Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

Robust High-speed Binocular 3D Eye Movement Tracking System Using a Two-radii Eye Model

Danjie Zhu, Mikhail Kunin, and Theodore Raphan

Abstract— **Existing video-based eye movement tracking systems measure three-dimensional eye orientations by assuming the eye is a sphere that rotates around its center at a fixed radius. We found this model inaccurate. We have developed a system that uses a two-radii eye model, which assumes that the eye rotates around two different centers with different radii horizontally and vertically. We found this two-radii model more accurate in estimating the three-dimensional eye positions than the traditional one-radius eye model.**

I. Introduction

Over the past decade, digital video techniques have increasingly been used for measuring three-dimensional (3D) eye movements [1-12]. Horizontal and vertical (2D) eye positions can be calculated from the pupil center coordinates, which can be determined using center of mass algorithm [3, 4, 6, 7, 10] or partial ellipse fits to the pupil boundary [11]. The torsional eye position can then be determined by tracking of natural or attached landmarks on the eye [2, 9, 13-15], utilization of variations of the polar cross-correlation method [1, 3, 4, 6, 7], or more robustly using a restrictive template-matching technique [12].

The algorithms that implement the eye tracking have traditionally been based on the assumption that the eye is a perfect sphere that rotates around its center at a fixed radius (one-radius eye model). We have found an inaccuracy in the 2D eye positions calculated using the one-radius eye model. Test results indicate that horizontal eye movements are uniformly larger in absolute values than the actual eye rotations and vertical eye movements are uniformly smaller (Fig. 2). The errors in 2D eye positions can be up to 1.5° (Fig. 2). These findings suggest that a more accurate model of eye rotation should consider that the eye rotates about two different centers with two different radii. Furthermore, if geometric compensation [7] is used in calculating the torsion, the errors in 2D eye positions would also affect the accuracy in the torsion calculation.

We have developed an algorithm using a two-radii eye model. The eye is assumed to rotate about two centers, one for horizontal rotation and the other for vertical rotation. The eye also rotates with different radii horizontally and vertically. Testing of this algorithm has shown that the two-radii model was 76% more accurate (Fig. 3) in estimating the horizontal and vertical components of the 3D eye positions than the traditional one-radius model.

This work was supported by DC 05222 and DC05204.

Danjie Zhu is with Department of Computer and Information Science, Brooklyn College of the City University of New York, 2900 Bedford Avenue, Brooklyn, NY 11210, USA (phone: 718-951-4194; fax: 718-951-4489; e-mail: danjie@ nsi.brooklyn.cuny.edu).

Mikhail Kunin is with Department of Computer and Information Science, Brooklyn College of the City University of New York, 2900 Bedford Avenue, Brooklyn, NY 11210, USA (phone: 718-951-4194; fax: 718-951-4489; e-mail: danjie@ nsi.brooklyn.cuny.edu).

Theodore Raphan is with Department of Computer and Information Science, Brooklyn College of the City University of New York, 2900 Bedford Avenue, Brooklyn, NY 11210, USA. He is also with Department of Neurology, Mt. Sinai School of Medicine, New York, NY 10029, USA

II. Eye tracking algorithm using two-radii eye model

A. Mathematical Basis of Algorithm

Mathematically, the eye can be considered as a sphere. As a first approximation, the iris and pupil can be modeled as a plane which intersects the sphere [16]. The visual axis of the eye can be approximated by the optical axis [17] which intersects the center of pupil. When projected onto an image plane, the pupil boundary is a closed contour. The eye orientation can be determined from the center of the contour projected on the image plane.

The following assumption were made in deriving the eye tracking algorithm based on the two radii model to simplify the mathematical development:

(1) The two centers are positioned along the optic axis and may not necessarily coincide with the center of the eye ball.

(2) The eye exhibits ideal ball and socket behavior, so that all eye movements are pure rotations around the two centers of the eye, with no translation of the centers.

B. Coordinate Frames

To compute the eye orientation, two coordinate frames have been used: 1) a head-fixed camera-coordinate-frame (X-Y-Z, Fig. 1) whose roll axis is normal to the image plane when the eye looks straight at the camera, and 2) an eye-fixed coordinate frame (X_e-Y_e-Z_e, Fig. 1) which is defined as the pitch, roll and yaw axes of the eye. It coincides with the camera-coordinate-frame when the eye is looking straight at the camera.

In the two-radii model, the eye rotates around two different centers (C_h and C_V, Fig. 1) respectively. The vertical rotation (θ) around C_v can be effectively described as a translation (d_t) from C_h to C_v, followed by a rotation (θ) about C_v, and followed by a reverse translation ($-d_t$) from C_v to C_h (Fig. 1). Therefore eye movements in the two-radii model are described as a rotation (φ) about the head-fixed yaw (Z) axis, followed by a translation (d_t) from horizontal rotation center (C_h) to vertical rotation center (C_v) along the rotated yaw axis (Y_e), followed by a rotation (θ) about the rotated pitch (X_e) axis, followed by a reverse translation ($-d_t$) from C_v to C_h along the rotated yaw axis (Y_e), and finally followed by a rotation (ψ) about the optic (Y_e) axis (Fig. 1). The three consecutive transformations (d_t, θ and $-d_t$) are equivalent in effect to the vertical rotation (θ) around the vertical rotation center (C_v) with a different radius.

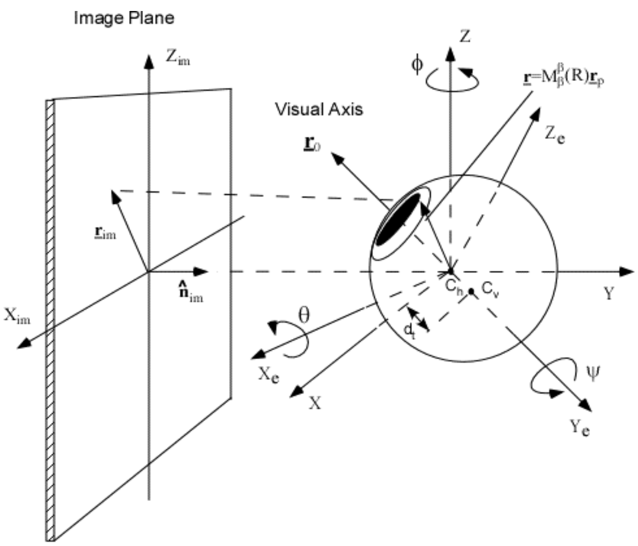

Fig. 1: Eye position in camera-coordinate-frame (X-Y-Z) with regard to eye-fixed coordinate frame (X_e-Y_e-Z_e) as seen by image plane (X_{im}-Z_{im}).

In practice, the camera may be tilted horizontally (φ_c), vertically (θ_c) and torsionally (ψ_c) relative to the reference position where the eye is looking straight ahead without any torsion. This camera offset can be determined during calibration.

C. Transformation Matrix

Eye movements in the head can be described by a transformation matrix given as a product of three rotation and two translation matrices given by Eq. 1. The three rotation matrices correspond to rotations about the head-fixed yaw (Z) axis ($\hat{R}_{\varphi}$), a rotated pitch (X_e) axis ($\hat{R}_{\theta}$), and the optic (Y_e) axis ($\hat{R}_{\psi}$) (Fig. 3). The two translation matrices correspond to a translation ($\hat{R}_{t1}$) from C_h to C_v along the optic (Y_e) axis and a reverse translation from C_v to C_h along the rotated optic (Y_e) axis. Since a 3×3 matrix does not provide translation transformation, 4×4 matrices and homogeneous coordinate representation of points have been used.

$$\hat{R}_{\phi}=\begin{pmatrix}\cos\phi & -\sin\phi & 0 & 0\\ \sin\phi & \cos\phi & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{pmatrix}$$

$$\hat{R}_{\theta}=\begin{pmatrix}1 & 0 & 0 & 0\\ 0 & \cos\theta & -\sin\theta & 0\\ 0 & \sin\theta & \cos\theta & 0\\ 0 & 0 & 0 & 1\end{pmatrix}$$

$$\hat{R}_{\psi}=\begin{pmatrix}\cos\psi & 0 & \sin\psi & 0\\ 0 & 1 & 0 & 0\\ -\sin\psi & 0 & \cos\psi & 0\\ 0 & 0 & 0 & 1\end{pmatrix}$$

$$\hat{R}_{t1}=\begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & d_t\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{pmatrix}$$

$$\hat{R}_{t2}=\begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & -d_t\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{pmatrix} \tag{1}$$

If camera offset is considered, an eye movement can be decomposed into two sub-movements: a first movement from the zero position where the eye is looking straight at the camera to the reference position where eye is looking straight ahead without any torsion, and a second movement from the reference position to the current position. When the eye is looking straight at the camera, Y_e axis (along the optic axis of the eye-fixed coordinate frame) is orthogonal to the image plane and X_e and Z_e axes of the eye-fixed coordinate system are parallel to X_{im} and Z_{im} axes of the image plane (Fig. 1). The transformation matrices describing the camera tilt are given by Eq. 2.

$$\hat{R}_{\phi_c} = \begin{pmatrix} \cos\phi_c & -\sin\phi_c & 0 & 0 \\ \sin\phi_c & \cos\phi_c & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\hat{R}_{\theta_c} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta_c & -\sin\theta_c & 0 \\ 0 & \sin\theta_c & \cos\theta_c & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (2)$$

$$\hat{R}_{\psi_c} = \begin{pmatrix} \cos\psi_c & 0 & \sin\psi_c & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\psi_c & 0 & \cos\psi_c & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Let P be a point on the pupil-iral plane. The homogeneous coordinates of P in the camera-coordinate-frame when the eye is at the zero and the current positions are (x, y, z, 1) and (x', y', z', 1) respectively.

Then the relationship between positions of point P before and after the eye rotation is given by:

$$\begin{pmatrix} x' \\ y' \\ z' \\ 1 \end{pmatrix} = \hat{R}_{\phi_c}\,\hat{R}_{\theta_c}\,\hat{R}_{\psi_c}\,\hat{R}_{\phi}\,\hat{R}_{t1}\,\hat{R}_{\theta}\,\hat{R}_{t2}\,\hat{R}_{\psi} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad (3)$$

In general the projection of the center of camera-coordinate-frame on the image plane is offset from the center of the image plane. Assume that the center of the camera-coordinate-frame projects to a point (x_c, z_c) in the image plane. Let (x', z') and (x_p, z_p) be the coordinates of point P in current position and its projection onto the image plane. Then the relationship between the projection of point P onto the image plane and point P in the camera-coordinate-frame is given below:

$$\begin{pmatrix} x_p \\ z_p \end{pmatrix} = \begin{pmatrix} x' \\ z' \end{pmatrix} + \begin{pmatrix} x_c \\ z_c \end{pmatrix} \quad (4)$$

D. Algorithm For Calibration

To develop the algorithm for computing the horizontal, vertical and torsional rotations of the eye, seven calibration parameters (φ_c, θ_c, ψ_c, x_c, z_c, $\mathbf{d_h}$, and $\mathbf{d_t}$) must be identified:

1) φ_c, θ_c and ψ_c are the Euler angles associated with the camera offset relative to the reference-coordinate-frame.

2) (x_c, z_c) are the coordinates of the projection of the center of the camera-coordinate-frame onto the image plane.

3) $\mathbf{d_h}$ is the distance between the horizontal rotation center of the eye and the pupil-iral plane.

4) $\mathbf{d_t}$ is the distance between the horizontal and vertical rotation centers.

The horizontal and vertical camera offset angles (φ_c and θ_c) can be omitted if they are less than 5° [7]. In practice, cameras can be mounted on goggles such that the "line of sight" of the camera is almost perpendicular (< 5°) to the "line of sight" of the eye when the eye is looking straight ahead. We now derive equations for computing the five calibration parameters ψ_c, x_c, z_c, $\mathbf{d_h}$ and $\mathbf{d_t}$ based on the coordinates of the projected pupil centers.

Since the torsional eye movement about the visual axis does not change the projection of the center of the pupil, ψ can be arbitrarily set to zero (ψ=0). When the eye is looking straight at the camera, the homogeneous coordinates of the pupil center is (0, -d, 0, 1) in the camera-coordinate-frame. Using (0, -d, 0, 1) for (x, y, z, 1), and substituting Eqs. 1 and 2 into Eq. 3, the coordinates of the pupil center in the camera-coordinate-frame when the eye is rotated by φ, θ, and ψ is given by:

$$\begin{aligned} x' &= (d_h + d_t)\cos\psi_c \sin\phi\cos\theta - (d_h + d_t)\sin\psi_c \sin\theta \\ &\quad - d_t \cos\psi_c \sin\phi \\ y' &= -(d_h + d_t)\cos\phi\cos\theta + d_t\cos\phi \\ z' &= -(d_h + d_t)\sin\psi_c \sin\phi\cos\theta - (d_h + d_t)\cos\psi_c \sin\theta \\ &\quad + d_t \sin\psi_c \sin\phi \end{aligned} \quad (5)$$

Using Eq. 4 and let d_v=d_h+d_t (Fig. 1), the projection of the pupil center onto the image plane is given by:

$$\begin{aligned} x_p &= d_v\cos\psi_c\sin\phi\cos\theta - d_v\sin\psi_c\sin\theta \\ &\quad - d_t\cos\psi_c\sin\phi + x_c \\ z_p &= -d_v\sin\psi_c\sin\phi\cos\theta - d_v\cos\psi_c\sin\theta \\ &\quad + d_t\sin\psi_c\sin\phi + z_c \end{aligned} \quad (6)$$

If we let:

$$\begin{aligned} &A = d_v\cos\psi_c && A' = -d_v\sin\psi_c \\ &B = -d_v\sin\psi_c && B' = -d_v\cos\psi_c \\ &C = -d_t\cos\psi_c && C' = d_t\sin\psi_c \\ &D = x_c && D' = z_c \\ &x = \sin\phi\cos\theta && y = \sin\theta \qquad z = \sin\phi \end{aligned} \quad (7)$$

we have:

$$\begin{aligned} x_p &= Ax + By + Cz + D \\ z_p &= A'x + B'y + C'z + D' \end{aligned} \quad (8)$$

In Eq. 8, (x_p, z_p) are the coordinates of the pupil center projected onto the image plane. They can be computed from the coordinates of the pupil boundary [11]. The Euler angles for each position, φ and θ, are known during calibration process. Therefore x, y and z can also be found. Theoretically the 8 parameters A, B, C, D, A', B', C' and D' can be solved from Eq. 8 by using 4 different calibration eye positions with 4 different (x_p, z_p) coordinates. In practice we found it to be more accurate to use Singular Value Decomposition (SVD) algorithm to calculate the 8 parameters. Five or more different calibration eye positions with their corresponding pupil center coordinates and the known euler angles are fit to Eq. 8 to calculate the 8 parameters. Once these parameters are found, ψ_c, x_c, z_c, $\mathbf{d_v}$, $\mathbf{d_t}$ and $\mathbf{d_h}$ can be obtained from Eq. 7 using the following equation:

$$
\begin{aligned}
\psi_c &= -\arctan\left(\frac{A'}{A}\right) \\
x_c &= D \\
z_c &= D' \\
d_v &= \sqrt{A'^2 + B'^2} = \sqrt{A^2 + B^2} \\
d_t &= SIGN(-C)\sqrt{C^2 + C'^2} \\
d_h &= \sqrt{A^2 + B^2} - d_t
\end{aligned}
\qquad (9)
$$

where d_h and d_v are the two radii for horizontal and vertical rotations, and $d_h = d_v - d_t$. There are two ways to calculate d_v. Since $d_h = d_v - d_t$, there are also two possible values for d_h. Theoretically they are the same, but in practice one ($\sqrt{A^2 + B^2}$) is more associated with horizontal rotation (d_h) and the other ($\sqrt{A'^2 + B'^2}$) with vertical rotation (d_v).

E. Algorithm For Determining Horizontal and Vertical Rotations

The coordinates of the projected center of the pupil (x_p, z_p) on the image plane are related to the pitch and yaw components (φ, θ) of eye orientation by Eq. 6. Once the calibration parameters ψ_c, x_c, z_c, $\mathbf{d_v}$, $\mathbf{d_t}$ and $\mathbf{d_h}$ are found (φ_c and θ_c are set to 0), φ, θ can be determined as follows:

$$
\begin{aligned}
\theta &= \arcsin\left(-\frac{(x_p - x_c)\sin\psi_c + (z_p - z_c)\cos\psi_c}{d_v}\right) \\
\varphi &= \arcsin\left(\frac{(x_p - x_c)\cos\psi_c - (z_p - z_c)\sin\psi_c}{(d_h + d_t)\cos\theta - d_t}\right)
\end{aligned}
\qquad (10)
$$

III. Results

Algorithms based on both traditional one-radius eye model and two-radii eye model were used to track the 3D eye positions. Video images of both eyes of a subject looking at 9 fixed points on a circular panel were obtained. The positions of the 9 points and the distance between the subject and the circular panel were chosen such that the horizontal and vertical angles of eye orientations for the 9 points are (-20°, 0°), (-10°, 0°), (0°, 0°), (10°, 0°), (20°, 0°), (0°, -20°), (0°, -10°), (0°, 10°) and (0°, 20°). 20 images were grabbed for each eye orientation except for the reference position (0°, 0°). For convenience of eye fixation, the reference position was used twice with 20 images grabbed each time. First the calibration parameters were calculated from images grabbed at 9 different positions using one-radius calibration algorithm [18]. Then the 3D eye positions were calculated using traditional one radius eye model algorithm. The calculated horizontal orientations overestimated, i.e., were larger in absolute value, than the actual eye position (Fig.2). The vertical orientations, underestimated the actual eye orientations and there was a tendency to deviate in the horizontal direction (Fig. 2). The average error was 0.45° horizontally and 0.47° vertically.

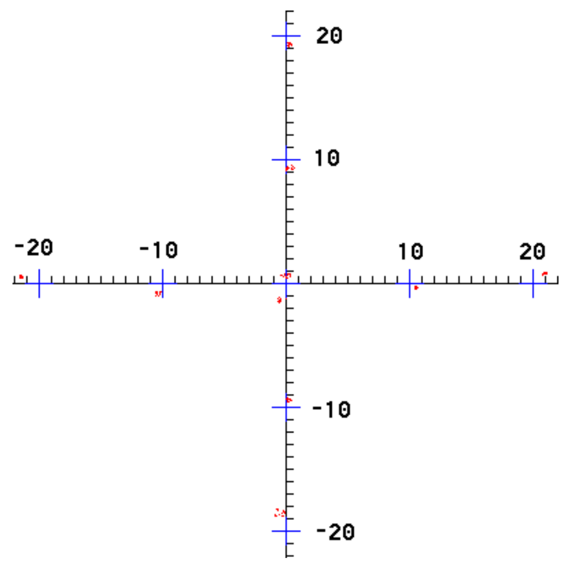

Fig. 2: 2D eye positions calculated using one-radius model. The red dots are the calculated 2D eye positions when subjects were asked to fixate points at 10° and 20° in the horizontal and vertical directions.

The same pupil center positions of all the images (20x10) were used to obtain the calibration parameters and to calculate the 3D eye orientations using the two-radii eye model algorithm (Eqs. 9, 10). The results for the 2D eye positions for the two-radii model (Fig. 3) indicate that there is improved accuracy. The average errors are 0.07° horizontally and 0.14° vertically.

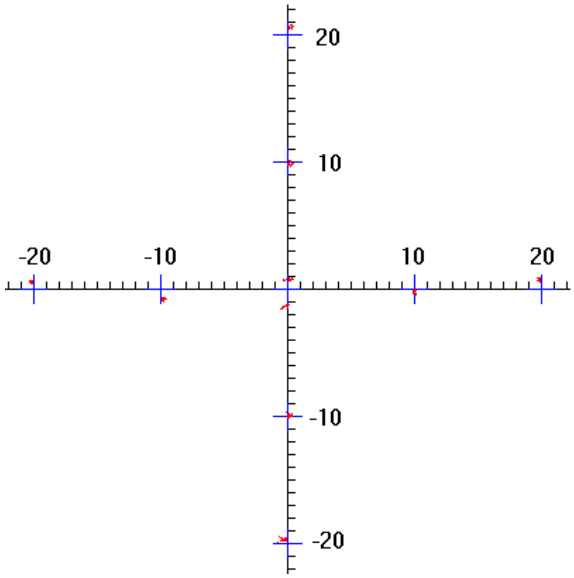

Fig. 3: 2D eye positions calculated using two-radii model. The red dots are the calculated 2D eye positions when subjects were asked to fixate points at 10° and 20° in the horizontal and vertical directions.

IV. Conclusions

The results indicate that there is an improvement of accuracy of 67% in computing eye position in two dimensions by using a two-radii model for the eye rotation. This affords the ability to compute eye rotations in three dimensions with greater accuracy and robustness.

References

[1] M. Hatamian and D. J. Anderson, "Design considerations for a real-time ocular counterroll instrument." IEEE Transactions on Biomedical Engineering, 1983. **BME-30** (5): p. 278-288.

[2] J. A. Parker, R. V. Kenyon and L. R. Young, "Measurement of torsion from multitemporal images of the eye using digital signal processing techniques." IEEE Transactions on Biomedical Engineering, 1985. **BME-32** (1): p. 28-36.

[3] T. Vieville and D. Masse, "Ocular counter-rolling during active head tilting in humans." Acta Otolaryngol, 1987. **103**: p. 280-290.

[4] A. H. Clarke, W. Teiwes and H. Scherer, "Video-oculography - an alternative method for measurement of three-dimensional eye movements," in Oculomotor Control and Cognitive Processes, R. Schmid and D. Zambarbieri, Editors. 1991, @Elsevier Science Publishers B. V.: North-Holand. p. 431-443.

[5] A.H Clarke, "Image processing techniques for the measurement of eye movement," in Eye Movements in Reading, J. Ygge and G. Lennerstrand, Editors. 1994, Pergamon: Oxford. p. 21-38.

[6] S. T. Moore, I.S. Curthoys and S. G. McCoy, "VTM-an image-processing system for measuring ocular torsion." Computer Methods and Programs in Biomedicine, 1991. **35**: p. 219-230.

[7] S. T. Moore, T. Haslwanter, I. S. Curthoys and S. T. Smith, "A geometric basis for measurement of three-dimensional eye position using image processing." Vision Res., 1996. **36**(No.3): p. 445-459.

[8] J. E. Bos and B. de Graaf, "Ocular torsion quantification with video images." IEEE Transactions on Biomedical Engineering, 1995. **BME-41**: p. 351-357.

[9] E. Groen, J. E. Bos, P. F. M. Nacken and Bernd de Graaf, "Determination of ocular torsion by means of automatic pattern recognition." IEEE Transactions on Biomedical Engineering, 1996. **BME-43** (No. 5): p. 471-478.

[10] J. B. Mulligan, "Image processing for improved eye tracking accuracy." Behavior Research Methods, Instruments and Computers, 1997. **29**(1): p. 54-65.

[11] D. Zhu, S. T. Moore and T. Raphan, "Robust pupil center detection using a curvature algorithm." Computer Methods and Programs in Biomedicine, 1999. **59**: p. 145-157.

[12] D. Zhu, S. T. Moore and T. Raphan, "Robust and real-time torsional eye position calculation using a template-matching technique." Computer Methods and Programs in Biomedicine, 2004. **74**: p. 201-209.

[13] K. Nakayama, "Photographic determination of the rotational state of the eye using matrices." American Journal of Optometry & Physiological Optics, 1974. **51**: p. 736-742.

[14] S. Yamanobe, S. Taira, T. Morizono, T. Yagi and T. Kamio, "Eye movement analysis system using computerized image recognition." Arch Otolaryngol Head Neck Surg., 1990. **116**: p. 338-341.

[15] D. Ott, F. Gehle, and R. Eckmiller, "Video-oculographic measurement of 3-dimensional eye rotations." Journal of Neuroscience Methods, 1990. **35**: p. 229-234.

[16] E. Wolff, "The anatomy of the eye and orbit." 2nd Edition ed. 1940, London: H.K. Lewis & Co.

[17] L. R. Young and D. Sheena, "Methods & Designs: survey of eye movement recording methods." Behavior Research Methods & Instrumentation, 1975. **7(5)**: p. 397-429.

[18] D. Zhu, "Video-based approach for tracking eye orientation in three fimensions", Ph.D. Thesis, Graduate School and University Center of City University of New York, New York, 1997.