Calibration-free eye tracking by reconstruction of the pupil ellipse in 3D space

Stefan Kohlbecher*
Stanislavs Bardins†
Klaus Bartl‡
Erich Schneider§
Ludwig-Maximilians-University Munich

Tony Poitschke¶
Markus Ablassmeier‖
Technical University Munich

Abstract

Most video-based eye trackers require a calibration procedure before measurement onset. In this work a stereo approach is presented that yields the position and orientation of the pupil in 3D space. This is achieved by analyzing the pupil images of two calibrated cameras and by a subsequent closed-form stereo reconstruction of the original pupil surface. Under the assumption that the gaze-vector is parallel to the pupil normal vector, the line of sight can be calculated without the need for the usual calibration that requires the user to fixate targets with known spatial locations.

Keywords: eye, gaze, pupil, stereo, tracking

1 Introduction

In general, modern video-based eye trackers map the position of the pupil in the image plane to the gaze-vector. This means that the system has to be calibrated by fixating a number of points every time prior to mounting the device, which can be a tedious task. Therefore a calibration-free operation would be a meaningful improvement.

Another problem with eye trackers is the impact of slippage. If the pupil in the camera image plane moves, it is not clear if a change in the line of sight occurred, or merely a translation of the whole eye with respect to the camera. This problem is normally addressed by additionally tracking infrared reflexions on the cornea, which roughly keep their position on eye movements but follow camera movements.

In this work a different approach is used to address these problems. With the help of two calibrated cameras the pupil is reconstructed in 3D space using a closed mathematical framework. Then, gaze is given by the pupil normal. If the position of the camera rig is known, this approach works without the need for user calibration, because the position and orientation of the pupil ellipse is explicitly known. This is also useful for compensating slippage. If the position and orientation of the eye tracker is being tracked in the world coordinate system, the gaze vector can be transformed to this coordinate system, too.

Previous work in this area has been done by [Wang et al. 2005]. They only used one camera though, which means they had to deal with ambiguities. Furthermore the limbus radius had to be calibrated to get true 3D gaze vectors. Both problems do not arise in a stereo camera approach. Another advantage of a stereo system is that it works with projections of ellipses, too (as opposed to circles). A system with a stereo camera that tracks the user through moveable mirrors has been proposed by [Beymer and Flickner 2003], however this work uses a model fitting algorithm and not a closed mathematical framework.

In the following section the algorithms used in this work are described. Section 3 shows the demonstrator that has been set up by incorporating the stereo reconstruction algorithm into a head mounted eye tracker (*EyeSeeCam* [eye 2007]) [Boening et al. 2006; Dera et al. 2006]. Section 4 shows the results of a simple experiment that has been conducted with this demonstrator.

2 Methods

This section presents the mathematical methods. First, the stereo camera calibration process is denoted. Then, the algorithm used for the stereo reconstruction of the pupil ellipse is given.

2.1 Calibration of the stereo rig

One prerequisite is that the parameters of the stereo camera setup must be known. These parameters can be obtained by a camera calibration procedure. We used the Camera calibration toolbox for MATLAB which is available on the internet [mat 2007]. The camera parameters have to be determined for each camera separately by taking snapshots of a checkerboard calibration pattern from various angles and distances with both cameras at the same time. An iterative algorithm based on [Zhang 1999] is used to calculate the intrinsic and extrinsic parameters of both cameras as well as the extrinsic parameters of the stereo rig. This process has to be performed only once after system assembly and no further user calibrations are required after this step.

2.2 Stereo reconstruction

The image of a pupil in the image plane is an ellipse. By laying a cone f_a through the camera center and the pupil image alone, one cannot determine the 3D position and orientation of the pupil, because many pupils are projected onto the same conic in the image plane. The situation changes if there is another projection of the same pupil on a second image plane. Now we can ray another cone f_b through the second camera center and pupil image. The intersection between the two cones then defines the pupil.

By reconstructing the original pupil ellipse from both projections, we can get position, size and orientation of the pupil in 3D space. A closed-form solution to this problem has been proposed previously [De Ma 1993], however, the solution has never been applied to eye trackers before.

Figure 1 shows the image planes, the camera centers (o_1, o_2), the cones (f_a, f_b) and their intersection, which is the pupil. The pupil is in the plane defined by the vectors x_w and y_w. The projections of

*e-mail: skohlbecher@nefo.med.uni-muenchen.de
†e-mail: sbardins@nefo.med.uni-muenchen.de
‡e-mail: kbartl@nefo.med.uni-muenchen.de
§e-mail: eschneider@nefo.med.uni-muenchen.de
¶e-mail: poitschke@tum.de
‖e-mail: ablassmeier@tum.de

ETRA 2008, Savannah, Georgia, March 26–28, 2008.
© 2008 ACM 978-1-59593-982-1/08/0003 $5.00

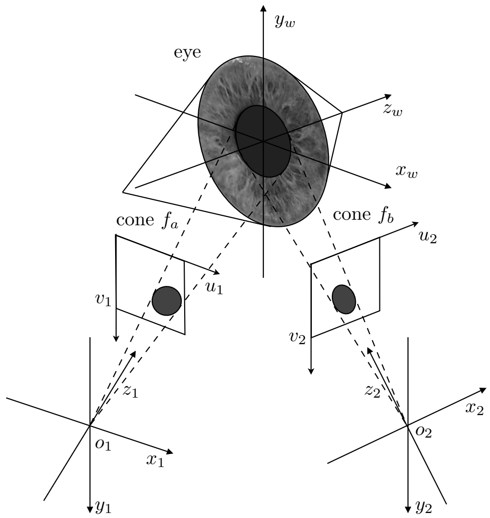

Figure 1: *Coordinate systems and two images of the pupil*

the pupils are in the normalized image planes, defined by the vectors u_1 and v_1 as well as u_2 and v_2. If the original pupil is assumed to be an ellipse (or a circle), its projection results in an ellipse again, because projective transformations applied to conic sections always result in conic sections again. The projected ellipses are defined by the two conics below:

$$\mathbf{x}^T\mathbf{A}_1\mathbf{x} = 0 \tag{1}$$
$$\mathbf{x}^T\mathbf{A}_2\mathbf{x} = 0 \tag{2}$$

Given the raw ellipse data, which is the lengths the two principal axes a and b, the position in the image plane t_x and t_y as well as the angle ϕ between the x-axis of the image plane and the first main axis of the ellipse, the conic matrix is obtained by applying an affine transformation $\mathbf{S}$ to a conic $\mathbf{H}$ in normal form as follows:

$$\mathbf{S} = \begin{pmatrix} \cos\phi & -\sin\phi & -t_x\cos\phi + t_y\sin\phi \\ \sin\phi & \cos\phi & -t_x\sin\phi - t_y\cos\phi \\ 0 & 0 & 1 \end{pmatrix} \tag{3}$$
$$\mathbf{H} = \begin{pmatrix} \frac{1}{a^2} & 0 & 0 \\ 0 & \frac{1}{b^2} & 0 \\ 0 & 0 & -1 \end{pmatrix} \tag{4}$$
$$\mathbf{A} = \mathbf{S}^T\mathbf{H}\mathbf{S} \tag{5}$$

The relation between the coordinate system c_i of camera i and the world coordinate system c_w is:

$$\mathbf{x}_i = \mathbf{R}_i\mathbf{x}_w + \mathbf{t}_i \qquad i = 1, 2 \tag{6}$$

For points in the pupil plane $\mathbf{x} = \left(x_w, y_w, 0\right)^T$ this can be written as:

$$\mathbf{x}_i = \mathbf{G}_i\mathbf{u}_w \qquad i = 1, 2 \tag{7}$$

with $\mathbf{u}_w = \begin{pmatrix} x_w & y_w & 1 \end{pmatrix}^T$ being homogenous coordinates on the pupil plane and $\mathbf{G}_i$ being a 3×3 matrix consisting of the first two columns of $\mathbf{R}_i$ and the translation vector $\mathbf{t}_i$.

$$\mathbf{G}_i = \begin{pmatrix} \mathbf{r}_{i1} & \mathbf{r}_{i2} & \mathbf{t}_i \end{pmatrix} \tag{8}$$

With $u_i = \frac{x_i}{z_i}$ and $v_i = \frac{y_i}{z_i}$ follows:

$$z_i\mathbf{u}_i = \mathbf{G}_i\mathbf{u}_w \qquad i = 1, 2. \tag{9}$$

An ellipse in the pupil plane is defined by

$$\mathbf{Q} = \begin{pmatrix} \frac{1}{a^2} & 0 & 0 \\ 0 & \frac{1}{b^2} & 0 \\ 0 & 0 & -1 \end{pmatrix} \tag{10}$$

and

$$\mathbf{u}_w^T\mathbf{Q}\mathbf{u}_w = 0 \tag{11}$$

and its projections by

$$\mathbf{u}_i^T\mathbf{A}_i\mathbf{u}_i = 0 \qquad i = 1, 2 \tag{12}$$

Inserting (9) in (12) yields:

$$\mathbf{u}_w^T\mathbf{G}_i^T\mathbf{A}_i\mathbf{G}_i\mathbf{u}_w = 0 \qquad i = 1, 2 \tag{13}$$

As equation (11) und (13) define the same conic $\mathbf{Q}$ we can write:

$$\mathbf{G}_i^T\mathbf{A}_i\mathbf{G}_i = k_i\mathbf{Q} \tag{14}$$

Thereby k_1 and k_2 are unknown scale factors, because $\mathbf{x}^T\mathbf{A}\mathbf{x} = 0$ and $\mathbf{x}^T k\mathbf{A}\mathbf{x} = 0$ describe the same conic.

The basic constraint is then given by:

$$\mathbf{G}_1^T\mathbf{A}_1\mathbf{G}_1 = k_1\mathbf{Q} \tag{15}$$
$$\mathbf{G}_2^T\mathbf{A}_2\mathbf{G}_2 = k_2\mathbf{Q} \tag{16}$$

with

$$\mathbf{G}_1 = \begin{pmatrix} \mathbf{r}_{11} & \mathbf{r}_{12} & \mathbf{t}_1 \end{pmatrix} \tag{17}$$
$$\mathbf{G}_2 = \begin{pmatrix} \mathbf{r}_{21} & \mathbf{r}_{22} & \mathbf{t}_2 \end{pmatrix} \tag{18}$$

Furthermore the relation between the coordinate systems c_1 and c_2 is known by calibration:

$$\mathbf{R}_2 = \mathbf{R}\mathbf{R}_1 \tag{19}$$
$$\mathbf{t}_2 = \mathbf{R}\mathbf{t}_1 + \mathbf{t} \tag{20}$$

The equations (15) and (16) provide 12 constraints, because they consist of two real symmetric 3×3 matrices with 6 parameters each.

As we only have 10 independent unknowns (three each in $\mathbf{R}_1$ and $\mathbf{t}_1$ as well as k_1, k_2, a and b) the system is overdetermined.

It has been shown that $\mathbf{R}_1$ and $\mathbf{t}_1$ can be solved for independently [De Ma 1993]. With $\mathbf{X}^{2\times2}$ being the upper left submatrix of $\mathbf{X}$ we can write:

$$(\mathbf{R}_1^T\mathbf{A}_1\mathbf{R}_1)^{2\times2} = k_1\mathbf{Q}^{2\times2} \tag{21}$$
$$(\mathbf{R}_2^T\mathbf{A}_2\mathbf{R}_2)^{2\times2} = k_2\mathbf{Q}^{2\times2} \tag{22}$$

Substituting equation (19) into equation (22) we can write after eliminating $\mathbf{Q}^{2\times2}$:

$$[\mathbf{R}_1^T(\mathbf{A}_1 - k\mathbf{R}^T\mathbf{A}_2\mathbf{R})\mathbf{R}_1]^{2\times2} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \tag{23}$$

$$\text{with } k = \frac{k_1}{k_2}$$

If the 2×2 upper left submatrix of a 3×3 matrix is equal to a zero matrix, this means that its determinant is equal to zero. Therefore (23) gives:

$$\det(\mathbf{A}_1 - k\mathbf{R}^T\mathbf{A}_2\mathbf{R}) = 0 \tag{24}$$

which is equivalent to

$$\det\left[(\mathbf{R}^T\mathbf{A}_2\mathbf{R})^{-1}\mathbf{A}_1 - k\mathbf{I}\right] = 0, \tag{25}$$

with $\mathbf{I}$ being the identity matrix. This means that k is the eigenvalue of the matrix $(\mathbf{R}^T\mathbf{A}_2\mathbf{R})^{-1}\mathbf{A}_1$. By calculating k and denoting $\mathbf{C} = \mathbf{A}_1 - k\mathbf{R}^T\mathbf{A}_2\mathbf{R}$, equation (23) can be written as:

$$(\mathbf{R}_1^T\mathbf{C}\mathbf{R}_1)^{2\times 2} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \tag{26}$$

Equation (26) only provides two independent equations because the matrices are 2×2 symmetric and $\det \mathbf{C} = 0$ has already been used for the solution of k.

Now $\mathbf{R}_1$ can be solved for: One of the eigenvalues of the matrix $\mathbf{C}$ is zero, because $\det(\mathbf{C}) = 0$. Given the two non-zero eigenvalues λ_1 and λ_2 with the corresponding eigenvalues $\mathbf{s}_1$ and $\mathbf{s}_2$ the third column of $\mathbf{R}_1$ can be calculated by:

$$\mathbf{r}_{13} = \pm\mathrm{norm}\left(\sqrt{|\lambda_1|}\mathbf{s}_1 \pm \sqrt{|\lambda_2|}\mathbf{s}_2\right) \tag{27}$$

Here $\mathrm{norm}(\mathbf{x})$ means that $\mathbf{x}$ is normalized to length one. The above case differentiations result in four different possible solutions for $\mathbf{r}_{13}$.

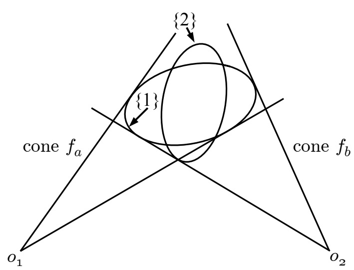

Figure 2: *Intersections of two cones*

This is due to the fact that two intersecting cones have two shared ellipses. As can be seen from Fig. 2 ambiguities can be resolved by considering two cases: In the first case (Fig. 2 $\{1\}$) both cameras are on the same side of the ellipse and in the second case (Fig. 2 $\{2\}$) one camera is on the front side and the other camera is on the back side of the ellipse. As there is no way to watch a pupil from both sides and we generally want the gaze-vector pointing away from the eyeball, we can rule out three of the solutions by ensuring that the z-components of both $\mathbf{r}_{13}$ and $\mathbf{r}_{23} = \mathbf{R}\mathbf{r}_{13}$ are positive. After having picked the correct $\mathbf{r}_{13}$ the remaining columns of $\mathbf{R}_1$ can be obtained by calculating the eigenvectors $\mathbf{r}_{11}$ and $\mathbf{r}_{12}$ of $\mathbf{H} = \mathbf{r}_{13}^T\mathbf{r}_{13}\mathbf{A}_1$.

In the last step, all remaining parameters can be solved:

$$\mathbf{R}_2 = \mathbf{R}\mathbf{R}_1 \tag{28}$$

$$\mathbf{t}_1 = \begin{pmatrix} \mathbf{r}_{11}^T\mathbf{A}_1^T \\ \mathbf{r}_{12}^T\mathbf{A}_1^T \\ \mathbf{r}_{21}^T\mathbf{A}_2^T\mathbf{R} \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 0 \\ -\mathbf{r}_{21}^T\mathbf{A}_2^T\mathbf{t} \end{pmatrix} \tag{29}$$

$$k_1 = -\mathbf{t}_1^T\mathbf{A}_1\mathbf{t}_1 \tag{30}$$

$$k_2 = \frac{k}{k_1} \tag{31}$$

$$a^2 = \frac{k_1}{\mathbf{r}_{11}^T\mathbf{A}_1\mathbf{r}_{11}} \tag{32}$$

$$b^2 = \frac{k_1}{\mathbf{r}_{12}^T\mathbf{A}_1\mathbf{r}_{12}} \tag{33}$$

3 Implementation

For the demonstrator, the 3D ellipse reconstruction algorithm has been incorporated into a head-mounted eye tracker (*EyeSeeCam*) [Boening et al. 2006; Dera et al. 2006].

Since previously only one camera per eye was used, the hardware setup was extended to support a second camera located next to the original one as depicted in figure 3.

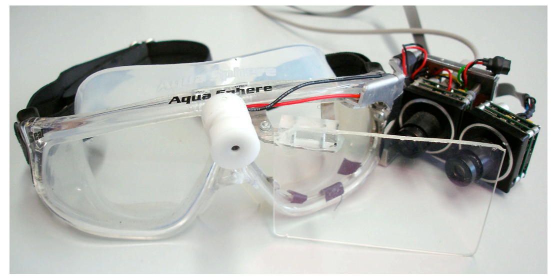

Figure 3: *Picture of the stereo goggles*

The cameras observe the eye through a semi-transparent mirror, which only reflects infrared light and the eye is lit by infrared LEDs on the inside of the goggle frame. Thus the cameras get a clear picture of the user's eye without affecting his field of view, because normal light passes straight through.

Being an experimental setup, only the left eye is tracked, with the option to add another pair of cameras later to do binocular measurements. The distance between the camera centers is 25 mm and the angle between them is 14°, so that the cameras center lines intersect at a distance of about 100 mm on the user's eye.

The system uses the image processing algorithm of the *EyeSeeCam* software which provides the five ellipse parameters per pupil that can be used for the ellipse reconstruction.

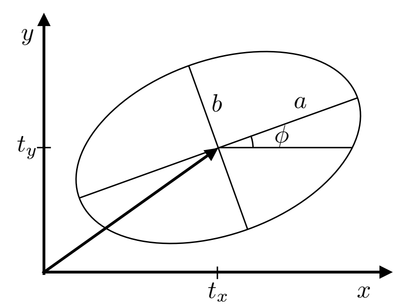

Figure 4: *Ellipse parameters*

These parameters – shown in figure 4 – have to be scaled and translated such that they are located in an image plane with a focal length of 1, thus taking the intrinsic parameters of the camera into account.

$$\hat{t}_x = \frac{1}{f}(t_x - c_x) \tag{34}$$

$$\hat{t}_y = \frac{1}{f}(t_y - c_y) \tag{35}$$

$$\hat{a} = \frac{a}{f} \tag{36}$$

$$\hat{b} = \frac{b}{f} \tag{37}$$

There f denotes the focal length of the camera. The point (c_x, c_y) is the intersection of the optical axis and the image plane. These parameters have been obtained through camera calibration.

The parameters $\hat{t}_x$, $\hat{t}_y$, $\hat{a}$, $\hat{b}$ and ϕ are used as input arguments for the stereo algorithm mentioned in section 2.

The result of the stereo reconstruction is the absolute length of the pupil's principal axes in millimeters. In the case of round pupils, the pupil radius is obtained. Additionally the normal vector of the pupil surface as well as the position of the pupil center are returned with respect to the left camera coordinate system.

4 Results

A simple experiment has been conducted with the head-mounted system. The subject had to fixate two horizontal rows with 5 points each at a distance of 8.5°. Figure 5 shows part of the input data, namely the position of the pupil center in both camera images.

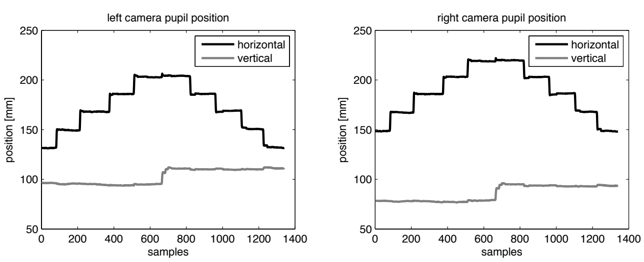

Figure 5: *Position of the pupil center in left and right camera image plane*

The results of the stereo reconstruction are depicted in figure 6. On the left side the position of the pupil center in 3D space is plotted. Note that for better readability 100 mm have been subtracted from the z-component. On the right side the angles of the gaze-vector in the horizontal and the vertical planes are plotted.

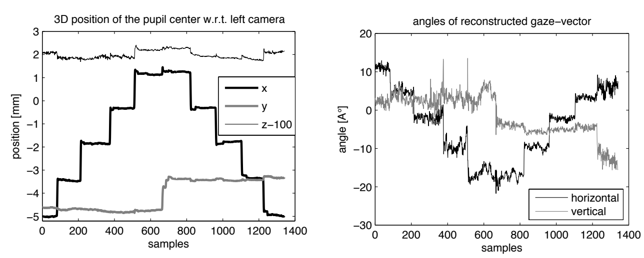

Figure 6: *Left: position of the pupil center in 3D space; right: angles of the gaze-vector in the horizontal and vertical plane*

With a standard deviation of about 0.02 mm the 3D position of the pupil can be extracted well. The gaze-vector, on the other hand, is noisy. In the horizontal plane the standard deviation varies between 0.5° and 1.6° depending on the point being fixated. In the vertical plane the standard deviation is between 0.4° and 2.2°.

5 Conclusion

The stereo reconstruction algorithm has proven useful for the implementation of a novel eye tracker with unprecedented functionality. Based on these results, eye trackers can be designed without the need for a lengthy fixation-based calibration procedure. The algorithm's closed-form solution guarantees real-time performance even with high sampling rates. One drawback, however, is that the ellipse parameters of the conics reconstruction algorithm are not well defined for gaze orientations around the primary position. In the current setup the resolution is up to 2.2° (RMS) which is due to the small pupil projections that lead to weakly defined ellipse parameters. This leads to considerable gaze-vector variabilities that are not sufficient for proper eye tracking.

Future work will address this problem by improving the image processing used to extract the pupil contours. Furthermore, methods of sensor fusion will be investigated, i.e., the ability to operate free of calibration will be combined with existing, but less variable tracking algorithms.

Another problem that has to be pointed out is the impact of corneal refraction, that distorts the pupil image. This is a systematic error that can be avoided by tracking the limbus instead of the pupil, or taken into account once the system produces more stable results.

A further drawback is the angular offset between the gaze vector and the pupil normal [Beymer and Flickner 2003]. This means that a calibration of the primary position is required. Additionally, the torsional eye movements will have to be tracked to calculate the correct gaze vector.

Acknowledgements

This work is supported in part within the DFG excellence initiative research cluster "Cognition for Technical Systems – CoTeSys", and by the Bavarian-Californian Technology Center (BaCaTeC).

References

BEYMER, D., AND FLICKNER, M. 2003. Eye gaze tracking using an active stereo head. *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on 2*.

BOENING, G., BARTL, K., DERA, T., BARDINS, S., SCHNEIDER, E., AND BRANDT, T. 2006. Mobile eye tracking as a basis for real-time control of a gaze driven head-mounted video camera. *Proceedings of the 2006 symposium on Eye tracking research & applications*, 56.

DE MA, S. 1993. Conics-based stereo, motion estimation, and pose determination. *International Journal of Computer Vision 10*, 1, 7–25.

DERA, T., BOENING, G., BARDINS, S., AND SCHNEIDER, E. 2006. Low-latency video tracking of horizontal, vertical, and torsional eye movements as a basis for 3dof realtime motion control of a head-mounted camera. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, 5191–5196.

2007. *EyeSeeCam*. `http://www.eyeseecam.com`.

2007. Camera calibration toolbox for matlab. `http://www.vision.caltech.edu/bouguetj/calib\_doc`, 06.

WANG, J., SUNG, E., AND VENKATESWARLU, R. 2005. Estimating the eye gaze from one eye. *Computer Vision and Image Understanding 98*, 1, 83–103.

ZHANG, Z. 1999. Flexible camera calibration by viewing a plane from unknown orientations. *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on 1*.