TR-H-255

A Three-Dimensional Eye Movement Measurement System

Max BERTHOLD

1998.8.24

ATR人間情報通信研究所
〒619-0288　京都府相楽郡精華町光台2-2　TEL: 0774-95-1011

ATR Human Information Processing Research Laboratories
2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan
Telephone: +81-774-95-1011
Fax : +81-774-95-1008

A three-dimensional eye movement measurement system

Investigation of a torsion estimation algorithm and computation of absolute viewing angles.

Max Berthold
July 1998

1. Introduction

The knowledge about the eye's physiology is very old, but it was not until Fick, (1854) and Helmholtz, (1867) presented their rotation matrices that the eye's movements could be described easily in an analytical way. Using their rotation matrices, the movements of the eye can be expressed in angles that are related to different kinds of coordinate systems, e.g., eye fixed coordinate system or head fixed coordinate system. At present, the research concerning eye movements is mainly related to high velocity, different viewing conditions and medical research. Research concerning high velocity investigates the effect of body motion on vision at high speeds, e.g., modern fighter plane and space travels. Research on viewing conditions investigates problems regarding stereo displays and new communication media. Medical research involves different ways of treating disabilities, as well as fundamental vision research.

Assuming a coordinate system with the origin in the eye's center, the rotation matrix enables us to express the eye's three-dimensional movements as three separate rotations about the x-, y- and z-axis. In order to measure these rotations, different kinds of techniques have been used, e.g., photographic techniques using landmarks on the eye, embedding coils in a lens and measuring induced electricity, and tracking iris pattern (see [Berthold, 1997] and [Sheena, 1975] for an overview). The systems used to be clumsy, uncomfortable, time-consuming, and interfered with the eye while measuring (invasive measurement). However, due to the development of image digitizers, CPUs and CCD-cameras, a new generation of video-based eye tracking systems have been developed during the last years, [Bos *et.al.*, 1994], [Bucher *et. al.*, 1990], [Clarke *et.al.*, 1991], [Moore *et.al.*, 1991] and [Ott, 1990].

This report contains results from studies during the development of a three-dimensional eye tracking system performed from June 1997 to April 1998 at the Human Information Processing Research Laboratories (HIP Lab) at the Advanced Telecommunication Research Institute, Japan. The system is video-based and will measure the eyes' movements non-invasively using different computer-imaging techniques. Results from the implementation of a torsion algorithm, as well as an algorithm to express image plane coordinates in a head fixed coordinate system i.e., absolute viewing angles, will be presented. Earlier tests investigating the evaluation of horizontal and vertical viewing angles as well as a system description, have been presented in [Berthold, 1997].

The torsion algorithm is based on the work of Hatamian. *et.al.*, (1983), and in order to achieve sub-pixel accuracy we apply a curve fitting technique (a second order polynomial using seven points). Because of the large amount of data that is generated when a video based system is used, the evaluation time for each frame is an important factor. We have therefore studied how to optimize the use of the track arcs regarding the estimation of the torsion angle. Since the system is meant to measure horizontal, vertical and torsion eye movements without constraints, the torsion was also investigated for different horizontal and vertical viewing angles i.e., free viewing conditions. The results showed that, for a view field less than $\pm 10^{\circ}$, an accuracy of $\pm 0.25^{\circ}$ could be achieved when using 10 track arcs and 3 pairs of min and max values from the cross-correlation were discarded.

Since the main part of the research at the Vision Group in the HIP Lab is related to different kinds of stereo viewing conditions, it is important that the system can express the evaluated eye movements in viewing angles using the head's coordinate system. A new transformation algorithm was derived and the initial results show that the algorithm is stable and the error is $<\pm 0.17^{\circ}$ for the horizontal viewing angles and $<\pm 0.05^{\circ}$ for the vertical viewing angles.

In *System description* the torsion algorithm as well as the algorithm that transforms from image plane coordinates to absolute viewing angels will be presented. Chapter 3, *Experiment,* describes the different *in-vitro* experiments and the results. The results have been validated by comparing estimated viewing angles with set viewing angles (the accuracy of the set viewing angles were 5'). In chapter 4, *Discussion,* the problem with image quality is discussed and some ideas on future work are briefly presented.

2. System description

2.1 Introduction

We are developing an eye tracking system capable of measuring the eye's three rotational movements, i.e., horizontal, vertical and torsion, non-invasively. The main parts of the system are glasses, which contain two CCD-cameras, two VCRs (storage medium) and a PC. During the experiments, the eyes' movements are recorded on the VCRs and later on, "off-line", the computer program calculates the eye's three different rotational movements from the video images.

In order to construct the evaluation program, different kinds of assumptions have been done. For example, the eye ball is assumed to have a perfect "ball-in-socket" behavior, which eliminate small translations, and the line of sight is assumed to coincide with the eye's optical axis. These simplifications reduce the eye's movements to pure rotational movements. It also enables us to use the two-dimensional projection of the pupil's center point onto the image plane as an indicator for the eye's horizontal and vertical movements. The third dimension, i.e., torsion, is calculated performing cross-correlation with data sampled from the iris pattern before and after rotation. The size of the shift from the cross-correlation corresponds to the eye's torsion.

Since the eye is projected onto the image plane, the evaluated rotation is relative to the coordinate system of the image plane. However, absolute viewing angles are also important, which is why an algorithm on how to transform from the image plane coordinate system to a head-fixed coordinate system is presented. The algorithm is based on the *Rotation Offset Matrix* by Moore *et.al.*, (1996). After implementation, their formula showed unstability and a different approach on how to calculate the offset angles in the *Rotation Offset Matrix* is therefore derived. The torsion algorithm and the algorithm for calculating the *Rotation Offset Matrix* will be presented in detail.

2.2 Coordinate systems and eye movement

The analytical formulas, used in the program, are based on a head-fixed, right-handed coordinate system $\{x_h, y_h, z_h\}$ and a camera-fixed coordinate system $\{x_c, y_c, z_c\}$, as is shown in figure 1. The coordinate systems are oriented so that when the subject looks straightforward x_h and $-z_c$ coincide, y_h and x_c are parallel to the horizontal axis, z_h and $-y_c$ are parallel to the vertical axis, and x_c- y_c in the image plane (u,v), are parallel to the y_h - z_h plane. The eye ball's position when looking straightforward is defined as the eye's reference position, $\underline{\mathbf{P}}_{ref} = r_p\{1,0,0\}$, where r_p denotes the eye's radius at the center of the pupil. This is the point where the eye's line of sight and the x_h axis coincide. The origin of the coordinate system in the image plane originates from the projection of the pupil center while looking straightforward, i.e., in the eye's reference position.

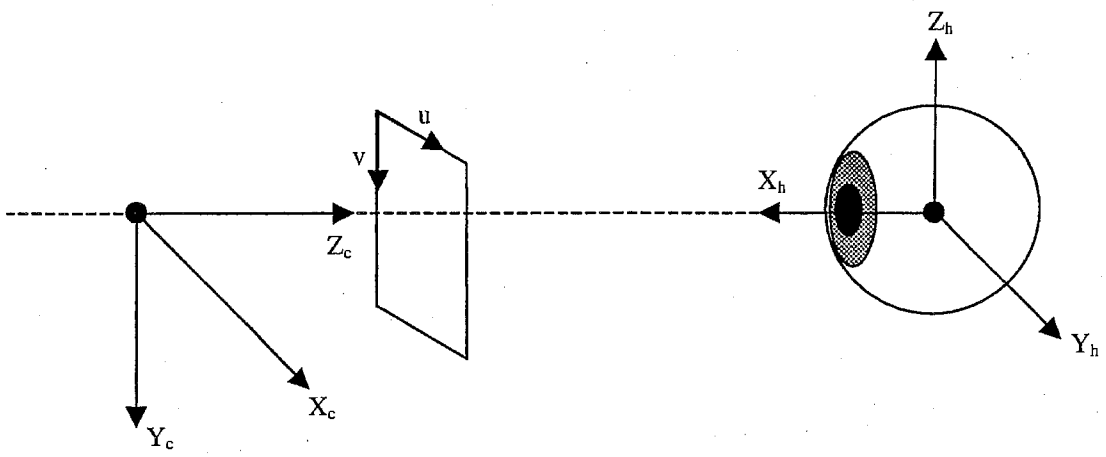


Figure 1. Coordinate system

Assuming a true "*ball-in-socket*" behavior for the eye enables us to express the eye's movements as pure rotations about the eye ball's center point. Using a rotation matrix $\mathbf{R}(\theta,\phi,\psi)$ makes it possible to describe any kind of rotation of the eye ball from the reference position $\underline{\mathbf{P}}_{ref}$ to a position $\underline{\mathbf{P}}_{eye}$. The new position of the pupil center, $\underline{\mathbf{P}}_{eye}$, after a rotation, is expressed by equation 1.

$$\mathbf{P}_{eye} = \mathbf{R}_{(\theta,\phi,\psi)} \bullet \mathbf{P}_{ref} \tag{1}$$

Equation 2 shows the elements of the rotation matrix $\mathbf{R}(\theta,\phi,\psi)$.

$$\mathbf{R}(\theta,\phi,\psi) = \begin{bmatrix} \cos\theta\cos\phi & \cos\theta\sin\phi\sin\psi-\sin\theta\cos\psi & \cos\theta\sin\phi\cos\psi+\sin\theta\sin\psi \\ \sin\theta\cos\phi & \sin\theta\sin\phi\sin\psi+\cos\theta\cos\psi & \sin\theta\sin\phi\cos\psi-\cos\theta\sin\psi \\ -\sin\phi & \cos\theta\sin\psi & \cos\phi\cos\psi \end{bmatrix} \tag{2}$$

Since the rotation is non-commutative, a change in the notation of the rotation order results in another rotation matrix. The algorithm uses a rotation sequence called Fick sequence [Fick, 1854], which is one of the most popular among eye movement researchers. For more detailed information on rotation matrices see [Berthold, 1997] or [Haslwanter, 1995].

2.3 New Implementation

2.3.1 Torsion

The third dimension of the eye's movement is the torsion, or, using the head-coordinate system, a rotation about the line of sight by ψ degrees. Instead of applying artificial landmarks on the eye, the iris pattern can be used for torsion estimation [Hatamian *et.al.*, 1983]. The iris pattern does not change when the eye rotates and will therefore contain the angular movement with respect to a reference position. The torsion is estimated by calculating the cross-correlation using two sequences of data sampled along the iris pattern. One arc is the reference arc and the other arc is continuously updated from the digitized images. Since the sampled arcs contain one-dimensional data, the cross-correlation is also one-dimensional and therefore easy to implement. A more sophisticated version of this algorithm was presented by [Haslwanter *et al.*, 1995], in which the shape of the track arc is corrected due to geometrical distorsion that occurs when looking at eccentric angles, i.e., a circle becomes elliptical. Applying this corrected algorithm reduces the torsion error significantly.

First, the pupil is segmented from the rest of the eye in the image by thresholding. Second, the pupil's center point in the image plane is calculated using a center of gravity function. A track arc's position and length on the iris is decided manually by the operator using a graphical interface. If s(n) and $s_r(n)$ denotes the iris data sampled along a track arc on radius r on the iris before and after torsion, the shift of the peak from the discrete one-dimensional cross-correlation function c(m), equation 3, corresponds to the torsion of the eye.

$$c(m) = \sum_{n=0}^{N-1} s(n)s_r(n+m) \qquad (3)$$

The iris luminance data is sampled with a 0.5° sample rate and the gray level values are calculated using a nonlinear interpolation function, which also works as a local low-pass filter. The cross-correlation uses the middle part of the sampled sequence that has been divided into three parts. For example, a track arc that is 60° and sampled with 0.5° contains 120 values which results in a sequence with 40 values used for cross-correlation. To achieve high accuracy in the peak detection, a least-square-fit of a second-order polynomial with seven points is used. In order to add further accuracy to the algorithm, several track arcs at different angels are used, which minimizes the error induced by pupil center point miscalculations, [Bos *et al.*, 1994], as well as stabilizing the calculated average value. Although all the track arcs are applied on the same image, the results from the cross-correlation will differ. However, sorting the results and discarding min and max values gives accurate torsion angle estimations.

Since video is used as storage medium, there will be a large amount of data to evaluate after each experiment. That makes the minimization of the evaluation time for each frame very important. As mentioned above, many track arcs increase the accuracy of the estimated value, but it also increases the evaluation time. Therefore, tests have also been done on how to estimate accurate angles, with an error of $<\pm 0.25^{\circ}$, using as few track arcs as possible for free viewing condition.

2.3.2 Rotation Offset Matrix

When looking at the screen, the three-dimensional movements of the subject's eyes are projected onto the camera's image plane, figure 2. The calculations will therefore initially be performed on data relative to the image plane's coordinate system instead of the head fixed coordinate system. Relative measuring is sufficient when for example examining the function of the vestibular system, but absolute viewing angles are preferable when investigation different view conditions. In order to express the viewing angles in absolute viewing angles i.e., viewing angles relative to the head fixed coordinate system, the image plane's viewing angles have to be transformed from the image plane's coordinate system to the head fixed coordinate system.

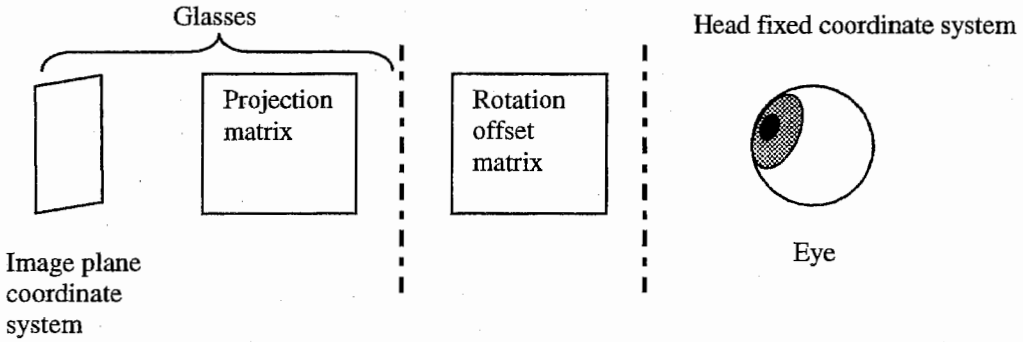


Figure 2. Components in the transformation from the eye to the image plane.

Equation 4 shows the relationship between the eye's position and the two-dimensional position in the camera's image plane. $\mathbf{R}_{offset}$ is the three-dimensional rotation matrix that contains the information on how the image plane is oriented relative to the head-fixed coordinate system. During experiment the subject keeps the head still in an upright position and looks straightforward on a vertical screen. The ideal case is when the image plane and viewing plane only differ in the vertical dimension, due to the reflection in the glasses half mirror. However, in the typical non-ideal case, the glasses will always slant a little depending on the test subject. Measuring the pupil's center point's coordinates in the image plane for five known viewing angles, the offset angles $\theta_{offset}, \phi_{offset}, \psi_{offset}$ can be calculated analytically. And when $\theta_{offset}, \phi_{offset}, \psi_{offset}$ in the $\mathbf{R}_{offset}$ are known, a point in the image plane can be expressed as viewing angles in the head fixed coordinate system. In equation 4, p denotes a

projection constant that contains information such as the camera's focal length, eye radius, different distances etc.

$$\mathbf{P}_{imageplane} = p \cdot \mathbf{R}_{offset} \bullet \mathbf{R}_{(\theta,\phi,\psi)} \bullet \mathbf{P}_{ref} \tag{4}$$

In our system, the formula from [Moore. *et.al.*, 1995] showed unstability when calculating the offset angles. The most likely reason is the use of a constant r_p (radius of the eye). The r_p is calculated, equation 5, and later used to calculate the offset angles. In equation 5 the y_ϕ and $y_{-\phi}$ -values correspond to v-coordinates in the image plane when looking at the calibration point at angle $\pm\phi_{cal}$ and y_r is the v-coordinate when looking at the reference point (straight ahead). The equations's parameters are very unstable since the nominators are small and therefore a shift of one pixel in the denominator will change the length of the eye radius drastically.

$$r_p^2 = \left(\frac{y_\phi + y_{-\phi} - 2\,y_r}{2\,(1-\cos(\phi_{cal}))} \right)^2 + \left(\frac{y_\phi - y_{-\phi}}{2\cos(\psi_{off})\sin(\phi_{cal})} \right)^2 \tag{5}$$

Instead of using the algorithm from Moore, (1995), a new set of equations are derived. The major difference is that we keep different parameters together in a constant, *scale constant (s)*. With a coordinate system depicted as in figure 1 and by using planar geometric projections, the pupil's center point in the image plane is given by equation 6a and 6b.

$$u = f\,\frac{\underline{\mathbf{p}}_{c1}}{\underline{\mathbf{P}}_{c3}} = f\,\frac{\underline{\mathbf{r}}_1 \bullet \underline{\mathbf{p}}_h + t_1}{\underline{\mathbf{r}}_3 \bullet \underline{\mathbf{p}}_h + t_3} \tag{6a}$$

$$v = f\,\frac{\underline{\mathbf{p}}_{c2}}{\underline{\mathbf{P}}_{c3}} = f\,\frac{\underline{\mathbf{r}}_2 \bullet \underline{\mathbf{p}}_h + t_2}{\underline{\mathbf{r}}_3 \bullet \underline{\mathbf{p}}_h + t_3} \tag{6b}$$

The rotation offset matrix's row vectors are $\underline{\mathbf{r}}_1$, $\underline{\mathbf{r}}_2$, $\underline{\mathbf{r}}_3$, and $\underline{\mathbf{p}}_h$ is the vector denoting the rotation of the eye and t is the translation of the camera coordinate system with respect to the head fixed system. Assuming $t_3 >> \underline{\mathbf{r}}_3 + \underline{\mathbf{p}}_h$ and $S = f/t_3$, we can write 6a and 6b as 7a and 7b.

$$u = s \cdot (\underline{\mathbf{r}}_1 \bullet \underline{\mathbf{p}}_h + t_1) \tag{7a}$$

$$v = s \cdot (\underline{\mathbf{r}}_2 \bullet \underline{\mathbf{p}}_h + t_2) \tag{7b}$$

In 7a and 7b the translation components are eliminated by subtracting two coordinates' values from each other, i.e., $\Delta u = u_1 - u_0$. The constant s is eliminated by dividing Δu with Δv, equation 8.

$$\begin{aligned}\Delta u &= s \cdot \underline{r}_1 \bullet \left(\underline{p}_{calpoint} - \underline{p}_{refpoint} \right) \\ \Delta v &= s \cdot \underline{r}_2 \bullet \left(\underline{p}_{calpoint} - \underline{p}_{refpoint} \right)\end{aligned} \quad \Rightarrow \quad \frac{\Delta u}{\Delta v} = \frac{\underline{r}_1 \bullet \left(\underline{p}_{calpoint} - \underline{p}_{refpoint} \right)}{\underline{r}_2 \bullet \left(\underline{p}_{calpoint} - \underline{p}_{refpoint} \right)} \quad (8)$$

This results in equation 9, where $\mathbf{R}_{offset}$ is the 3x3 rotation matrix $\mathbf{R}(\theta_{off},\phi_{off},\psi_{off})$ (eq. 4).

$$(\Delta v, -\Delta u, 0) \bullet \mathbf{R}_{offset} \bullet \left(\underline{p}_{calpoint} - \underline{p}_{refpoint} \right) = 0 \quad (9)$$

This equation system becomes nonlinear, and Newton-Raphson's method for nonlinear systems of equations is used to solve the angles [Vetterling *et. al.*, 1997]. Since Newton-Raphson's method requires an initial guess for each angle to be solved, a certain knowledge about the glasses orientation is required in order to enter good initial values. Newton-Raphson is also used when calculating the eye's viewing angles from an image point coordinate (u,v). In this case, the viewing angles in the image plane are used as initial guess values.

The algorithm for transformation from image plane coordinates to absolute viewing angles comprises two parts:

- Calculate rotation offset matrix and scale constants.

1) Input known calibration angles, their u and v coordinates, and initial guess of the three angles.
2) Newton-Raphson's method is used to solve the non-linear equation system.
3) Calculate the projection constants using the results from (2).

- Transform from image plane coordinates to head fixed coordinate system.

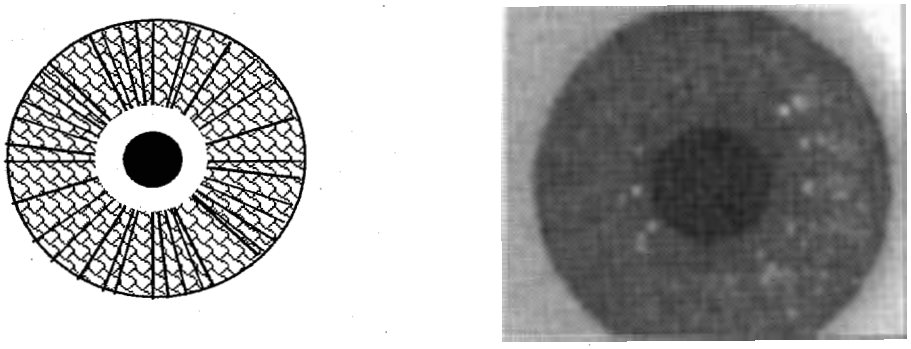
1) Calculates the eye's viewing angles from an image screen point (u,v) using Newton-Raphson's method.

3. Experiment

3.1 Torsion

The first part of the experiment was to verify the implemented cross-correlation algorithm, the second part investigated evaluation time, track arc optimization and accuracy for free-viewing conditions. In order to achieve an accurate value, cross-correlation can be performed with several track arcs and then calculating the average. However, assuming a five minutes experiment recording both eyes, 18 000 images will be generated. Therefore, we want to reduce our system's computation-time for each image while maintain high accuracy.

The two different eye patterns used to evaluate the cross-correlation algorithm are depicted in figure 3. The camera was a Teli CCD CS3500. The digitizer is a Matrox Pulsar frame grabber that delivers a 640x480 pixels 8 bits gray level image to a personal computer, a MMX Pentium 230 MHz. The patterns' horizontal, vertical and torsion angles were set by hand and the accuracy was 5' (0.08°). Due to the glasses's design, the view field is restricted to $<\pm 10^{\circ}$ for the vertical and horizontal viewing angles.



a) b)

Figure 3. Eye pattern, (a) simple and (b) artificial.

The torsion angle was calculated for five images in each position, and the pattern was rotated 0.5° at the time. The results from the simple pattern are shown in figure 4. The variation in the estimated torsion angles is small and the error is within our limit ($<\pm 0.25^{\circ}$). The cross-correlation algorithm is therefore assumed to work sufficiently well.

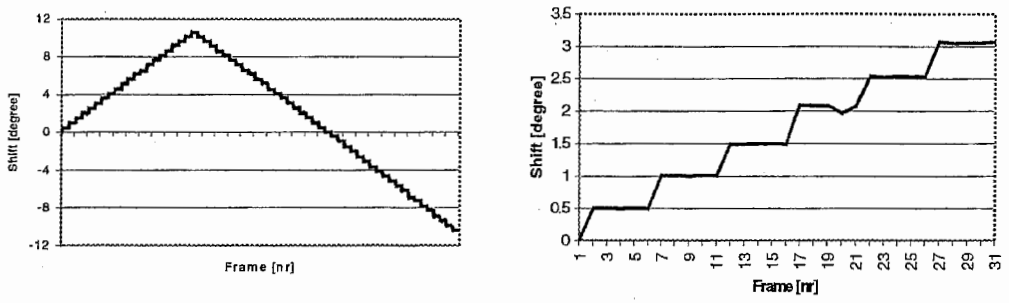


Figure 4. Result from cross-correlation (simple eye pattern).

The pattern was changed to the artificial eye, and the track arcs' dynamic gray level range decreased from approximately 180 to 20 values (figure 5). The results from the cross-correlation are shown in figure 6. In figure 6a, we see the average result from five arcs. The large variation in the calculated shift can be explained with bad image quality due to large dark current in the camera, low dynamic gray level range and a simple sampling method. In order to reduce the noise in the image, two video filters 3.2 MHz and 4.0 MHz, were used (fig. 6b). Methods to handle peaks as in 6b will be presented below.

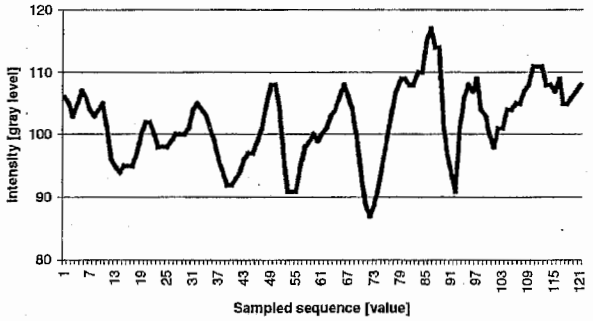
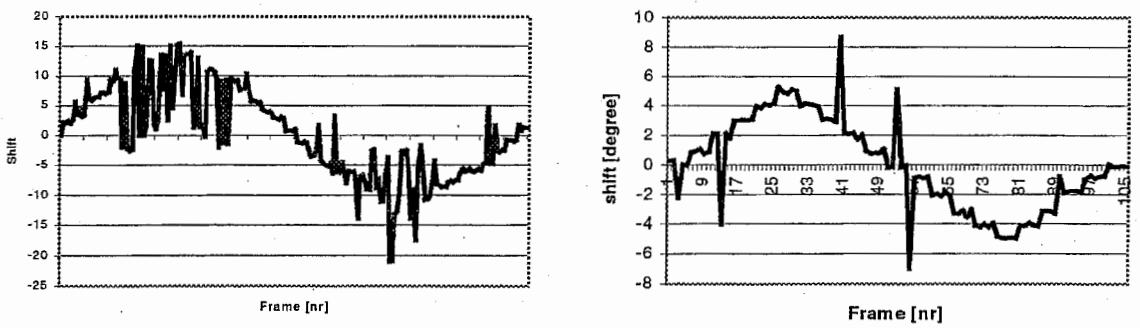


Figure 5. A sequence of sampled iris illumination, artificial eye.



a) b)

Figure 6. Results from cross-correlation, using an artificial eye. a) Average and b) with video filters (3.2 MHz and 4.0 MHz).

Figure 7 shows the result with video filters, and when rejecting the max and min values and calculating the average. The accuracy is improved and the test shows good results. Further improvements are expected with the new sampling function.

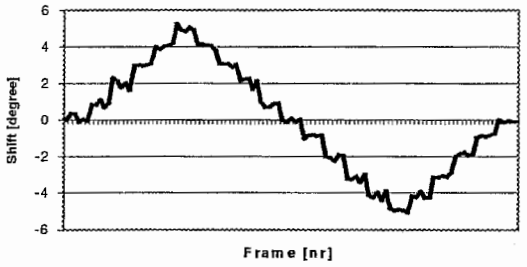


Figure 7. Result when using video filters and rejecting min and max values.

A new nonlinear method to sample along the track arcs was also implemented. The sampled gray level values are decided using the gray levels from the four surrounding pixel values ($C_{k,l}$, $C_{k+1,l}$, $C_{k,l+1}$and $C_{k+1,l+1}$) (fig. 8) and the distances between the four pixels and the sampling point (eq. 10).

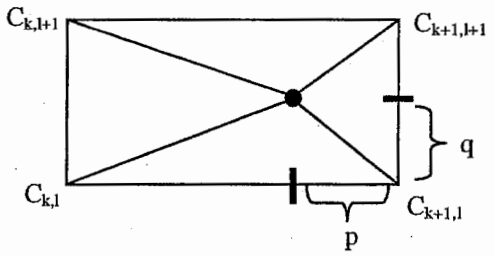


Figure 8. Non-linear sampling method.

$$\text{Gray value} = (1-p)(1-q)C_{k,l}+(1-p)qC_{k+1,l}+p(1-q)C_{k,l+1}+pqC_{k+1,l+1} \quad (10)$$

The second part of the test investigated evaluation time, accuracy, and optimization of the use of track arc.

1) *Correlation between the length of the track arc's and the estimated torsion values.* The artificial eye was positioned vertical = horizontal = 0°. 11 track arcs were evenly positioned on the reference image and the iris intensity was sampled along each arc. The torsion was set to 5°, a new image was digitized and intensity levels sampled. Cross-correlation was then performed using different lengths of the track arcs. Figure 9 shows a clear correlation between the arc's length and the accuracy. When performing cross-correlation with a sequence shorter than 40 values (i.e. a track arc less than 60°) the result from the cross-correlation is very unstable. However using 45 values or more gives an accurate result. In order to minimize the evaluation time we choose to use a sequence containing 50 values (i.e. a 75° track arc) in the following experiments.

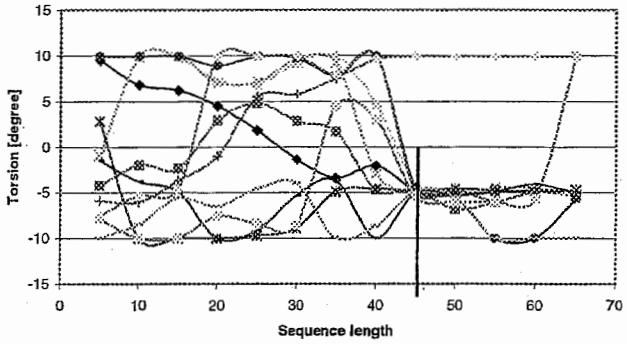


Figure 9. The torsion estimation due to different arc length.

2) Computation time.
The time for sampling along a 75° track arc 1000 times and performing cross-correlation required approximately 14sec. This means that the evaluation time for a 5 minute video sequence for both eyes takes 42 minutes when using 10 arcs, and 62 minutes when using 15 arcs (fig. 10). Since 90% of the time was spent during the sampling, the result clearly indicates that optimizing the use of the track arcs will save time.

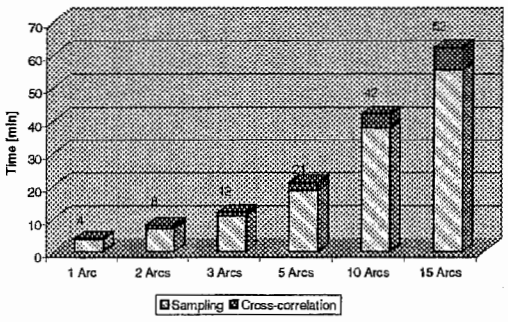
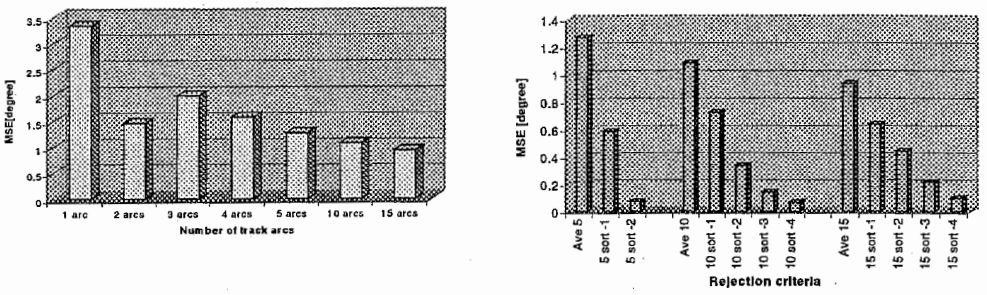


Figure 10. Time for sampling and cross-correlation with different numbers of track arcs (18 000 images = 5 min of video for both eyes).

3) *How does the amount of track arcs and rejection criteria affect the result?*
Even though data is sampled along track arcs from the same image, the results of the cross-correlation function vary for all the track arcs. This is because some positions of the iris are affected by other factors, such as reflection from illuminating light sources, eyelid interference or noise. The effect of these error sources can be minimized if the numbers of track arcs are increased in combination with (a) calculating the average value of the cross-correlation or (b) sorting the cross-correlation results for each image and rejecting a certain number of maximum and minimum values, and then calculate the average. In order to reduce induced torsion error due to center point miscalculation, at least two arcs, separated in the angular dimension, should be used (for further error analysis see [Bos *et. al.*, 1994] and [Hatamian *et. al.*, 1983]).

70 track arcs were randomly positioned on the iris, each 75° and the torsion set to +10 °, 0 ° and –10°, V=H=0°. The torsion was calculated using (a) average for groups of 2,3,4,5,10 & 15 arcs and (b) sorting and discarding 1,2,3 or 4 pair of values (containing values from the top and the bottom) and calculating the average on the remaining values using groups with 3,4,5,10 & 15 arcs. The track arcs were randomly chosen from the 70 arcs and in each case, the torsion was calculated for 10 different groups of track arcs.



a) b)

Figure 11. The MSE of torsion estimation using (a) average only and (b) different rejection criteria (torsion +10°, 0° & -10°). In (b) Ave 5 means average 5 track arcs and e.g. 5 sort –1 means that 5 track arcs are used and one pair of maximum and minimum values is discarded.

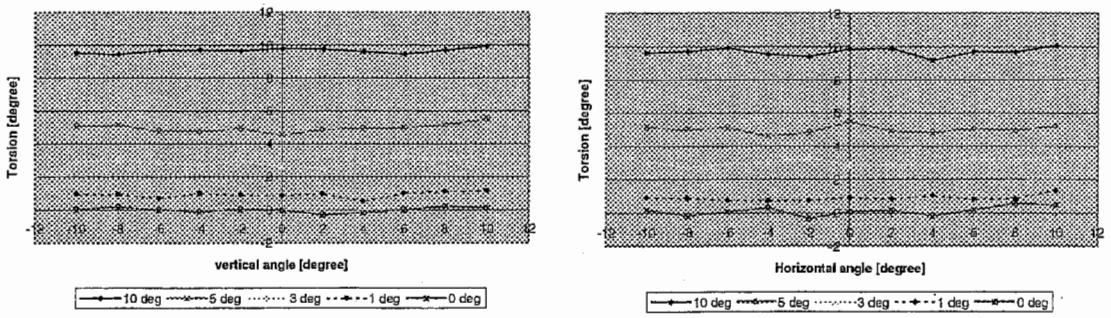
The results for the three angles (+10°, 0° & -10°) were combined and the results are shown in figure 11. Figure 11a shows the relation between the number of used track arcs and the accuracy of the estimated torsion when calculating the average. Even though 15 track arcs were used the mean square error (MSE) was four times (0.94°) our demanded accuracy (<±0.25°). Figure 11b shows the result when discarding minimum and maximum values. Both the accuracy and the precision were reduced largely compared to the results when calculating average only.

Since we have specified the error of our system to be within <±0.25°, looking in figure 11b we see that suitable estimation methods are 5 sort –2, 10 sort –3, 10 sort –4, 15 sort –3 and 15 sort –4. However, keeping the induced torsion error due to center point miscalculations in mind, the method 5 sort –2 is rejected. Furthermore, taking the sampling time into account, the methods using 15 track arcs were discarded. We therefore chose to use 10 track arcs in our algorithm. The method 10 sort –4 has a smaller error than 10 sort –3, but since the result from 10 sort –3 is based on an average from 4 values we assume this method to be more stable.

4) Different view angles.

In order to test the stability of the torsion estimation algorithm, 10 sort –3, when looking in various directions, the torsion angles were calculated for three new viewing conditions. The conditions were:

(a) Horizontal view angle = 0°, vertical view angle = -10° to +10 °, step 2°
(b) Vertical view angle = 0°, horizontal view angle = -10° to +10 °, step 2°
(c) Vertical view angle = horizontal view angle = 0°, torsion = -10° to +10 °, step 1°



a) b)

Figure 12. The estimated torsion angles when looking in various directions when using the method 10 sort -3. (a) Horizontal view angle = 0°, vertical view angle = -10° to +10 °, step 2° and (b) Vertical view angle = 0°, horizontal view angle = -10° to +10 °, step 2°.

Figure 12 shows the relationship between the torsion angle and different view angles for viewing condition (a) and (b) (see above). As we see in figure 12a and 12b, accurate and precise torsion estimation was maintained for viewing angles within ±10°. The large variation in figure 12b originates from reflexes from the illumination sources.

3.2 Rotation Offset Matrix

In order to perform initial tests on the transformation algorithm, image plane coordinates were measured for a ±5° view field, in steps of 1°. The coordinates were used for calculating the offset-angles, as well as checking the image plane coordinates expressed in absolute viewing angles. In order to investigate the stability of the transformation algorithm, a ±1 pixel noise in the u and v coordinates was added. Furthermore, the influence of the scale constant (SC) was also investigated, i.e., separated (s_u and s_v) or common ($s^2 = (s_u)^2+(s_v)^2$). As expected, the use of separated scale constants gave better results. Table 1 shows some results from the transformation when using separated scale constants and common scale constant. The effect on the viewing angles due to noise (variations) in the center point's coordinates is shown in figure 13.

Image coordinates	Common SC	Separated SC	Set view angle
391,202	3, -1.90	2.83, -1.95	3, -2
244,116	-5.33, -4.65	-5.17, -5.0	-5, -5
293,311	-1.27, 3.82	-1.15, 4.01	-1, 4

Table 1. Transformation from image plane coordinates to absolute view angles.

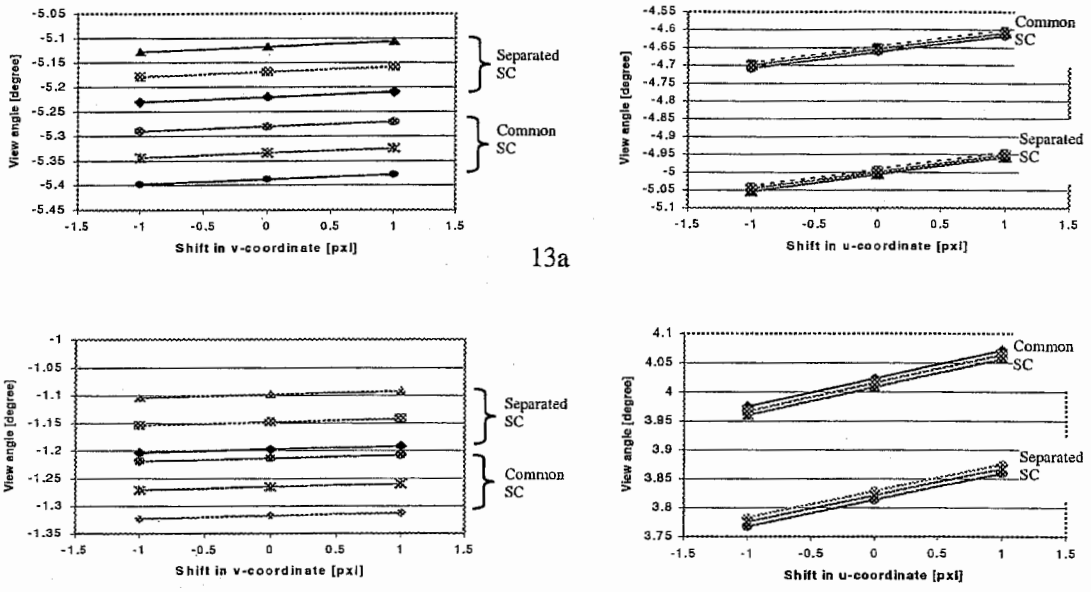


Figure 13. The result of the transformation for two points in the image plane and also the influence of ±1 pixel noise. 13a) is (-5°, -5 °) and 13b) is (-1 °, 4 °).

In figure 13a the viewing angles are set to (-5°,-5°). The left diagram shows the horizontal viewing angles, where the three upper lines belong to the use of separated projection constant and the three lower when using common projection constant. The right diagram shows the vertical viewing angles (Note! The projection constants' positions in the right diagram are the opposite compared to the left diagram). In figure 13b, the viewing angles were set to (-1°, 4°). The induced noise, ±1 pixel, in the image plane's coordinates (u,v) induced an error less than ±0.02° in the horizontal viewing angle while the vertical viewing angle's error was less than ±0.05°. The error for the two viewing angles when combining the results from,

(a) keep u-coordinate still and vary the v-coordinate
(b) keep v-coordinate still and vary the u-coordinate

Are <±0.17° for the horizontal viewing angle and <±0.05° for the vertical viewing angle.

3.3 Comments

Performing a test to check the computation-time using different numbers of track arcs clearly indicated that the number of track arcs was an important factor (fig.10). Therefore, further experiments were performed in order to optimize the length of the track arcs as well as the amount of arcs to be used due to computation-time and accuracy. Our results show that there exists a break point regarding the track arc length, and that there also exists an optimized amount of track arcs to be used when

estimating the torsion. We have decided to use 10 sort-3 as default but the operator will be able to change the criteria manually.

The transformation algorithm from image coordinates to absolute viewing angles is stable. The error is $<\pm 0.17^{\circ}$ for the horizontal viewing angle, and $<\pm 0.05^{\circ}$ for the vertical viewing angle. So far, the tests have included small viewing angles. The next step will therefore be large viewing angles, as well as investigating how the absolute viewing angles differ from the viewing angles calculated in the image plane.

4. Discussion

4.1 Introduction

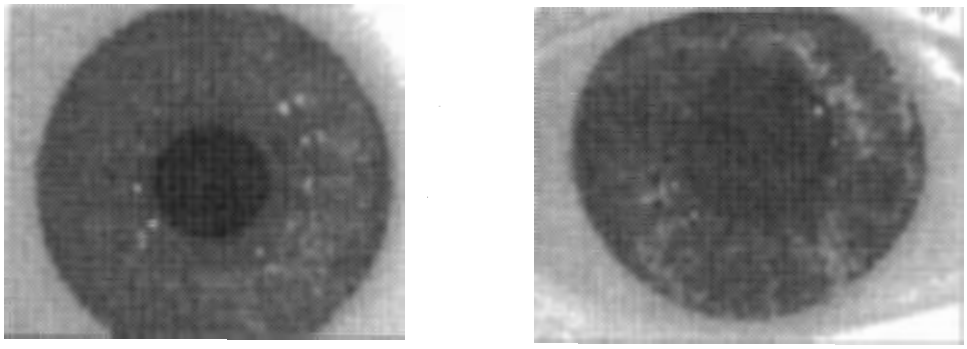
The work on developing the eye tracking system has been focused on implementing and testing the algorithms for measuring horizontal, vertical and rotational eye movements. The implementation of the torsion algorithm and the transformation matrix from the image plane to the head-fixed coordinate system shows good results. This is the first step, which means that the algorithms have to be further tested and maybe improved in order to get a program that is easy to use, with accurate and reliable results. The system has an error less than $\pm 0.25^{\circ}$ for the estimated torsion angles, for a $\pm 10^{\circ}$ view field when using 10 track arcs and rejecting 3 pairs of min and max values. The transformation algorithm from image coordinates to absolute viewing angles is stable and the error is less than $\pm 0.17^{\circ}$ for the horizontal viewing angle and less than $\pm 0.05^{\circ}$ for the vertical viewing angle.

We now have the necessary parts to construct a system that calculates the eye's three dimensional movements using video images. Testing the different modules separately is much easier than testing the whole system. However, the tests so far show that our implemented algorithms are stable if the quality of the images is good. Therefore, the next problem to solve, when the modules are put together, will be how to get stable images.

4.2 Problems - Image quality

The eye's movements can be categorized as rotation and translation. However in order to simplify the algorithms, an ideal "*ball-in-socket*" behavior is assumed (i.e. translation is eliminated) and only the three rotational movements horizontal, vertical and rotation (i.e. torsion) exists. The viewing field of the eye is less than $\pm 30^{\circ}$ and the accuracy on the measured angles should be larger than $\pm 0.25^{\circ}$ for free viewing conditions, and even higher when torsion only is measured. The evaluation could be performed rather easily if the image conditions were good. However, the video images contains different kind of noise such as reflection from the illumination source, shadings, unfocused area, interference from eye lashes, etc.. All these problems make it difficult to make a system that is easy to use, non-invasive, fast and accurate.

The image quality is the single most important factor for accuracy and reliability of the estimated values. The present glasses have IR-diodes mounted in semi-circles that illuminate the eye balls However, this is not a suitable method for our system since the reflexes (white dots, fig. 12) from the illumination source will ruin the sampled data from the iris. Therefore, a new pair of glasses are constructed where the IR-diodes (λ_{peak}=750nm) positions are moved, and the viewing angles are increased. One alternative solution is to track the reflexes and not to use data sampled along a track arc ruined by reflections. However, this is time consuming and not as reliable as moving the position of the diodes.



a) b)

Figure 12. Iris pattern for a) artificial eye and b) human eye. Note! The white dots are reflexes from the illumination source

The image quality also depends on the focal depth (*f*) of the camera. The focus of the camera is set when the eye is looking straightforward. When the eye is moving around the image of the eye becomes unfocused and this adds an error to mainly the torsion evaluation. This error can be minimized with a camera with a larger focal depth (today *f*=12mm). The size of the dynamic gray level range also affects the cross-correlation. The present camera delivers an 8 bit gray level value. Increasing it to 10 bits would make the cross-correlation more stable for noise in the image. The current range of the dynamic gray levels is 20-25 values, and this is expected to increase when we start testing on human eyes (see fig. 12b). In order to reduce the effect of the shading of the eye ball that occurs when looking in different view angles, the cross-correlation will be performed with filtered data. The present sampling algorithm of the iris works as a high pass filter, but we also need a low pass filter, [Faes *et.al.*, 1994].

4.3 Future work

- Absolute viewing angles

The first set of verifications of the transformation algorithm from image plane coordinates to absolute viewing angles has been done, and shows promising results. However, further tests have to be done to verify if it works for larger viewing angles as well. Several projection constants might be needed depending on the size of the viewing angles.

- Ellipse detection

An ellipse detection algorithm that calculates the center point using fragmented ellipses are necessary to handle the problem with eye lashes.

- The wave length of the illumination source

Different wavelengths give different iris patterns, which affect the range of the dynamic gray level.

- New track arcs

Today the arc is sampled along a circle, but in order to change the shape when looking in eccentric angles the shape of the track arc has to be calculated using, for example, bicubic splines.

- Video interface

A method for loading images from the video into the PC's RAM when evaluating the angles will be implemented.

4.4 Conclusions

Thanks to the development of small CCD-cameras, computer imaging has become very interesting to use when tracking the eye's movements. However, the video technique is not only good, it also causes new problems that have to be solved. The solutions are often based on different computer imaging algorithms and/or 3D geometry. Our modules are based on known algorithms, but some modifications were required.

Our work has been concentrated on the basic modules, and studies have been performed on one module at a time. All the modules evaluate viewing angles that are within the error limit, less than $\pm 0.25^{\circ}$, for free viewing conditions. The next major step is a complete system check using the artificial eye, i.e., an image is digitized and all the three angles calculated. When the modules are put together the biggest problem will be the stability in the evaluation. The stability depends mainly on the image quality. There are many ways to improve the images using software algorithms. However, since computer imaging takes time and the enhanced images will never be better than the in-data, the best solution is to look at the image recording devices and the digitizer. Therefore, a new pair of glasses with a larger view field and the positions of the illumination sources moved will be used. The present glasses use IR-diodes, but we are also interested in testing other wavelengths in order to get a more distinguishable iris pattern.

5. Reference

Berthold Max, "Development of a Three Dimensional Eye Measurement System based on Computer Imaging", ATR Technical Report TR-H-220, June, 1997.

Bos E. Jelte and De Graaf Bernd, "Ocular Torsion Quantification with Video Images", IEEE Transactions on Biomedical Engineering,Vol. 41, No 4, 1994.

Clarke H. Andrew, Teiwes Winfried and Scherer Hans, "Video-oculography - an method for measurement of three-dimensional eye movements", Oculomotor Control and Cognitive Process, Elsevier Science Publishers B.V (North Holland), 1991.

Enright J.T., "Saccadic anomalies: Vergence induces large departures from ball and socket behavior", Vision Research, Vol. 24, 1984.

Faes J.C., Govaerts H.G., Tenvoorde B.J. and Rompelman O., "Frequency synthesis of digital filters based on repeatedly applied unweighed moving average operations", Medical & Biological Engineering & Computing, Nov 1994.

Fick. A, "Die Bewegungen des menschlichen Augapfels", Zeitschrift fur rationelle Medizin, 4, 109-128, 1854.

Groen Eric, Bos E. Jelte, Nacken F. M. Peter and De Graaf Bernd, " Determination of Ocular Torsion by Means of Automatic Pattern Recognition", IEEE Transactions on Biomedical Engineering,Vol. 43, No 5, 1996.

Haslwanter Thomas, "Mathematics of Three-dimensional Eye Rotations", Vision Research, Vol. 35, 1995.

Haslwanter Thomas and Moore T. Steven, "A Theoretical Analysis of Three-Dimensional Eye Position Measurement Using Polar Cross-Correlation", IEEE Transactions on Biomedical Engineering,Vol. 42, No 11, 1995.

Hatamian Mehdi and Anderson J. David, "Design Considerations for a Real-Time Ocular Counterroll Instrument", IEEE Transactions on Biomedical Engineering, No 5 , 1983.

Von Helmholtz H., " Handbuch der Physiologischen Optik", Leipzig, 1867, English translation: Southall J. P. "Treatise on physiological optics", New York, Optical Society of America, Reprinted 1962.

Moore T. Steven, Haslwanter Thomas, Curthoys S. Ian and Smith T. Stuart, "A Geometric Basis for Measurement of Three-Dimensional Eye Position Using Image Processing", Vision Research Vol 36, No 3, 1996.

Moore T. S., Curthoys S. I. and McCoy S.G., "VTM - an image-processing system for measuring ocular torsion", Compute Methods and Programs in Biomedicine, Vol 35, 1991.

Ott Dietmar, Gehle Frank and Eckmiller Rolf,"Video-oculographic measurement of 3-dimensional eye rotations", Journal of Neuroscience Methods, vol 35, 1990.

Young L. R. and Sheena D, "Survey of eye movement recording methods.", Behavior Research Methods and Instrumentation, Vol. 7, 1975.

Vetterling T. W, Press H. W, Teukolsky A. S and Flannery P. B, " Numerical Recipes", Cambridge University Press, 2nd edition, 1997.