

Study on the interest of hybrid fundamental matrix for head mounted eye tracker modeling

Sacha Bernet¹
sacha.bernet@uha.fr

Peter Sturm²
peter.sturm@inria.fr

Christophe Cudel¹
christophe.cudel@uha.fr

Michel Basset¹
michel.basset@uha.fr

¹ MIPS Université de Haute Alsace
12, rue des frères Lumière
68093 Mulhouse cedex, France

² INRIA Rhône Alpes
Inovallée, 655 Avenue de l'Europe
38334 Saint Ismier cedex, France

Abstract

In this paper, we propose a new approach for calculating a user's real point of regard (POR) using a head mounted binocular eye tracking system. We show that the coupling of one of the user's eyes and the scene camera can be considered as a stereovision system. We then propose a new description for binocular eye trackers, in which we model them via pairs of stereovision systems having the scene camera in common. In this model, we use a hybrid fundamental matrix which allows to take into account the spherical shape of the eyeball model we use. Whatever the distance our model does not require any a priori information on the user's morphology to calculate the POR.

1 Introduction

Eye tracking systems are increasingly used for human behavior analysis as well as a means for interaction and communication for disabled persons. Many applications are intended with a screen and so Remote Eye-tracking Devices (RED) [1] are presumably the best solution. However, there are situations where the analysis is done when the user can freely move the head up and down and from right to left. For this application, the best solution is to use a Head-mounted Eye-tracker Device (HED) [2]. The head mounted eye tracker has a scene camera, supposed to give an image similar to the one seen by the user and one or two eye cameras, placed near the eyes without affecting the field of view of the user (see figure 3 for an example). For eye gaze analysis and to extract the center of the pupil we can work with only one eye for the monocular eye trackers, or use two eyes for binocular systems. Monocular systems use a point-to-point relation to compute the point of regard (POR) in the scene image from the pupil center detected in the eye camera. Monocular Eye Tracking is sufficient when the analysis or the interaction is done with a plane [3]. But when we are lead to work with a three-dimensional scene, depth can be recovered by the position and convergence of the two eyes to the POR [4].

In this paper, an HED is studied for its use in the automotive field, to analyze the driver's POR. The driver's gaze is pointing inside as well as outside of the vehicle, hence there are significant distances between viewed objects. Since the scene camera cannot be exactly placed in the same position as the user's eye, the viewpoint difference between the user and the scene camera causes a parallax error during the POR estimation. That is why we propose a model of a binocular eye tracking system based on a hybrid fundamental matrix.

Compared to existing works in this research area, we propose a different approach to model the binocular eye tracker. Binocular eye trackers have been developed to obtain the gaze direction or the POR in real environments. These methods follow the principle enunciated by [4] where the intersection between two lines obtained from the gaze directions of the two eyes, is calculated. This approach works, but calibration needs some physical measurements, like the diameter of the eyes and the distance between the optical centers of the eyes. In [5], an approach with neural networks to model the system and calculate a 3D gaze direction, was proposed.

An interesting approach had been proposed in [6], in which the concept of stereovision is introduced. The authors were interested in the error caused by using an eye tracker outside the distance of calibration. The eye tracker is modeled as a trifocal system with two eye cameras and one scene camera. The spherical shape of the eye is handled here by projecting the eye image onto a virtual plane, treated then as a perspective camera. This projection is only possible if the optical center of the user's eye is known; the authors of [6] propose to obtain it during the calibration phase (the user should look at the optical center of the eye camera). This approach is probably the closest to our proposed method in this paper, where we suggest to introduce the concept of hybrid fundamental matrix, initially described in [9] in the context of catadioptric cameras, in order to take into consideration the spherical shape of the eye.

This paper starts by describing Eye Tracking via the use of a stereovision model. This model reveals that there is no longer a point-to-point relationship with a monocular eye tracker, but a point-to-line relation, i.e. to each position of the pupil will correspond a line in the scene camera (an epipolar line). Therefore, we suggest using a binocular eye tracker to determine the POR by the intersection of two epipolar lines. Then we show that the spherical shape of the eye must be taken into account in this model and the geometry of the eye tracker is finally very close to a hybrid stereovision system, combining a perspective and a catadioptric camera. The paper concludes with simulated results; these confirm the potential applicability of the proposed model and method.

2 Binocular Eye Tracking and Stereovision

2.1 Epipolar geometry

Before explaining how an eye tracker system can be modeled as a stereovision system, we will remind the principles of epipolar geometry for perspective cameras. Epipolar geometry (Fig. 1) expresses the fact that for a point p_e in the left image, the matching point p_s in the right image must lie on a line. This line D_s is called the epipolar line of point p_e in the right image. If the 3D scene point P moves along a line formed by the optical center and the point

p_e , its projection on the left image will not move but the projection on the right image moves. This movement traces out the epipolar line D_s .

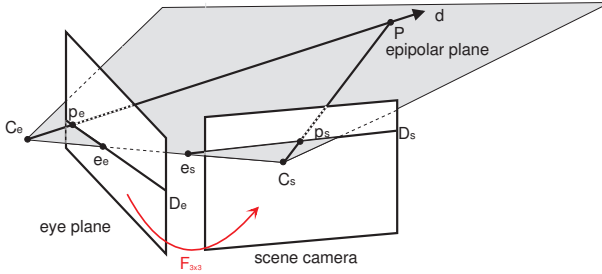


Figure 1: Epipolar constraint. C_e and C_s are the optical centers of the two cameras. P is a 3D scene point and p_e and p_s are its projections. D_e and D_s are the associated epipolar lines and e_e and e_s the epipoles.

The fundamental matrix noted F is the algebraic representation of the epipolar geometry. It only depends on the camera configuration (intrinsic and extrinsic parameters), not on the objects in the scene. For every point p_e in the left image the corresponding epipolar line is:

$$D_s = F p_e \quad (1)$$

The dimensions of the matrix F is 3×3 . Knowing one of the two points p_e or p_s , the following relation is equivalent to the corresponding epipolar line and constrains the matching of points:

$$p_s^t F p_e = 0 \quad (2)$$

In our scenario, we can take into consideration that the scene and eye cameras form a stereovision system. Unfortunately, we are of course not able to acquire the real image delivered by the eye and thus the fact of using the eye camera brings us to have an indirect stereovision system.

2.2 Eye tracker model

In a first approach, we approximate the spherical shape of the eye as a plane (Fig. 2). We can consider in this situation, that the eye camera sees the point P by its reflection in the eye plane. The eye can be considered as a kind of mirror, where the pupil center characterises the projection of P , or in other words, the projection of the point seen by the user. Under this condition, with an indirect stereovision system, all classic stereovision tools [1, 2] can be used. Notwithstanding, this system is a particular one since only one point interests us – the center of the eye’s pupil.

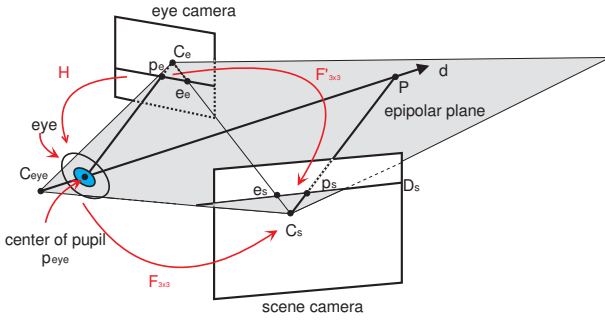


Figure 2: Model of the Eye-tracker system and the eye. Here, the eye’s surface is considered to be planar.

The relation between the eye plane and the eye camera can be defined as a homography as follows:

$$p_{eye} = H p_e \quad (3)$$

If F denotes the fundamental matrix between the scene camera and the eye, then the epipolar geometry between the eye camera and the scene camera, can be obtained by replacing p_e in equation (2), by the definition of p_{eye} as of equation (3):

$$p_s^t \underbrace{F H}_{F'} p_e = 0 \quad (4)$$

This formulation gives the possibility to compute the epipole in the scene camera as well as epipolar lines. Let p_e be the center of the eye pupil, extracted in the eye camera. The associated epipolar line passes through the user’s POR (cf. figure 2). It is computed as:

$$D_s = F' p_e \quad (5)$$

The obtained result is interesting since it shows the change of relation if we want to take into account the phenomenon of parallax into our model; the relation is not point-to-point but point-to-line. The first epipolar line is obtained by the combination of scene camera and right eye camera, and the second one that of the scene camera and the left eye camera. The POR is then computed by intersecting the two epipolar lines. Overall, our model thus consists of two stereo systems, having the scene camera in common.

2.3 Results

We have done different tests with a binocular eye tracking system, in order to assess if the assumption of a planar eye shape allows us to estimate fundamental matrices $F'_{3 \times 3}$ for the two stereo systems, and finally the POR, without much error. Figure 3 illustrates the experimental setup. The matches required to compute the fundamental matrices, are obtained by the protocol described in section 4.2.

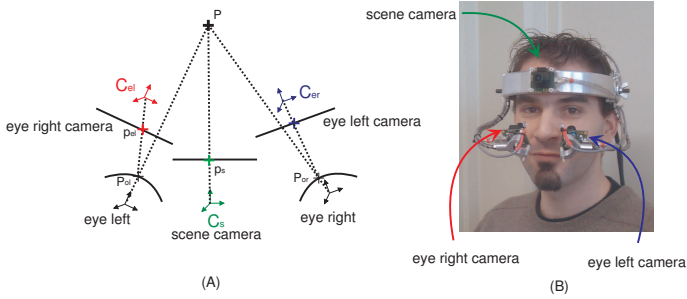


Figure 3: (A) Top view schematic of binocular eye tracker used in simulations; (B) Experimental prototype.

Figure 4 shows a typical example of estimated POR. For this experiment, the user was asked to focus the head of the person in the scene. Although the estimated POR is approximately in the right place, the localisation error is about 40 pixels, corresponding to about 6 degrees in terms of gaze direction, which is too large for our applications. We believe that one explanation is the fact that the eye cameras are placed close enough to the eyes such that the assumption of a planar eye shape is a too coarse approximation. In the next section, we thus extend our model by assuming a spherical eye shape.

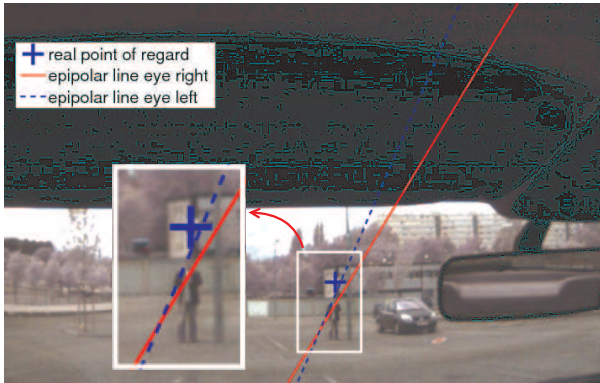


Figure 4: POR estimated with the approach described in this section. The distance to the real POR is 40 pixels; the angular difference between associated gaze directions is about 6 degrees.

3 Hybrid fundamental matrix

We propose in this section to extend the previous stereo model by assuming a spherical eye shape. Although this remains an approximation to the true shape, it is much better than a planar model. Also, we know from previous works [10] that the pupil's displacement is well described by the eyeball model. It means that in any situation, the pupil center is located on

a sphere. Figure 5 gives a corresponding representation of the system.

Like in section 2, the goal here will be to estimate and use the epipolar geometry between the scene camera and the system composed of the eye and the eye camera. This will then be used to estimate the POR in the scene camera, from its epipolar geometries with the two eye cameras. In the following, we described how to derive the epipolar geometry in our case; it will be applied to the detected position of the eye’s pupil center in the eye camera. Suppose for now that we know the complete geometry of our scenario, i.e. camera parameters (intrinsic and extrinsic) and the center and radius of the sphere representing the eye’s shape. Having detected the position of the pupil in the eye camera, we can first back-project this image point to 3D and intersect the back-projection line with sphere, to compute the pupil’s position in 3D. Then, we can compute the line of sight of the eye, simply by spanning the line between the sphere center and the point on the sphere that pupil’s 3D position. The image of this line of sight, in the scene camera, is the epipolar line we are seeking.

The process just described requires full knowledge of the system’s 3D geometry. However, like for perspective stereo systems, the epipolar geometry can also be expressed directly between the 2D images. To do this in our case, we first observe a perfect analogy of the system composed of eye and eye camera, with a catadioptric camera: the above back-projection procedure (from the pupil’s image in the eye camera, onto the sphere and then to the line of sight in 3D) is exactly the same as in the unified model for catadioptric cameras, first introduced in [1]. The only minor difference is that for most catadioptric cameras, the position of the perspective camera (which corresponds to the eye camera in our case) is usually inside the sphere in that unified model, whereas in our case, it is outside. However, algebraically there is no difference.

In [1], it was shown that the epipolar geometry between a central catadioptric camera and a perspective one, can be described by a 6×6 “hybrid” fundamental matrix. We can transpose this directly to our case. Let us first write down the associated expressions and then explain their meaning and application. Let P be the POR and $p_s = (x_s, y_s, 1)^t$ its image in the scene camera (we sometimes call this POR too). Let $p_e = (x_e, y_e, 1)^t$ be the image of the pupil center in the eye camera. From the above observations and [1], it follows that there exists a fundamental matrix F of size 6×6 such that the epipolar constraint between p_s and p_e can be written as:

$$\begin{pmatrix} x_s^2 & x_s y_s & y_s^2 & x_s & y_s & 1 \end{pmatrix} F_{6 \times 6} \begin{pmatrix} x_e^2 \\ x_e y_e \\ y_e^2 \\ x_e \\ y_e \\ 1 \end{pmatrix} = 0 \quad (6)$$

This expression can be interpreted as follows. For a given point p_s in the scene camera, the possible matching points p_e in the eye camera, must lie on a conic (the above equation is quadratic in the coordinates of p_e), the epipolar conic. The other way round is similar; however, the epipolar conic in the scene camera, degenerates into a pair of lines.

The calibration were carried out with 40 correspondences obtained from 3D points projected on both cameras following the model of figure 5. No noise was added here.

Figure 6 shows a calculation’s simulation of the POR using our 6×6 hybrid fundamental matrices. Epipolar curves in the scene camera are degenerate conics as expected. The intersection of two degenerate conics (a conic for each eye) gives four possible crossings whereof one corresponds to the real POR. In general, we cannot predict which ones among the four lines are the true and the parasite solutions. However, it is likely that simple heuristics may be used to differentiate these in practice. For our evaluation we determine, for the time being, among the four possible solutions for the POR the one closest to the ground truth.

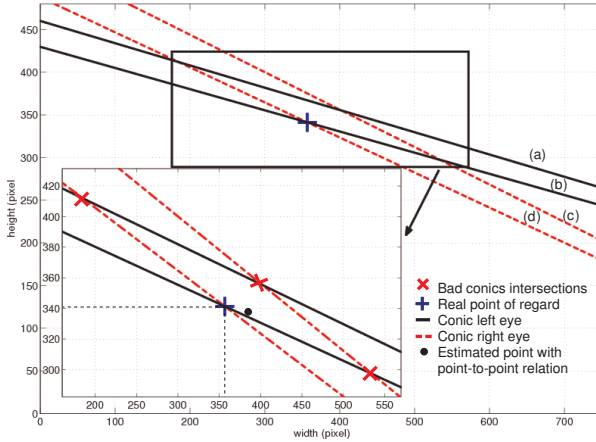


Figure 6: Calculation’s simulation POR using a matrix F . Two lines (a) and (b) correspond to the conic of the left eye. Same for the right eye. The lines intersections give the 4 solutions for the POR.

4.2 First experimental results with real images

Figure 7 shows a result obtained with real images. The calibration has been done with 40 matches between scene and eye cameras. To obtain these, the user was asked to focus a target moving in the scene. The pupil center was extracted automatically in the eye camera images; in the scene camera image, the target center was selected manually. From these matches, the two hybrid fundamental matrices, between the scene camera and the two eye cameras, were estimated with a linear least squares approach. It is very likely that 40 matches are not sufficient for a good stability but for a proof-of-concept it may be enough.

As expected (see section 3), the epipolar conics computed from the estimated fundamental matrices, are not degenerate. Their four intersection points are in the vicinity of the true POR, the closest one being about 30 pixels away. It is of course premature to draw detailed conclusions. However, the fact that the true POR is close to both epipolar conics, is promising, especially since the fundamental matrices were estimated from quite few matches. With many more matches, results should become better.

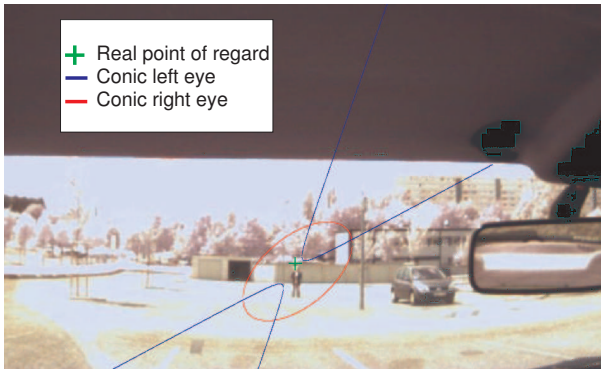


Figure 7: POR estimation with hybrid fundamental matrices and real images.

5 Conclusion

We proposed a method for a binocular eye tracker system based on modeling the camera system as two stereovision systems with one camera in common. Due to the morphology of the human eye, each stereo system is modeled by a hybrid fundamental matrix $F_{6 \times 6}$. We have validated this approach with simulations and first experimental results with real images are going in the right direction. Several explications can be given to comment these first results. First, our eye model is still simple – in reality the eye’s optical axis does not correspond exactly to the visual axis; there is an angle offset of around 4 degrees. We believe that this offset is taken into account in our model, but we have to check if this is actually true. Second, an important source of error is introduced with the real data used for the calibration, for which the position of the pupil and the target are noisy. To reduce the noise effect we project to compute the fundamental matrix by a non-linear method instead of a DLT [14] as it is done actually, and by using more matches. This work uses 6×6 fundamental matrices, which are the correct model under the assumptions made. However, we think that matrices of size 3×6 or 3×4 could give a better compromise between generality and stability and at the same time, lead to algebraically less complex formulations and fewer required matches for the estimation.

References

- [1] K. Barreto, J. and Daniilidis. Epipolar geometry of central projection systems using veronese maps. volume 1, page 1258–1265, 2006.
- [2] Martin Bohme, André Meyer, Thomas Martinetz, and Barth Erhardt. Remote eye tracking: State of the art and directions for future development. In *The 2nd Conference on Communication by Gaze Interaction*, 2006.
- [3] Christophe Cudél, Sacha Bernet, and Michel Basset. Fast and easy calibration for a head mounted eye tracker. In *COGAIN Communication, Environment and Mobility Control by Gaze*, pages 11–14, 2008.

- [4] Andrew T. Duchowski. *Eye Tracking Methodology: Theory and Practice*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2007. ISBN 1846286085.
- [5] Kai Essig, Marc Pomplun, and Helge Ritter. A neural network for 3d gaze recording with binocular eye trackers. *International Journal of Parallel, Emergent and Distributed Systems*, 21:79 – 95, 2006.
- [6] G Geyer and K Daniilidis. A unifying theory of central panoramic systems and practical applications. In *Proceedings of the 6th European Conference on Computer Vision*, pages 445–461, 2000.
- [7] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, New York, NY, USA, 2003. ISBN 0521540518.
- [8] Stefano Marra and Fiora Pirri. Eyes and cameras calibration for 3d world gaze detection. *Lecture Notes in Computer Science*, Volume 5008/2008:216–227, 2008.
- [9] Luis Puig, Josechu Guerrero, and Peter Sturm. Matching of omnidirectional and perspective images using the hybrid fundamental matrix. In *The 8th Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras - OMNIVIS*, 2008.
- [10] Arantxa Villanueva, Rafael Cabeza, and Sonia Porta. Eye tracking: Pupil orientation geometrical modeling. *image and vision computing*, 24 (7):663–679, 2005.
- [11] Andrew E. Welchman and Julie M. Harris. Tasks demands and binocular eye movements. *Journal of vision*, 3:817–830, 2003.
- [12] Zhengyou Zhang. Determining the epipolar geometry and its uncertainty: A review. *Int. J. Comput. Vision*, 27(2):161–195, 1998. ISSN 0920-5691. doi: <http://dx.doi.org/10.1023/A:1007941100561>.