A Robust Neural Network Approach for Determining 3D Gaze Position

Mei Xiao, Tyler Garaas, Marc Pomplun

Department of Computer Science, University of Massachusetts at Boston
100 Morrissey Boulevard, Boston, MA 02125-3393, U.S.A.
Email: meixiao@cs.umb.edu, tgaraas@cs.umb.edu, marc@cs.umb.edu

Abstract. This paper presents a neural-network based calibration method for measuring the 3D eye-gaze position in a stereoscopic display. The method is based on depth computation through local non-linear regression and interpolation using a parameterized self-organizing map (PSOM). A novel, efficient 3D calibration method for this application is introduced. We report an experiment in which we compare the performance of the neural-network based method with a simple geometric method. The results show that the PSOM-based method provides significantly better accuracy of measurement than the geometric method.

Keywords: eye tracking, 3D eye gaze, neural network, 3D calibration, stereoscopic 3D

1 Introduction

Using eye tracking for the investigation of visual attention has become increasingly popular during the last few decades. Nevertheless, only a small number of eye-tracking studies have employed three-dimensional displays, although such displays would closely resemble our natural visual environment. The main reason for the paucity of research using three-dimensional displays is the problem of computing a participant's current 3D gaze position based on the measured binocular gaze angles. Determining the depth of the gaze position requires very precise measurement of a subjects' vergence, i.e., the angle between their viewing axes. However, the best video-based systems reach an absolute precision of approximately 0.3° to 0.5° of visual angle source (e.g., Stampe, 1993), which may deteriorate over time due to small shifts of the headset.

Binocular video-based eye trackers compute this pupil-to-gaze mapping separately for each eye, and the calibration marker only appears on the two-dimensional screen. While this approach is well suited for 2D on-screen gaze tracking, it is less appropriate for measuring 3D gaze positions. Clearly, for 3D tracking the calibration should include marker positions at various depths within the 3D stimulus presentation area. Furthermore, since determining 3D gaze positions requires the combined gaze angle information from both eyes, the pupil-to-gaze mapping should be based on this combined information instead of using separate mappings for each eye. Such a binocular mapping, however, would be even more complex than the monocular one

for the 2D case and may produce large measurement errors due to inaccurate interpolation.

In a previous study, the parameterized self-organizing map (PSOM, see Ritter, 1993), which is a fast-learning variant of Kohonen's self-organizing maps (SOMs, see Kohonen, 1990) was applied to the 3D gaze-tracking task (Essig, Pomplun, & Ritter, 2006). The PSOM received the binocular gaze data obtained by a 3×3×3 calibration as its input and learned the mapping of binocular gaze positions on the screen to 3D gaze positions. While this approach was able to greatly reduce the error in 3D gaze-position measurement compared to geometrical, 2D calibration-based methods, subsequent experiments (e.g., Garaas, Xiao, & Pomplun, 2006) revealed a shortcoming of this method. It was found that in situations when no drift correction can be performed for durations longer than about 20 to 30 seconds, the small errors accumulating in the binocular gaze positions can cause substantial deviation in the computed 3D gaze position. The reason for this effect is that the PSOM recursively approximates the most likely 3D gaze coordinates for a given pair of binocular gaze positions as a general mapping from $\mathbf{R}^4$ to $\mathbf{R}^3$ without considering the underlying geometry at all. Consequently, small deviations in measurement, especially those affecting the vergence angle, may not only lead to large error in the computation of the z-coordinate (depth), but can also considerably affect precision in the x- and y-coordinates.

To tackle this problem, we devised a new approach that both considers the underlying geometry of the eye-tracking setup and uses a PSOM for the interpolation of calibration data. Moreover, we introduced a novel calibration method yielding a large amount of data within a reasonable amount of time.

2 The 3D Calibration Procedure

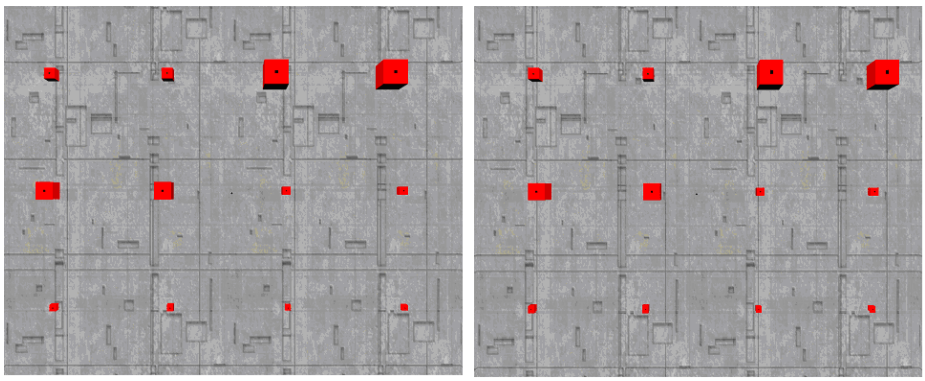


Fig. 1: Stereo image-pair showing a sample screen used for measuring the precision of the neural network and geometric methods. If you cross your visual axes to fuse the two images into one, objects should be perceived as floating in 3D stereo.

As shown in Figure 1, the full 19-inch screen, divided into 12 sections, was used to display the virtual 3D scene. During the calibration procedure, however, only one red

cube at a time was shown. It appeared in the center of every section and moved back and forth once, spanning the depth of the 37 cm (width) × 30 cm (height) × 27 cm (depth) cuboid within which 3D gaze position was to be measured. Every red cube had a black dot (fixation marker) in the center of its front plane. The subjects were instructed to look at the black dot and keep track of it while the red cube moved along its path. This procedure was conducted for every section of the screen. It allowed us to collect a large number of binocular gaze-position samples across the depth dimension of the tracking cuboid.

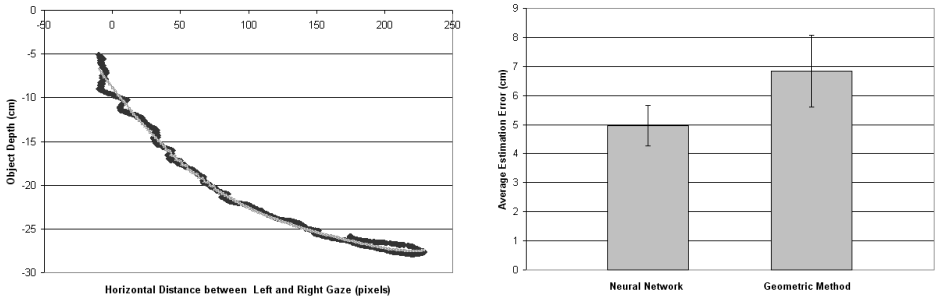
3 Description of the Calculation Method

Since the greatest challenge in the measurement of 3D gaze position is the computation of the z-coordinate (depth), we will focus on the computation of this variable. In an ideal experimental setup, gaze depth only depends on the vergence angle. However, in a real setup there are systematic distortions and noise in the measurement. While our approach assumes that locally, i.e., for similar gaze angles, depth entirely depends on vergence, it maintains that for different gaze angles the mapping from vergence to depth may vary. To account for this, we use non-linear regression on the calibration data to obtain an individual vergence-to-depth mapping for each of the 12 calibration gaze angles. Our algorithm then uses a PSOM to interpolate between these 12 positions so that vergence can be translated into gaze depth for any given horizontal and vertical gaze angles.

In order to determine a subject's vergence angle, we use the standard 2D on-screen gaze-position measurement provided by most standard eye-tracking systems. Greater vergence, i.e., more pronounced crossing of the visual axes, is then indicated by a greater distance $d = s_l - s_r$ between the horizontal on-screen gaze positions s_l and s_r measured for the left and the right eye, respectively. Our regression function for computing gaze depth z in each of the 12 local regressions was chosen to account for the underlying physical geometry:

$$z(d) = -v \cdot \frac{d}{e+d} + c,$$

where the variables v, e, and, c are determined in such a way that they minimize the MSE between the gaze samples obtained during calibration and the function $z(d)$ (see Figure 2a). If v is set to the distance between the subject and the screen, e is set to the distance between the subject's eyes, and c is set to 0, then the above formula turns into the ideal geometric computation of gaze depth. The adaptation of these three variables allows the algorithm to account for variations in the vergence-to-depth mapping for different gaze angles. The variables v, e, and c are interpolated across the range of relevant gaze angles using a 4 x 3 PSOM (see Ritter, 1993). We tested the improvement in 3D gaze-measurement accuracy over the standard 2D geometrical computation in an empirical study described in the following section.



a b

Fig. 2: (a) Object depth as a function of horizontal distance of the binocular eye gaze positions (dark line) and the regression function (bright line). (b) Average error in depth estimation by the PSOM neural network and the geometric method.

4 Empirical Evaluation of the Method

Participants
Ten students from the University of Massachusetts at Boston were tested individually. All participants had normal or corrected-to-normal vision. They were paid for their participation.

Apparatus
Eye movements were recorded with the SR Research Ltd. EyeLink-II system, which operated at a sampling rate of 500 Hz and measured the subjects' gaze position with any average error of less than 0.5 degrees of visual angle. The stimuli were presented on a 19'' DTI autostereoscopic (using no glasses) 3D display with 1280 x 1024 pixels resolution and 60 Hz refresh rate.

Materials and Procedure
At the start of each experimental session, the 3D calibration procedure was performed. The subject was instructed to focus on a fixation marker in the center of the screen to do a drift correction. Then the subject was asked to focus on the black marker on the front plane of the red cube, which was shown in one of the 12 sections of the screen, and its virtual depth was aligned with the physical screen. The cube moved from its initial position towards the subject, and thereby subtended a visual angle increasing from 1.5 to 7 degrees. When it reached its maximum size, it moved back on the same path. This procedure was performed once per screen section.

After the calibration, sample data for the evaluation of our method were collected. The subject was asked to focus on the black dot on the center of the screen again to do the drift correction. Now the screen showed 12 cubes at a time at random positions and depths. While such a testing set was displayed, the subject had to focus sequentially on the black rectangle on each of the red cubes and click the cube using a

mouse pointer to indicate the selection of the object. The object then disappeared. This procedure was performed 11 times.

Results
One of the subjects' data were excluded from analysis, because the subject did not follow the instructions correctly. The data of the remaining nine subjects showed that the PSOM neural-network method gave us a significantly better estimation of the objects' depth (average error of 4.96 cm) than using only the geometric calculation (6.85 cm) in such a virtual 3D environment, $t(8) = 2.863$, $p < 0.05$ (see Figure 2b).

5 Discussion

As predicted, the present neural-network based method of 3D gaze-position estimation outperformed the purely geometric solution by adapting to each user's binocular gaze characteristics and compensating for system-related measurement distortions. By incorporating the underlying physical geometry of the experimental setup, the new method also overcame a problem observed with a previous neural-network approach (Essig, Pomplun, & Ritter, 2006). The latter approach computed an abstract mapping of binocular gaze angles onto 3D gaze positions, which made it very sensitive to small 2D measurement errors, as they are commonly introduced by headset shifts. The current approach, slightly constrained by assumptions about the geometry of the experimental setup, is more robust and yields reasonable accuracy even if noise is introduced. The ability to estimate a user's gaze depth with a reliable precision of about 5 cm is a step toward useful 3D gaze selection in human-computer interfaces. Three-dimensional displays would allow more effective interaction, for example, when robots or cameras need to be remotely controlled by human operators.

References

1. Essig, K., Pomplun, M. & Ritter, H. (2006). A neural network for 3D gaze recording with binocular eye trackers. International Journal of Parallel, Emergent, and Distributed Systems, 21 (2), 79-95.
2. Garaas, T., Xiao, M. & Pomplun, M. (2006). Implicit and Explicit learning as it relates to machine vision systems. Research poster presented at the 2006 ACM SIGGRAPH Conference, Boston, Massachusetts.
3. Kohonen, T. (1990). The self-organizing map. Proceedings of IEEE, 78, 1464-1480.
4. Ritter, H. (1993). Parametrized self-organizing maps. ICANN93-Proceedings, Berlin: Springer, pp. 568-577.
5. Stampe, D. (1993). Heuristic filtering and reliable calibration methods for video-based pupil tracking systems. Behavior Research Methods, Instruments, & Computers, 25, 137-142.