirements introduce a high uncertainty in the locations of the camera and screen with respect to the user. Therefore, the gaze estimation technique employed must be as generic as possible, and should not assume any specific configuration of the hardware components. The gaze estimation method implemented is an interpolation technique that uses a pair of second-order polynomials (one for each screen axis) [Morimoto et al. 1999]. A 12-point calibration procedure allows to calculate the coefficients that map the center of the pupil to the gaze coordinates on the screen. The calibration procedure takes approximately 20 seconds.

The system was demonstrated at the CHI 2009 Conference. A total of 28 delegates tried out the system mounted on a piece of balsa wood. 22 had normal vision or wore contact lenses, and 6 wore glasses. All 28 were able to get a successful calibration and control the cursor with their eyes.

3 Gaze Typing with the ITU GazeTracker

Text input is a very important task for people who use gaze to communicate. The design of typing systems is one of the most well researched areas within the field of gaze interaction [Majaranta and Räihä 2002]. Gaze typing systems are often evaluated experimentally with standardized performance measurements such as words per minute (WPM) and error rates.

The two eye typing applications used in this experiment have different approaches to handle noise. GazeTalk [Hansen et al. 2004] has 10 large buttons with a hierarchical organization of the alphabet. The user activates the keys by looking at them for a set duration (i.e., dwell time activation). StarGazer [Hansen et al. 2008] is a 3D interface that uses pan/zoom to enlarge objects of interest, thereby making them easier to select. The full alphabet is visible in the initial state of the system. The user looks at a desired character and zooms towards it. Selections are made when passing through the character.

3.1 Experimental Procedure

Seven participants (2 female and 5 male, ranging from 20 to 35 years) participated in the experiment. All participants had normal or corrected-to-normal vision. Four of them had prior experience with gaze interaction. The typing interfaces (GazeTalk and StarGazer) were presented on a 17" monitor with a resolution of 1280×1024. Participants used the webcam-based gaze tracker to interact with the typing applications, and, in the conditions where it was required, an optical mouse to perform selections. Participants sat approximately 60 cm away from the screen. Word prediction was deactivated on both systems but the built-in letter prediction in GazeTalk showed the six most likely letters on the main interface.

Participants were asked to type a total of 60 sentences (with an average of 25 characters each) as quickly and accurately as possible. The sentences were Danish translations of the phrase set by Mackenzie and Soukoreff [2003]. No participant entered the same sentence more than once. After completing a sentence, the next was presented. A break was offered after each block of five sentences. Before each block the participants were encouraged to adjust the activation time. At the end of each block the subjects were asked to evaluate the physical strain on eyes, mouth/jaw, face, neck and head on a seven-point scale.

The within-subjects experiment tested the following four conditions repeated over three sessions: StarGazer with zoom activation, StarGazer with mouse click activation, GazeTalk with dwell time activation and GazeTalk with mouse click activation. A total of 420 sentences were typed (7 participants $\times$ 3 sessions $\times$ 4 conditions $\times$ 5 sentences). The conditions were counterbalanced using a balanced Latin square design to neutralize learning effects.

The dependent variables were typing speed, measured in words per minute (WPM), and accuracy, measured as the minimum string distance (MSD). The MSD expresses the minimum number of basic actions (insertions, deletions and substitutions) between the produced sentence and the target sentence [Levenshtein 1966].

3.2 Results

Sentences with an MSD value above 6 (49 out of the total of 420) were considered outliers and removed prior to analysis. One-way ANOVAs were conducted to compare the effects of the different typing methods (StarGazer zoom, StarGazer mouse, GazeTalk dwell, GazeTalk mouse).

The results obtained for WPM and MSD are shown in Figure 3. The grand mean for WPM was 5.02. There was a significant effect of typing method on WPM, $F(3, 18) = 30.67$, $p < 0.05$. Both of the GazeTalk conditions showed significantly higher WPM than the StarGazer conditions in an LSD post-hoc analysis. GazeTalk with click activation had a mean WPM = 6.78 (SD = 2.72), and a mean WPM = 5.62 (SD = 1.98) with dwell activation. StarGazer with click had a mean WPM = 3.91 (SD = 1.63), and a mean WPM = 3.56 (SD = 1.58) with zoom activation. The grand mean for MSD errors was 1.61% of all characters typed. Typing method did not have a significant effect on MSD error, $F(3, 18) = 0.56$, $p > 0.05$.

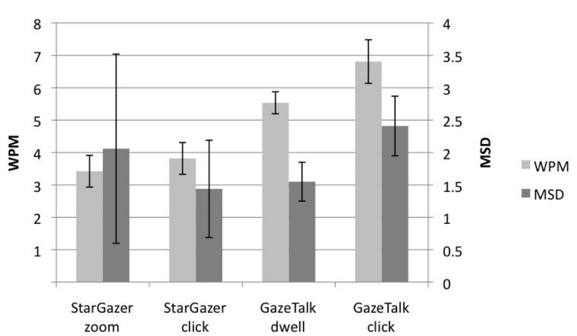

Figure 3: *Words per minute and MSD error for each typing method. Error bars show the standard error of the mean.*

The average settings chosen by the participants in the last session were 1375 ms for StarGazer with zoom activation and 403 ms for GazeTalk with dwell time activation.

Participants reported biting the piece of balsa wood to be quite unpleasant after the first session, with an average value of 2.4 on a scale from 1 (“most uncomfortable”) to 7 (“no strain at all”). However, by session 3 this rating had gone up to 4.6, indicating only moderate discomfort.

3.3 Discussion of the Results

Gaze typing with the low-cost gaze tracker was possible in both of the noise tolerant interfaces. The relatively small amount of errors in the final sentences indicates that participants controlled the applications properly and were able to correct their mistakes adequately. This suggests that a low-cost, webcam-based gaze tracker holds potential for gaze interaction with interfaces that have rather large targets, as is the case of GazeTalk and StarGazer.

In our experiment the final typing speed for GazeTalk with dwell time activation was 6.56 WPM in the third session. This compares to the typing speed of 6.26 WPM for equally trained subjects in the experiment by Hansen et al. [2004] who used a commercial gaze tracking system. Presumably, the noise tolerance given by GazeTalk’s big buttons allows to remove any possible difference between the commercial gaze trackers and the low-cost ITU Gaze Tracker. Actually, in the present experiment the participants were a bit faster than in the previous experiment from 2004 where a commercial system was used. The slight difference may be explained by a small difference in dwell time: in the experiment by Hansen et al. [2004] this was set to 500 ms while the subjects in our experiment adjusted it to an average of 403 ms. Text-entry speed in the present experiment is much lower than in other gaze typing studies, where typing speeds of 15 WPM are reported (e.g., [Tuisku et al. 2008]). However, the low-cost gaze tracker would not work well with the small on-screen keyboard buttons used in the other experiments, which were all done with commercial systems.

After calibration of the mouth-mounted gaze tracker, even the slightest head movement introduced a large offset relative to the screen. Since the camera is mounted on the user’s head, the two will move together. Therefore, a relative movement of the head/camera with respect to the screen will take place under head movements. As a result, when the user maintains his gaze on the same point on the screen and moves the head, the estimated gaze coordinates will be affected. The participants in our experiments got used to this behavior rather quickly and started to take advantage of the ability to adjust for offsets by head movements.

Carrying the camera and biting the balsa wood felt uncomfortable, at least in the beginning. Eventually, it just came to rest in the bite marks and then it did not feel strenuous. However, it prevents the user from talking and it looks quite strange. Hence, in the current version we do not consider mouth mounting to be feasible in social settings outside the private home or the experimental lab. However, reduced size and weight of future webcams might make it possible to mount them firmly on for example eyeglass frames.

4 Case Study

The main objective of the case study was to see if the low-cost gaze tracker was flexible enough to accommodate individual setups. The system was set up in the living room of Birger Bergmann Jeppesen, a person who has been living with ALS for more than 10 years and has been gaze typing for more than 5 years. The interface was projected onto a wall and the camera was placed on an arm close to his eye. Birger’s wheelchair was placed in front of the wall, approximately 3 meters away. Figure 4 shows the setup.

The trial revealed that completing a calibration was consistently achievable and subsequently Birger was able to interact with GazeTalk and produce text. The typing speed was very low - less than a word per minute. However, it must be noted that although Birger can be considered an expert user of GazeTalk, he is in a late stage of ALS and does not have good control of his eye movements anymore. Furthermore, he no longer has the ability to blink, and the LEDs mounted on the camera would dry out his eyes after a while. Therefore, they needed to be moistened by the helper a few times.

When asked after the trial what he thought the system could be used for, he answered “For poor people”, referring to the fact that this approach would be beneficial for users who cannot afford an expensive system. His wife, who was present during the trial, was very enthusiastic about the setup. She liked that her husband could now be part of conversations around the living room table and everybody would be able to see what he was typing, thereby eliminating the need for a synthetic voice. Furthermore, the disassociation

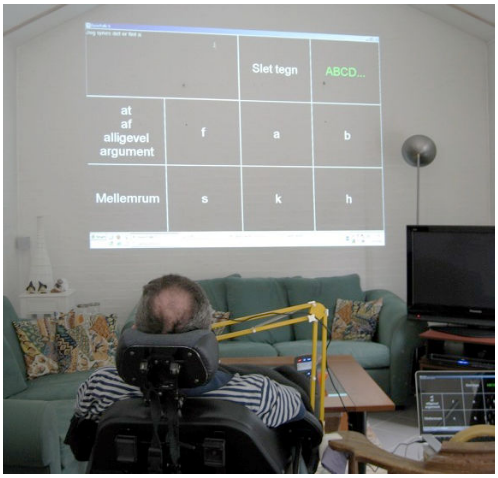

Figure 4: *The typing application (GazeTalk) is projected onto the wall. Birger sits on his wheelchair approximately 3 meters away.*

of the camera with respect to the screen gives freedom in regard to screen format, thereby allowing a wall projection or a regular monitor to be controlled with the eyes.

This initial user trial highlighted some issues regarding the setup. First, care is needed when placing the camera in a fixed position so that it does not block the field of view of the user. Second, the camera was sensitive to light; drawing the blinds helped increase the stability. There is still a need for more robustness and flexibility but this case study demonstrated the potential of low-cost gaze trackers.

5 Conclusions and Future Work

A low-cost, open-source gaze tracker has been described and evaluated in this paper. The system offers a high flexibility regarding both the software, which can be easily extended by modifying the source code, and the hardware, which can be composed of different cameras and light sources.

The results obtained in the evaluation indicate that a gaze tracking system built from inexpensive components can be robust and accurate enough to make gaze typing possible for people who cannot afford the more expensive commercial systems. The user interface needs to have big buttons, which will reduce the productivity, but typing speeds at more than 5 words per minute are found to be achievable when the typing interface is tolerant to noise.

The main limitation of the webcam-based system is the lack of head-pose invariance. However, some participants would often use this as a feature and compensate for calibration offsets by making small adjustments of their head position at each selection. Nevertheless, in normal everyday usage it might be difficult to keep the head still for long periods of time.

One situation where the head-pose invariance does not play a role is when the user is not able to move the body, for example due to a locked-in syndrome. The webcam can be mounted on the user's wheelchair or bed and placed close to the eye. Once the calibration procedure is completed, the accuracy is maintained due to the immobility of the user. The pilot study described in this paper indicates that this is a feasible solution; however, a more thorough experiment will be carried out in the busy environment of a hospital or care home to see if the user will in fact not move and to see if it is realistic to assume that the camera stays fixed.

As of January 2009, more than 3500 people have downloaded the software, showing a high interest in the technology. Furthermore, more than 200 people have taken part in on-line discussions on how to use the software, and the community is collaborating in the development of the new version. Thanks to the simplicity of the hardware setup and usage of low-cost components, we expect more enthusiasts to become interested in the project.

Acknowledgments

This work was supported by the COGAIN European Network of Excellence, funded under the FP6/IST program of the European Commission. Thanks to the anonymous people who have contributed to the further development of the source code.

References

BABCOCK, J. S., AND PELZ, J. B. 2004. Building a lightweight eyetracking headgear. In *Proceedings of the 2004 symposium on Eye tracking research & applications*, ACM, San Antonio, Texas, 109–114.

HANSEN, J. P., TORNING, K., JOHANSEN, A. S., ITOH, K., AND AOKI, H. 2004. Gaze typing compared with input by head and hand. In *Proceedings of the 2004 symposium on Eye tracking research & applications*, ACM, San Antonio, Texas, 131–138.

HANSEN, D. W., SKOVSGAARD, H. H. T., HANSEN, J. P., AND MOLLENBACH, E. 2008. Noise tolerant selection by gaze-controlled pan and zoom in 3D. In *Proceedings of the 2008 symposium on Eye tracking research & applications*, ACM, Savannah, Georgia, 205–212.

LEVENSHTEIN, V. 1966. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady 10*, 6, 707–710.

MACKENZIE, I. S., AND SOUKOREFF, R. W. 2003. Phrase sets for evaluating text entry techniques. In *CHI '03 extended abstracts on Human factors in computing systems*, ACM, Ft. Lauderdale, Florida, USA, 754–755.

MAJARANTA, P., AND RÄIHÄ, K. 2002. Twenty years of eye typing: systems and design issues. In *Proceedings of the 2002 symposium on Eye tracking research & applications*, ACM, New Orleans, Louisiana, 15–22.

MORIMOTO, C., KOONS, D., AMIT, A., FLICKNER, M., AND ZHAI, S. 1999. Keeping an eye for HCI. In *Proceedings of the XII Symposium on Computer Graphics and Image Processing, 1999.*, 171–176.

MULVEY, F., VILLANUEVA, A., SLINEY, D., LANGE, R., COTMORE, S., AND DONEGAN, M. Exploration of safety issues in eyetracking. In *COGAIN Deliverable 5.4. http://www.cogain.org/w/images/c/c8/COGAIN-D5.4.pdf.*

SAN AGUSTIN, J., SKOVSGAARD, H., HANSEN, J. P., AND HANSEN, D. W. 2009. Low-cost gaze interaction: ready to deliver the promises. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, ACM, Boston, MA, USA, 4453–4458.

TUISKU, O., MAJARANTA, P., ISOKOSKI, P., AND RÄIHÄ, K. 2008. Now dasher! dash away!: longitudinal study of fast text entry by eye gaze. In *Proceedings of the 2008 symposium on Eye tracking research & applications*, ACM, Savannah, Georgia.

ZIELIŃSKI, P., 2007. Opengazer: opensource gaze tracker for ordinary webcams. http://www.inference.phy.cam.ac.uk/opengazer/.