Bidimensional eye position measurement using video-oculographic systems: close form solution and error analysis
F. Fioravanti, P. Bruno, P. Inchigolo
Unità di Bioingegneria, Dipartimento di Elettrotecnica, Elettronica ed Informatica, University of Trieste, Trieste, Italy

ABSTRACT

In recent years, the Video-Oculographic technique (VOG) has been increasingly adopted for the measurement of eye movements. For most of the existing VOG systems, the rotation angles of the eye are evaluated by detecting the position of the centre of the pupil image: this technique, however, assumes orthogonal projection for the optical system and needs an estimation of the diameter of the pupil and of the radius of the globe, which is normally performed using subjective calibration sessions, where the patient must fixate a set of known position targets. Such procedure is not suitable for recordings on strabismic subjects or on children, and does not allow an accurate evaluation of the precision of the system. This study introduces a precise model of the image formation process, allowing the error analysis both analytically and with computer simulations. An alternative solution for the measurement of the eye rotation angles requiring no subjective calibration is also presented.

INTRODUCTION

The Video-Oculographic eye position measurement systems (VOG) permit multi-dimensional, drift-free and non-invasive recording of eye movements. However, in order to evaluate the eye position, a relationship between the rotation angles of the globe and some characteristics to be measured on the images of the eye should be priorly defined. For most of the existing VOG systems, this relation is defined describing the image formation process as an orthogonal projection from the eye to the surface of the sensing device [1,2,4]. With this assumption, the rotation angles can be evaluated by detecting the position of the centre of the pupil image; in addition, some subjective parameters of the eye (distance of the pupil from the centre of rotation of the
eye and/or diameter of the globe) must be priorly estimated with a calibration procedure that requires the fixation of some targets placed in known positions. The detection of the pupil centre is usually performed by evaluating the centre of gravity of the black pixels of the image or by detecting the position of the pupil edge [1,4]. This technique, however, is not suitable for a precise error analysis, since its simplifications do not take into account most of the noise sources affecting the recording system, like the non-linearities due to the presence of the optical system, and include errors due to the imperfect fixation of the calibration targets. Moreover, it must be considered that the fixation task cannot be performed at all by strabismic subjects and with low precision by most of the young children. In addition, the diameter of the pupil must be evaluated: since mydriasis produces changes of the pupil diameter up to 60% [1], the measurements based on the estimation of this parameter are affected by large errors. In order to evaluate the maximum global error on the measurement of the rotation angles, a precise model of the whole system must be considered, taking into account the presence of the optical system and the effects of spatial discretisation due to the digital image sampling. The numerical conditioning of the identification algorithms, with respect to errors in the parameter measurement on the pupil image, can be finally studied by introducing typical noise sources for VOG systems (corneal reflection, grey level quantization, partial occlusion of the eye-lids, etc.). By introducing into the system also the spatial discretisation noise, the relationship between the precision of the system and the spatial sampling frequency can be clearly defined.

ANALYTICAL MODEL OF THE VOG SYSTEM

The movement of a rigid body has in general six degrees of freedom: three of them are related to translation, the remaining to rotation. However, since the movements of the eye are governed by mechanical and neural constraints, the rotation centre of the globe can be assumed stationary: this approximation is generally accepted, since translations of the eye can be considered negligible in normal subjects. With this assumption, the degrees of freedom of the eye are reduced to three, and the eye movement can be described by a system of three angular co-ordinates. The choice of an appropriate co-ordinate system normally depends on the recording set-up, since the description of the movement in terms of a different co-ordinate system can be obtained with simple substitutions. In this study we adopted Pick's co-ordinate system, which is one of the most popular in many fields and allows a clear understanding of eye rotations in the space [5]. Measuring these angular co-ordinates with respect to a fixed orthogonal co-ordinate system with the origin in the rotation centre of the eye (craniotopic reference, CR), the position of the eye will be described by a sequence of three rotations around the axis of a floating co-ordinate system defined on the eye-ball (Eye-Ball reference, ER), as shown in figure 1. The CR and the ER will be described by the axis \((X,Y,Z)\) and \((X_e,Y_e,Z_e)\) respectively. The position of the ER with respect to the CR can be identified by three rotations: \(\alpha_e\) around axis \(Y_e\) (longitude), followed by a rotation \(\beta_e\) around \(X_e\)
Computer Simulations in Biomedicine 477

(latitude) and \( \gamma_e \) around \( Z_e \) (torsion). The identification of the third co-ordinate (ocular torsion) exceeds the purpose of this paper, since the identification of the rotation angles in the horizontal and vertical plane is independent of rotations around the fixation axis \( Z_e \). Therefore, the torsion angle \( \gamma_e \) can be assumed equal to zero without lack of generality, and the consecutive rotations around the ER horizontal and vertical planes can be described by two rotation matrixes.

\[
\theta = \begin{bmatrix}
\cos(\alpha_e) & 0 & -\sin(\alpha_e) \\
-\sin(\alpha_e)\sin(\beta_e) & \cos(\beta_e) & \cos(\alpha_e)\sin(\beta_e) \\
\cos(\beta_e)\sin(\alpha_e) & \sin(\beta_e) & \cos(\alpha_e)\cos(\beta_e)
\end{bmatrix}
\]

According to the non-commutative product of rotation matrixes, the relationship between the two co-ordinate systems \((X,Y,Z)\) and \((X_e,Y_e,Z_e)\) is given by the following expression:

\[
\bar{x}_e = E(\alpha_e,\beta_e) \cdot \bar{x} = E^X(\beta_e) \cdot E^Y(\alpha_e) \cdot \bar{x}
\]

Figure 1: Definition of the Cartesian co-ordinate systems for the VOG model: Craniotopic Reference (CR), the Eye-Ball Reference (ER) and the Sensing Device Reference (SR).

According to the non-commutative product of rotation matrixes, the relationship between the two co-ordinate systems \((X,Y,Z)\) and \((X_e,Y_e,Z_e)\) is given by the following expression:

\[
\bar{x}_e = E(\alpha_e,\beta_e) \cdot \bar{x}
\]

where

Since the surface of the sensing device can generally be assumed as a plane \((X_s,Y_s)\), a Sensing Device Reference (SR) can be defined by an orthogonal third axis \( Z_s \) through the centre of the sensing area: in this case, the relationship between the co-ordinate systems \((X,Y,Z)\) and \((X_s,Y_s,Z_s)\) is a roto-translation and can be described as follows:

\[
\bar{x}_s = S(\alpha_s,\beta_s,\gamma_s) \cdot \bar{x} + \delta_s = S^Z(\gamma_s) \cdot S^X(\beta_s) \cdot S^Y(\alpha_s) \cdot \bar{x} + \delta_s
\]
with evident meaning for symbols $\alpha_s, \beta_s, \gamma_s$. The vector $\delta_s$ defines the translation of the origin of the SR with respect to the CR. The choice of these references is conditioned by some considerations about the recording set-up. In particular, the angles $\alpha_e, \beta_e, \gamma_e$ represent the measurement of the position of the eye with respect to a craniotopic co-ordinate system, while $\alpha_s, \beta_s, \gamma_s$ and $\delta_s$ identify the displacement of the sensing device with respect to patient's head. These quantities can be easily evaluated by a patient-independent calibration procedure: in particular, the translation of the sensing device with respect to the CR can be identified by studying the corneal reflection of the incident light field, while the coefficients of the rotation matrix $S(\alpha_s, \beta_s, \gamma_s)$ can be evaluated by measuring the position of some light spots (LED) placed in known positions on the head of the patient and defining by themselves the orientation of the CR axes. A more detailed discussion of the calibration procedure will be presented elsewhere.

Combining equations (2) and (3), and considering that the inverse of a rotation matrix is simply its transposed, we obtain a direct relationship between the ER and the SR co-ordinate systems:

$$\tilde{x}_s = S \cdot E^T \cdot \tilde{x}_e + \delta_s = M \cdot \tilde{x}_e + \delta_s$$

(4)

The pupil can be described, in the ER co-ordinate system, as a circle of radius $R_p$ laying on a plane $z_e = Z_p$; this description is valid in normal subjects, and should be corrected in case of some pathologies of the central nervous system: however, the analysis of such particular cases exceeds the purpose of this paper and will not be analysed. For the following analysis, the pupil edge can be more clearly described in terms of parametric equations:

$$\bar{\Gamma}_e(\varphi) = (\Gamma^x_e(\varphi), \Gamma^y_e(\varphi), \Gamma^z_e(\varphi))^T = (R_p \sin(\varphi), R_p \cos(\varphi), Z_p)^T$$

(5)

where the parameter $\varphi$ is defined in the interval $[0, 2\pi]$.

The description of the edge in the SR co-ordinate system can be obtained by substituting the parametric equations (5) in equation (4):

$$\bar{\Gamma}_s(\varphi) = M \cdot (R_p \sin(\varphi), R_p \cos(\varphi), Z_p)^T + \tilde{\delta}_s$$

(6)

It can be shown that for a typical recording set-up the optical system can be described by the paraxial rays approximation or, in other words, with the thin lenses equations. Under this hypothesis, we will make the following assumptions: 1) the sensing surface of the recording device lays on the secondary plane of the lenses; 2) the centre of the lenses lays on the Z axis of the SR, and therefore it can be described by the point $V_s = (0,0,F)$ in the SR co-ordinate system. The image formation process can now be described by a conic projection of each point of the source image (the pupil edge) on the sensing surface: the vertex of the cone is placed in the centre of the optical system $V_s$ (figure 2).
Figure 2: Conic projection from the pupil plane to the sensing surface. The vertex of the cone is in the centre of the optical system.

In the SR co-ordinate system the parametric equations of the cone projecting the pupil edge are described by the following expression:

\[ \Sigma_s(\sigma, \phi) = (\Sigma_s^x, \Sigma_s^y, \Sigma_s^z)^T = \sigma \left[ \Gamma_s(\phi) - \bar{V}_s \right] + \bar{V}_s \]  

where the parameters \( \sigma \) and \( \phi \) are defined in the intervals \([-\infty, +\infty]\) and \([0, 2\pi]\) respectively. The intersection of the cone with the sensing surface (plane \( z_s = 0 \)) directly gives the parametric description of the pupil edge projection on the sensing surface:

\[ \Sigma_s^x(\phi) = \frac{\Gamma_s^x(\phi) \cdot F}{F - \Gamma_s^z(\phi)} \quad \Sigma_s^y(\phi) = \frac{\Gamma_s^y(\phi) \cdot F}{F - \Gamma_s^z(\phi)} \]  

By substituting equations (7) in the above parametric system we obtain the description of the edge of the pupil image on the sensing device as a function of the rotation angles of the eye and of the displacement of the sensing device:

\[ \Sigma_s^x(\phi) = \left( \sqrt{m_{11}^2 + m_{12}^2} R_p \sin(\phi + \arctan(m_{12}/m_{11})) + m_{13} Z_p + \delta_s^x \right) G \]  

\[ \Sigma_s^y(\phi) = \left( \sqrt{m_{21}^2 + m_{22}^2} R_p \sin(\phi + \arctan(m_{22}/m_{21})) + m_{23} Z_p + \delta_s^y \right) G \]  

\[ G = \frac{F}{F - \sqrt{m_{31}^2 + m_{32}^2} R_p \sin(\phi + \arctan(m_{32}/m_{31})) - m_{33} Z_p - \delta_s^z} \]
It can be clearly seen that, if $G$ is constant for different values of $\varphi$, the parametric equations (9) and (10) describe an elliptical edge: this assumption is not valid in general, but permits the identification of the rotation angles of the eye by measuring some parameters on the image of the pupil, like the centre of the elliptic contour and its axes. Since this approach is the most suitable for a real-time detection of the eye position, it is commonly accepted in most of the VOG realisations. In the next paragraph we will introduce a new approach to the bi-dimensional measurement of the eye position, based on the assumption of elliptical contour for the image of the pupil edge. To this purpose, the gain $G$ will be assumed constant and equal to the magnification of the optical system, defined as follows:

$$G_0 = F / (F - Z_P - \delta_s^Z)$$

The effects of variations in the gain $G$ will be discussed in the last paragraph.

**DETECTION OF THE EYE POSITION**

By fitting an elliptic contour to the image of the pupil edge, we obtain the following parametric equations:

$$\hat{\Sigma}_s^x (\varphi) = X_0 + A_x \sin(\varphi + \Omega_x) \quad \hat{\Sigma}_s^y (\varphi) = Y_0 + A_y \sin(\varphi + \Omega_y)$$

and, matching the co-ordinates of the centre and the axis amplitudes to the corresponding terms in equations (9) and (10), we have

$$A_x = \sqrt{m_{11}^2 + m_{12}^2} \quad R_P \quad G_0 \quad A_y = \sqrt{m_{21}^2 + m_{22}^2} \quad R_P \quad G_0$$

$$X_0 = \left( m_{13} Z_P + \delta_s^x \right) G_0 \quad Y_0 = \left( m_{23} Z_P + \delta_s^y \right) G_0$$

where the first members of each equation are known. In particular, our interest is in the evaluation of coefficients $m_{ij}$, which are directly related to the rotation angles of the eye-ball around the $X_e$ and $Y_e$ axes. On the other hand, however, the term $R_P$ should be avoided, since it can not be assumed constant over the recording interval. By introducing the quantities:

$$K_1 = \frac{A_y}{A_x} = \sqrt{\frac{m_{21}^2 + m_{22}^2}{m_{11}^2 + m_{12}^2}} \quad K_2 = \frac{X_0 - G_0 \delta_s^x}{Y_0 - G_0 \delta_s^y} = \frac{m_{13}}{m_{23}}$$

we can eliminate both the subjective parameters $R_P$ and $Z_P$. The constants $K_1$ and $K_2$ are known since they depend on the measured parameters $X_0$, $Y_0$, $A_x$ and $A_y$, on the gain of the optical system $G_0$ and, finally, on the displacement of the sensing device.
Considering the properties of rotation matrixes and solving with respect to \( m_{ij} \), we obtain:

\[
\begin{align*}
    m_{23} &= \pm \sqrt{\frac{1-K_1^2}{1-K_1^2 K_2^2}} \\
    m_{13} &= \pm K_2 \sqrt{\frac{1-K_1^2}{1-K_1^2 K_2^2}}
\end{align*}
\]  

(17)

The uncertainty in the sign of equations (17) can be eliminated taking into account the equations (15) and considering that \( z_p \) is a positive quantity:

\[
\text{Sign}(m_{13}) = \text{Sign}(X_0 - \delta_s^x) \\
\text{Sign}(m_{23}) = \text{Sign}(Y_0 - \delta_s^y)
\]  

(18)

The term \( m_{33} \) can be obtained directly from the properties of rotation matrixes, and its sign is not relevant since it produces an uncertainty of a factor \( \pi \) in the rotation angles, whose values is in any case defined in the range \([-\pi/2, +\pi/2]\).

Recalling the definition of matrix \( M \) we have the following linear equation system:

\[
\begin{align*}
    s_{11} e_{31} + s_{12} e_{32} + s_{13} e_{33} &= m_{13} \\
    s_{21} e_{31} + s_{22} e_{32} + s_{23} e_{33} &= m_{23} \\
    s_{31} e_{31} + s_{32} e_{32} + s_{33} e_{33} &= m_{33}
\end{align*}
\]  

(19)

Solving the system of three equations in the three unknowns \((e_{31}, e_{32}, e_{33})\) yields, with some trigonometric substitutions:

\[
\alpha_e = \text{ArcSin} \left( \frac{e_{31}}{\sqrt{1-e_{32}^2}} \right) \\
\beta_e = \text{ArcSin} (e_{32})
\]  

(20)

In conclusion, the detection of the rotation angles of the eye-ball in the horizontal and vertical planes can be obtained by detecting the centre and the axes of the ellipse describing the image of the pupil edge. This solution is based only on objective parameters, like the sensing device displacement, and does not require the measurement of subjective parameters like the pupil radius.

**ERROR ANALYSIS AND COMPUTER SIMULATION**

In the preceding paragraph we assumed a constant value \( G_0 \) for the factor \( G \) of equation (11). This hypothesis is verified only in the case:

\[
m_{31}^2 + m_{32}^2 = 0
\]  

(21)
This condition is true if $\alpha_s = \alpha_e$ and $\beta_s = \beta_e$ or, in other words, when the sensing surface is normal to the fixation axis of the eye. In this case the gain $G$ matches the gain $G_0$ of the optical system under the hypothesis of paraxial rays, as it can be clearly seen comparing equation (11) with equation (12). Neglecting the changes in the gain $G$, the image of the pupil edge can be described by an elliptic contour, and the image formation process is therefore approximated by an orthogonal projection with constant magnification $G_0$. The error introduced by this approximation propagates in the identification algorithm presented in the previous paragraph, since it implements an indirect measurement of the rotation angles of the eye based on the measurement of the centre of the pupil image and of its axes. The co-ordinates of the pupil centre $C$ in the ER are $(0,0,Z_p)$, and the projection $\tilde{C}$ of this point on the sensing surface can be easily obtained from equations (4):

$$
\tilde{C}_x = \frac{m_{13} Z_p + \delta^x_s}{F - (m_{33} Z_p + \delta^z_s)} F
$$

$$
\tilde{C}_y = \frac{m_{23} Z_p + \delta^y_s}{F - (m_{33} Z_p + \delta^z_s)} F
$$

(22)

![Figure 3: Relative error (percent) on the identification of the pupil centre.](image)

However, under the hypothesis of constant gain ($G = G_0$), the image of the pupil centre is identified by the centre $(X_0, Y_0)$ of the elliptic contour on the sensing surface and, with some manipulations, the relative error is equal both for the $X$ and $Y$ co-ordinates and can be expressed as follows (figure 3):

$$
\Delta X_C = \Delta Y_C = \frac{Z_p (m_{33} - 1)}{F - Z_p - \delta^z_s}
$$

(23)

It can be clearly seen from figure 3 that the relative error affecting the pupil centre estimation depends on the deviation from the fixation axis and the $Z$ axis of the sensing surface: this error propagates to the indirect measurement of the
rotation angles according to the differential of equations (20). Expanding the equation for the vertical rotation angle $\beta_e$, in fact, we have:

$$\beta_e = \arcsin(m_{23} \cos(\beta_s) \cos(\gamma_s) + m_{33} \sin(\beta_s) + m_{33} \cos(\beta_s) \sin(\gamma_s))$$ (24)

and the maximum absolute error on the measurement of this angle is given by the following expression:

$$\Delta \beta_e = \frac{1}{\cos(\beta_s)} \left( \left| \frac{\partial \beta_e}{\partial x} \right| \Delta x_C + \left| \frac{\partial \beta_e}{\partial y} \right| \Delta y_C + \left| \frac{\partial \beta_e}{\partial A_x} \right| \Delta x_A + \left| \frac{\partial \beta_e}{\partial A_y} \right| \Delta y_A \right)$$ (25)

where $\left( \Delta x_C, \Delta y_C, \Delta x_A, \Delta y_A \right)$ are the maximum absolute errors on the measurement of the pupil image centre and of its axes. It can be observed that the partial derivatives of $\beta_e$ with respect to $X_0$, $Y_0$, $A_x$, $A_y$ introduce a bad numeric conditioning, since these variables are present in the denominator of terms $(m_{13}, m_{23}, m_{33})$. For example, expanding $m_{13}$ we have:

$$m_{13} = X_0 \cdot \sqrt{\frac{A_x^2 - A_y^2}{(A_x X_0)^2 - (A_y Y_0)^2}}$$ (26)

It can be clearly seen that the partial derivatives in equation (25) become very large when the eccentricity of the elliptic contour is close to 1 and its centre is close to the origin of the SR (figure 6, left panel). In this case, the numeric conditioning of the algorithm becomes bad, and absolute errors of 30µm in the estimation of the pupil image centre yield to errors up to 8deg in the identification of the rotation angles. In general, this problem can be avoided by choosing an appropriate displacement of the sensing device, but a more suitable solution can be obtained by introducing a calibration procedure. Let us consider a calibration session where the rotation angles are detected with the method presented in the preceding paragraph: during this session, the subject randomly moves his eyes within the recording field. For each position of the eye, the distance $Z_p$ of the pupil from the rotation centre of the eye-ball can be evaluated using one of the following expressions:

$$Z_p = \frac{1}{m_{13}} \left( \frac{X_0}{G_0} - \delta_x^s \right)$$

$$Z_p = \frac{1}{m_{23}} \left( \frac{Y_0}{G_0} - \delta_y^s \right)$$ (27)

A precise estimation of $Z_p$ can be obtained by averaging over a large population (100 to 200 eye positions). During the following recording session the quantities $(m_{13}, m_{23}, m_{33})$, which revealed to be responsible of the bad numeric
conditioning of the method, can be directly evaluated with the following expressions:

\[
m_{13} = \frac{1}{Z_p} \left( \frac{X_0}{G_0} - \delta_s^x \right) \quad m_{23} = \frac{1}{Z_p} \left( \frac{Y_0}{G_0} - \delta_s^y \right)
\]

Figure 4: Relative Error (percent) on the identification of parameter \(Z_p\) during the calibration procedure. The test has been performed with different values \(N_p\) of the sensing-device spatial-resolution (\(N_p \times N_p\) pixels).

With this approach, the identification of the rotation angles of the eye ball results much more stable with respect to the errors in the measurement of the elliptic contour parameters, and the maximum absolute error presents very small variability within the recording field (figure 6, right panel).

To test our algorithms and the overall precision of the method, we implemented a computer simulation of the system (figure 5). The program generates the image of the eye acquired by the sensing device by projecting each point of a high resolution digital image of the eye according to equations (9), (10) and (11). The high resolution image is rotated in the space around the axes of the ER, and then weighted by a non-linear light-intensity distribution.

Figure 5: Block diagram of the simulation program for the VOG-precision test.
The resulting function is projected on the sensing surface according to the projective model of the optical system, processed by a low-pass Finite Impulse Response (FIR) filter and resampled to fit the sensing-device resolution (\(N_p\)x\(N_p\) pixels). The low resolution image, is finally converted to a 8-bit grey-level scale and processed by the identification algorithms, returning the estimated rotation angles \(\alpha_e\) and \(\beta_e\). The identification of the elliptic contour of the pupil has been performed by an iterative least-mean-square fitting algorithm [4]. The simulation starts a calibration session for the identification of the \(Z_p\) parameter, and consequently evaluates the rotation angles using equations (28) instead of equations (17).

![Figure 6: Maximum Absolute Error in the identification of the horizontal rotation angle (\(N_p=128\)). Left panel: effects of the bad numeric conditioning when using equations (17); right panel: effects of the calibration procedure (\(N_p=1024\)) when using equations (28).](image)

To account for the degradation of the signal-to-noise ratio due to the image acquisition system, an additive Gaussian noise source is also introduced in the simulation. The simulation program allows to change the displacement of the sensing device and to set different values for its resolution. In order to reveal the conditions for the best performance, the number of pixels can be differently set during the calibration and the recording sessions.

![Figure 7: Maximum Absolute Error (MAE) affecting the measurement of the rotation angles as a function of spatial resolution of the sensing device (\(N_p\)) and of the precision in the estimation of the parameter \(Z_p\) after the calibration.](image)
The results of the error analysis obtained for \( N_p = 128 \) are shown in figure 6, where the effect of the calibration procedure on the numeric conditioning of the algorithms is evident.

Testing the overall precision of the system by choosing different values of the sensing device resolution \( (N_p) \) for the calibration and for the recording sessions, we showed that a consistent reduction of the data storage during the recording session can be reached by using high resolution images of the eye during calibration and reducing the number of pixels in the following eye-movement recording. As shown in figure 7, in fact, the maximum absolute error (MAE) depends much more on the precision of the calibration procedure than on the resolution of the sensing device during the recording session. This result suggests a solution to by-pass the difficulties related to the high data-transfer rate required for high precision VOG systems when a temporal sampling rate of up to 300 frames/s is needed [4]: considering the diagrams of figure (4) and (7), it can be seen that a precision of 0.1 deg can be obtained using a 128x128 pixels sensing-device during the recording session and running the calibration procedure on 1024x1024 pixels frames. Such different values of resolution in the two procedures could be obtained using a high resolution device and reducing the spatial sampling frequency during the recording session; however, a more suitable solution can be obtained using a low-resolution device and a robust algorithm for the resolution enhancement during the calibration session [4].

Acknowledgements: Work supported by the Italian Ministry of the University and of the Scientific and Technological Research (MURST), National Project "Bioimmagini e Neuroscienze" and by the European Community, project "Oculomotor Plasticity in Relation to Binocular Vision and Strabismus", contract #SCI-CT91-0747(TSTS)

REFERENCES