3D Model-Based Gaze Estimation in Natural Reading: a Systematic Error Correction Procedure based on Annotated Texts

Andrea Mazzei* Shahram Eivazi† Youri Marko‡ Frederic Kaplan§ Pierre Dillenbourg¶
School of Computer and Communication Sciences, École Polytechnique Fédérale de Lausanne
School of Computing, University of Eastern Finland

Abstract

Studying natural reading and its underlying attention processes requires devices that are able to provide precise measurements of gaze without rendering the reading activity unnatural. In this paper we propose an eye tracking system that can be used to conduct analyses of reading behavior in low constrained experimental settings. The system is designed for dual-camera-based head-mounted eye trackers and allows free head movements and note taking. The system is composed of three different modules. First, a 3D model-based gaze estimation method computes the reader's gaze trajectory. Second, a document image retrieval algorithm is used to recognize document pages and extract annotations. Third, a systematic error correction procedure is used to post-calibrate the system parameters and compensate for spatial drifts. The validation results show that the proposed method is capable of extracting reliable gaze data when reading in low constrained experimental conditions.

CR Categories: H.5.2 [Information Interfaces and Presentation]: User Interfaces—Input devices and strategies;

Keywords: 3D Gaze Estimation, Systematic Error Correction

1 Introduction and Motivation

Eye movements in reading have been intensely studied on visual display units (VDU), using monitor-integrated eye tracking technology, typically accompanied with chin- and head-rest. Most of these investigations were characterized by experimental controls and constraints, necessary to accurately measure the task performance. Designing low-constrained experimental setups to study eye movements in natural reading on printed instructional material still remains a challenge. Reliable gaze data can be acquired by limiting participants movements or the printed material manipulability. However, depending on the experimental focus, these impositions may negatively affect the concentration of the subjects. In this paper, we describe an eye tracking method to study eye movements in natural reading in low constrained experimental conditions. The system: 1) is capable of accurately measuring the readers' gaze points, 2) does not require commercial hardware setups for acquiring gaze data, 3) does not require experimental settings that may interfere with reading activity and comprehension. Participants are allowed to read on paper, to freely move and to take notes while reading. Under low-constrained experimental conditions, the gaze data extraction suffers from systematic errors and experimental inconvenients. We here propose a method to address this problem.

*andrea.mazzei@epfl.ch
†shahram.eivazi@uef.fi
‡youri.marko@me.com
§frederic.kaplan@epfl.ch
¶pierre.dillenbourg@epfl.ch

2 Related Works and Our Contribution

The majority of the eye tracking studies of reading have been characterized by rigid and rigorous experimental constraints that, on one side, limit the participants' movements and range of actions, and, on the other, improve the accuracy of data collection. In eye tracking studies based on display-units participants were usually encouraged to remain still at a fixed distance from the monitor while performing the task [Schotter et al. 2012; Rayner et al. 2010; Pieters and Wedel 2004]. In such experimental setting chin- and headrests were usually introduced in order to minimize head movements. A similar setup was used to study spatial attention shifts on print advertisements [Pieters and Wedel 2004].

Eye movements in reading were also studied through a comparative analysis between visual display units and traditional paper [Velde and Grunau 2003]. In the paper reading condition, the authors attached the printed instructional material on a vertical monitor. All subjects were invited to use a head holder during the experiment. A similar experimental control was introduced by [Siegenthaler et al. 2011] in an analysis focusing on reading patterns and visual fatigue between e-reader displays and printed paper.

Researchers also have used head mounted eye tracking setups for reading research. This type of setup allows free movements during the experimental task. However, in many studies, experimenters still felt appropriate to introduce a chin-rest and other forms of control in the experimental setting [Rayner et al. 2008; Holsanova et al. 2009]. A head-mounted eye tracker coupled with a computer monitor was used by [Holsanova et al. 2009] to investigate how people look at the textual and pictorial contents of advertisements. The distance between the participant's head and the monitor was controlled.

Newspapers reading was studied using a head-mounted eye tracker [Holsanova et al. 2009]. In that study a newspaper was attached to the table in order to maintain the same position recorded during the calibration procedure. This setting gave some freedom of movements to the user but limited the manipulation of the text.

When studying natural reading and its underlying cognitive processes from an eye tracking perspective, most of the experimental constraints should be relaxed in order to avoid any interference with note taking practices and reader's concentration. However, conducting an experiment under these conditions may introduce systematic errors into the data collection process. In addition, the experimental procedure should be calibration free and resistant to unexpected violations of experimental rules, which may require a recalibration of the eye tracking system. These experimental conditions may increase the risk of exposure to data inaccuracies that may change during the experimental time, and that can be quantified as drift of gaze-point locations [Weigle and Banks 2008].

There are various techniques to post correct gaze data. For exam-

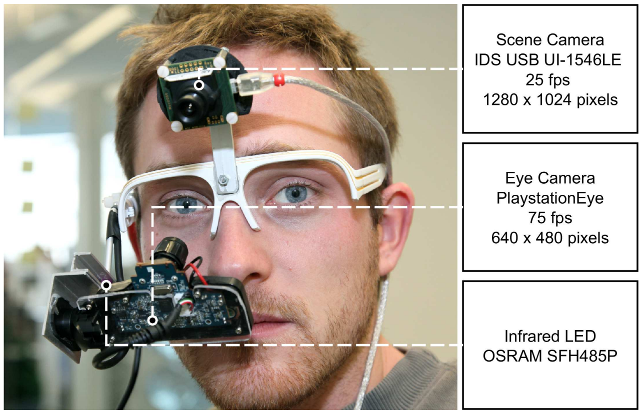


Figure 1: *Self-made head-mounted eye tracker*

ple, [Hornof and Halverson 2002] focused on the assumption that experimenters can reliably define where the user's visual attention is focused on at specific moments of the experimental task. Then the system can automatically shift the fixation points to the true locations. A similar procedure was designed to exploit the regularity of the distribution of reading-related fixation points over the text, to semi-automatically compensate for vertical errors [Hyrskykari 2006]. Completely automatic techniques for systematic error correction have been investigated by [Zhang and Hornof 2011]. Their procedure preliminary measures the disparity between the fixation locations and their nearest areas of interest and then estimates the magnitude and orientation of the correction.

In this paper our contribution is twofold. First, we describe a 3D model-based gaze estimation method that -applies to head-mounted dual camera setups- can be used to study natural reading and note taking with low experimental control. Under these conditions, systematic errors and unexpected experimental events could affect the data collection process. Therefore, our second contribution is a post-calibration procedure, that can be used after the experiment, to correct for spatial drifts of the gaze pattern.

3 Apparatus

Reading and note-taking are activities that require certain changes in the body and head postures as well as manipulation of the printed text. In our work we realized a self-made head mounted setup to meet these requirements (see Figure 1). Our method applies to any dual-camera based head mounted setup, equipped with an IR camera taking close-ups of the eye-pupil and a second camera pointing at the printed texts in front of the reader.

4 Method

The proposed eye tracker provides a video stream containing close-ups of the reader's eye and another video stream containing the printed documents in front of the reader. Given these two video streams, we have developed a processing pipeline (see Figure 2) for: 1) reconstructing the 3D dynamics of the reading activity 2) estimating the point of gaze with respect to the printed stimuli, 3) correcting potential spatial drift that could derive from a wrong estimation of the system parameters. The system is organized in two phases. During the experiment it records the two video streams provided by the eye tracker. After the experiment it post-process the videos in order to extract the gaze data. The following sections describe the modules of the proposed system.

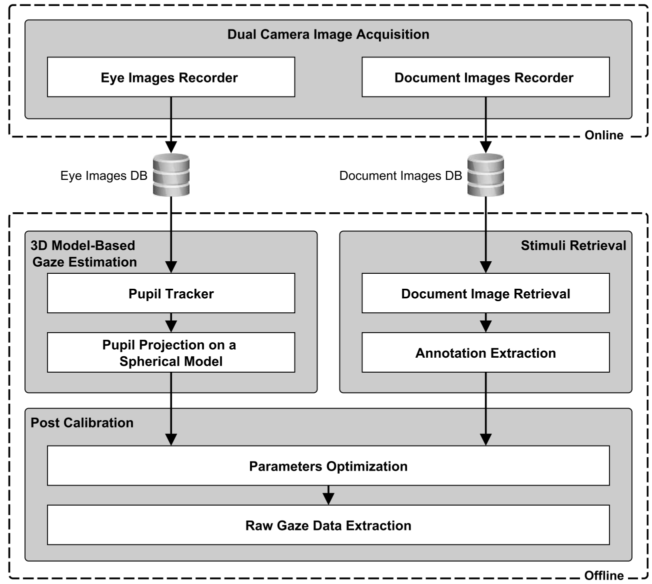


Figure 2: *Processing pipeline.*

4.1 3D Model-Based Gaze Estimation

We employed a feature-based method for gaze detection, based on a spherical modeling of the eyeball (see Figure 3). The image is grabbed by a single camera, under a single IR Led source. The image results into a sharp close-up of the reader's eye under the dark pupil effect. Detecting and localizing the pupil is a common feature that is typically tracked in the video-based eye tracking setups. After detecting the eye-pupil [Agustin et al. 2010], the system reprojects the pupil centroid onto a plausible spherical model of the eyeball [Wang et al. 2005] (see Figure 3).

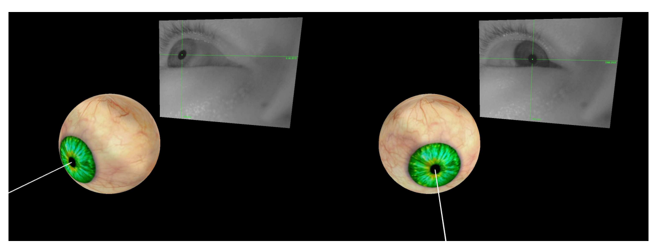


Figure 3: *Eye-pupil tracker and 3D gaze estimation.*

4.2 Stimuli Retrieval

Two tasks are performed to extract the gaze data on printed documents and spatially map them to the contained information units. First, the document(s) in front of the reader are recognized and their 3D position is estimated. Second, the annotations are extracted and their morphological features are computed (see Figure 4). A document page contained in an image can be recognized by overlaying fiducial markers on it or by simply processing the pattern of the text. Fiducial markers are graphical artifacts, characterized by high contrast patterns, designed to be easily recognized by machines. We discarded any fiducial marker-based approach for recognizing document images for two reasons. First, extraneous high contrast artifacts may affect spatial attention and introduce cognitive biases. Second, to achieve robust optical recognition and tracking, many

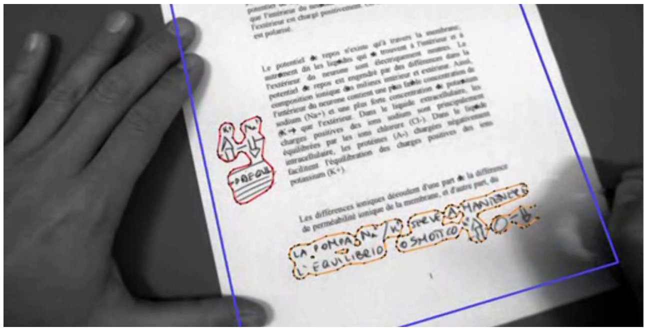


Figure 4: *Document image retrieval and pose estimation (blue quadrilateral). Annotation extraction (two color contours).*

fiducial markers should be printed onto the page margin, which, consequently, could reduce the available space for taking notes. Therefore, we opted for a marker-less algorithm for document image recognition [Nakai et al. 2006]. The method can robustly recognize and 3D track textual patterns in a video stream. On top of this method a connected component analysis is performed in order to segment the handwritten annotations, by classifying each information unit into handwritten or machine printed [Zheng et al. 2004].

4.3 Geometrical Framework and Post Calibration

Figure 5 shows the geometrical model of our head-mounted eye tracking system. There are two reference systems. The global reference system O_G is centered on the scene camera. The O_{EYE} reference system is centered on the eye camera. In our geometrical framework some parameters have to be post-calibrated in order to 3D reconstruct the reading session and extract the raw gaze data. Ten parameters are estimated for each page read by each user.

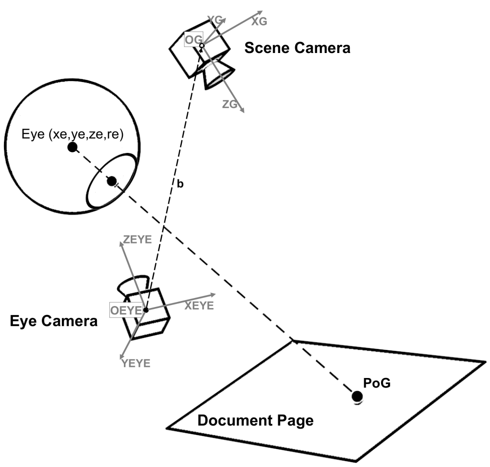


Figure 5: *A geometrical model of reader-page.*

- $P_e(x_e, y_e, z_e)$ and r_e: the relative position of the reader's eye with respect to the coordinate system O_{EYE} and the radius.
- $b = [R|t]$: the homogeneous transformation matrix from the O_{EYE} to the O_G coordinate system.

Estimating these ten parameters is a problem of ten-dimensional bounded search with fitness function $C(x) = C(x_1, x_2, ..., x_{10})$ that can be approached through random optimization techniques. To launch the optimization procedure, the experimenter has to make safe associations between the uncalibrated fixation points with their expected location. For this task we elaborated a graphical user interface, conceptually inspired by the one proposed by [Noris et al. 2011]. First, using the interface, the experimenter can visualize the uncalibrated fixation points; they may appear spatially drifted because of an incorrect estimation of the system parameters (see Figure 6(a)). Second, the experimenter can click on one or more fixation points (they get surrounded by concentric black circumferences) and associate them with their expected location (connected by a segment ending with a black filled circle) (see Figure6(b)). The success of this procedure relies on the ability of the experimenter to recognize the overall distribution of the shifted fixation points with respect to the stimuli appearance. Finally, the procedure perturbs iteratively all the parameters, within their confidence interval, in order to find a value of the fitness function below a predetermined threshold τ. The fitness function measures the distance between the iteratively recalculated fixation locations and their expected position. A typical result is shown in Figure 6(c).

5 Results

We conducted an empirical evaluation of the system with a calibration sheet composed of 32 dots. A grid of 4 by 4 dots was used to estimate the system parameters. An additional set of 16 randomly positioned dots was used to measure the accuracy and precision of our method. Participants were asked to look at all the dots for 3 seconds each. No postural constraint was imposed. System accuracy was measured as the average distance between the dot and the measured gaze position. System precision was measured as the average standard deviation of the dot-related fixations. We measured a precision of 0.87 degrees and an accuracy of 1.01 degrees at an average distance of 50 cm from the printed stimuli. We also studied the convergence of the procedure for the parameters estimation. In Figure 7, we show the convergence of the accuracy of the extracted fixations in function of the number of iterations (in log_{10} scale). The data correspond to the optimization performed on the eye-gaze data collected for a preselected annotated page. Two external experimenters were recruited and instructed to conduct the experimental test. Four associative conditions were considered: 4, 8, 12, and 16 random associations of uncalibrated fixation points with their expected position. For each of these four conditions the experimenters repeated the calibration procedure 10 times. First, from a preliminary graphical analysis of the convergence, it can be concluded that after 1000 iterations there is an overall accuracy of the estimated parameters below 1 degrees of visual angle. By using a general purpose computer the computational time of this procedure is about 16 seconds. We finally observed that the experimenters produced 64% associations involving annotations and 36% involving textual elements. This enforced our intuition that post correction procedures could be shaped around the experimental setting.

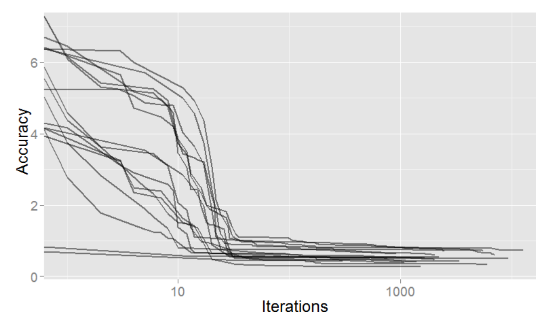
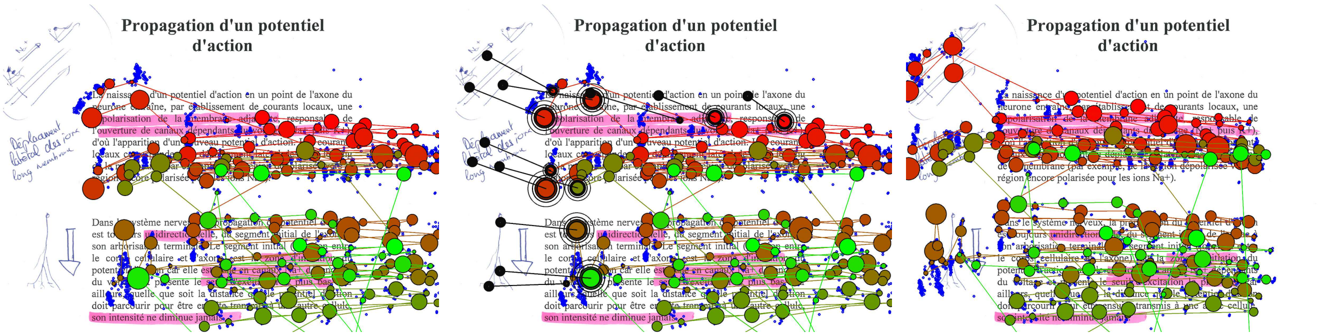


Figure 7: *System parameters convergence.*



(a) Before Optimization (b) First Optimization (c) After Optimization

Figure 6: *An interface to associate drifted fixations (surrounded by black circumferences) and their expected position (black filled circles).*

6 Limitations and Future Work

The future work of our system will derive from its limitations.

First, although major experimental constraints can be relaxed in our solution, we introduced a head mounted eye tracker to the experimental procedure itself. The potential influence of these devices on reading comprehension should be addressed through formal hypothesis testing.

Second, in the systematic error correction procedure, the experimenter has to identify safe associations between drifted fixation and their expected position. The quality of the estimation of the system parameters relies on whether the experimenter manages to make correct associations. We observed that the majority of these associations were involving annotations. This result could be a manifestation of the "isolation effect": the experimenter tends to focus on the information components that are easely distinguisheable. Performing the same procedure for annotation free texts is expected to be more difficult or require more iterations.

Third, the procedure has to be repeated for each page, sequentially read during the experimental task. We can speculate that this granular page-based post-calibration procedure can compensate for changes of system parameters. Potentially, this procedure could also absorb incremental effects of gaze displacement that incrementally increase over time, during the same experimental run. However this procedure can demand a considerable amount of time to the experimenter regarding the extraction of the experimental data. As future work we intend to automatize this procedure by detecting the magnitude and orientation of spatial drifts in the fixation distribution with respect to the closest areas of interest.

Fourth, we observed that our experimenters took advantage of produced annotations in the post-calibration procedure. This preliminary result suggests that procedures for post correcting eye gaze data could be shaped around the experimental task.

References

AGUSTIN, J. S., SKOVSGAARD, H., MLLENBACH, E., BARRET, M., TALL, M., HANSEN, D. W., AND HANSEN, J. 2010. Evaluation of a low-cost open-source gaze tracker. In *ETRA*, ACM, 77–80.

HOLSANOVA, J., HOLMBERG, N., AND HOLMQVIST, K. 2009. Reading information graphics: The role of spatial contiguity and dual attentional guidance. *Applied Cognitive Psychology 23*, 9, 1215–1226.

HORNOF, A. J., AND HALVERSON, T. 2002. Cleaning up systematic error in eye-tracking data by using required fixation locations. *BRM 34*, 4, 592–604.

HYRSKYKARI, A. 2006. Utilizing eye movements: Overcoming inaccuracy while tracking the focus of attention during reading. *Computers in human behavior 22*, 4, 657–671.

NAKAI, T., KISE, K., AND IWAMURA, M. 2006. Use of affine invariants in locally likely arrangement hashing for camera-based document image retrieval. In *DAS '06*, Springer, Springer, 541–552.

NORIS, B., KELLER, J. B., AND BILLARD, A. 2011. A wearable gaze tracking system for children in unconstrained environments. *CVIU 115*, 4, 476–486.

PIETERS, R., AND WEDEL, M. 2004. Attention capture and transfer in advertising: Brand, pictorial, and text-size effects. *Journal of Marketing*, 36–50.

RAYNER, K., MILLER, B., AND ROTELLO, C. M. 2008. Eye movements when looking at print advertisements: the goal of the viewer matters. *Applied Cognitive Psychology 22*, 5, 697–707.

RAYNER, K., SLATTERY, T. J., AND BÉLANGER, N. N. 2010. Eye movements, the perceptual span, and reading speed. *Psychonomic bulletin & review 17*, 6, 834–839.

SCHOTTER, E. R., ANGELE, B., AND RAYNER, K. 2012. Parafoveal processing in reading. *AP&P 74*, 1, 5–35.

SIEGENTHALER, E., WURTZ, P., BERGAMIN, P., , AND GRONER, R. 2011. Comparing reading processes on e-ink displays and print. *Displays 32*, 5, 268–273.

VELDE, C., AND GRUNAU, M. 2003. Tracking eye movements while reading: Printing press versus the cathode ray tube. *Perception*, 1297–1326.

WANG, J.-G., SUNG, E., AND VENKATESWARLU, R. 2005. Estimating the eye gaze from one eye. *CVIU 98*, 1, 83–103.

WEIGLE, C., AND BANKS, D. 2008. A comparison of the perceptual benefits of linear perspective and physically-based illumination for display of dense 3d streamtubes. *TVCG 14*, 6, 1723–1730.

ZHANG, Y., AND HORNOF, A. J. 2011. Mode-of-disparities error correction of eye-tracking data. *BRM 43*, 3, 834–842.

ZHENG, Y., LI, H., AND DOERMANN, D. 2004. Machine printed text and handwriting identification in noisy document images. *TPAMI 26*, 3, 337–353.