A probabilistic approach for the estimation of angle kappa in infants

Dmitri Model
University of Toronto
dmitry.model@gmail.com

Moshe Eizenman
University of Toronto
eizenm@ecf.utoronto.ca

Abstract

This paper presents a probabilistic approach for the estimation of the angle between the optical and visual axes (angle kappa) in infants. The approach assumes that when patterned calibration targets are presented on a uniform background, subjects are more likely to look at the calibration targets than at the uniform background, but it does not require accurate and continuous fixation on presented targets. Simulations results show that when subjects attend to roughly half of the presented targets, angle kappa can be estimated accurately with low probability (<1%) of false detection. In experiments with five babies who attended to the calibration target for only 47% of the time (range from 26% to 70%), the average difference between repeated measurements of angle kappa was 0.04 ± 0.31°.

Keywords: Eye Tracking, Remote Gaze Estimation, Calibration-Free, Infants, Uncooperative Subjects.

1 Introduction

State-of-the-art Remote Eye Gaze Tracking (REGT) systems that use a stereo pair of video cameras can estimate the center of curvature of the cornea, c, and the optical axis of the eye, ω, without any user calibration [Guin et al. 2007; Shih and Liu 2004]. However, since human gaze is not directed along the optical axis, but rather along the visual axis, a subject-dependent angle between the optical and visual axes (angle kappa) should be estimated. In cooperative subjects, a one-point calibration procedure can be used to estimate angle kappa. During this calibration procedure, the subject is required to look at a single calibration target, and it is assumed that the visual axis intersects with the center of the target. However, in studies with uncooperative subjects (e.g., young children or mentally challenged people), even such a simple procedure cannot be performed reliably.

Several “user-calibration-free” approaches for remote gaze estimation have been suggested in the literature. Shih et al. assumed that the visual and optical axis coincide (angle kappa is zero) [Shih et al. 2000], while Nagamatsu et al used a midpoint of intersection of the optical axes of both eyes as approximation for the Point-of-Gaze (PoG) [Nagamatsu et al. 2010]. However, such approaches result in significant subject-dependent gaze-estimation errors [Model and Eizenman 2010a]. Model and Eizenman suggested an automated method for the estimation of angle kappa [Model and Eizenman 2010a; Model and Eizenman 2010b]. However, for accurate results, such method requires a very accurate REGT system (RMS error in the estimation of the optical axis < 0.1°) or observation surface that consists of multiple planes (please see [Model and Eizenman 2010a; Model and Eizenman 2010b] for details), which limits the practical utility of such approach.

This paper presents an automatic calibration procedure that is designed for infants. It is based on inherent tendency of babies to look at patterned stimuli presented on a uniform background [Fantz et al. 1975]. Therefore, if several stimuli (animated images) are presented at different locations on a uniform background, there is a higher likelihood that infants will look at the presented stimuli than at any other point. The probabilistic approach to the estimation of angle Kappa is based on this observation. During the presentation of the visual stimuli the suggested Model-Eizenman algorithm (M-E algorithm, hereafter) continuously estimates the angle between the optical and visual axes under the assumption that the infant fixates the presented point (see Section 2.1). At the end of the presentation, the resulting estimates of angle Kappa form a distribution from which a determination of the true angle Kappa can be obtained. If there is no correlation between the infant’s gaze and the positions of the visual stimuli, the distribution of the measured angles will tend to be uniform. On the other hand, if the infant did look at the presented targets from time to time, there will be a dominant peak in the distribution of the measured angles that corresponds to an angle between the optical and visual axes. Therefore, the M-E algorithm consists of two stages: Estimation and Validation. The M-E algorithm is described in detail in Section 2.2. Since the M-E procedure is probabilistic in nature, the confidence measure in the outcome of M-E is derived in Section 2.3. Experimental results with five infants are presented in Section 3. Finally, discussion and conclusions are presented in Section 4.

2 Method

In the following analysis, all points are represented as 3-D column vectors (bold font) in a right-handed Cartesian World-fixed Coordinate System (WCS).

Figure 1 presents a simplified schematic diagram of the eye [Model and Eizenman 2010a]. The line connecting the center of curvature of the cornea with the center of the pupil defines the optical axis. The line connecting the center of the fovea with the center of curvature of the cornea defines the visual axis or the line-of-gaze. The average magnitude of the angle between the optical and visual axes is 5°. This angle has both horizontal
(nasal) and vertical components, which exhibit considerable inter-personal variation [Carpenter 1977].

Two coordinate systems are used in this paper. The first coordinate system is a stationary right-handed Cartesian World Coordinate System (WCS) with the origin at the center of the main display, the $X_w$-axis in the horizontal direction, the $Y_w$-axis in the vertical direction and the $Z_w$-axis perpendicular to the display (see Figure 1). This coordinate system is used to define PoG relative to fixed objects in space (e.g., computer monitor).

**Figure 1. A simplified schematic diagram of the eye (not to scale). The optical axis, $\omega$ (the axis of symmetry of the eye) passes through the center of curvature of the cornea, $c$, and the center of the pupil, $p$. The visual axis of the eye (the line-of-sight), $v$, connects the fovea (the region of the highest visual acuity on the retina) with $c$.**

The second coordinate system is a non-stationary right-handed Cartesian Eye Coordinate System (ECS), which is attached to the eye, with the origin at the center of curvature of the cornea, the $Z_{eye}$ axis that coincides with the optical axis of the eye and $X_{eye}$ and $Y_{eye}$ axes that, in the primary gaze position, are in the horizontal and vertical directions, respectively. The $X_{eye}$-$Y_{eye}$ plane rotates according to Listing’s law [Helmholtz 1924] around the $Z_{eye}$ axis for different gaze directions.

In the ECS, the unknown 3D angle between the optical and the visual axes of the eye can be expressed by the horizontal, $\alpha$, and vertical, $\beta$, components of this angle (see Figure 1). The unit vector in the direction of the visual axis with respect to the ECS, $v_{ECS}$, is then expressed as:

$$v_{ECS}(\alpha, \beta) = \begin{bmatrix} -\sin(\alpha) \cos(\beta) \\ \sin(\beta) \\ \cos(\alpha) \cos(\beta) \end{bmatrix}$$

The unit vector in the direction of the visual axis with respect to the WCS, $v$, can be expressed as:

$$v(\alpha, \beta) = R v_{ECS}(\alpha, \beta)$$

where $R = [i_{ECS}, j_{ECS}, k_{ECS}]$ is the rotation matrix from the ECS to the WCS and $i_{ECS}$, $j_{ECS}$, and $k_{ECS}$ are the unit vectors in the direction of the $X_{eye}$, $Y_{eye}$, and $Z_{eye}$ axes with respect to the WCS, respectively. Note that $i_{ECS}$, $j_{ECS}$, and $k_{ECS}$ are independent of $\alpha$ and $\beta$ and $R$ can be readily calculated from the orientation of the optical axis of the eye and Listing’s law [Guestrin and Eizenman 2010; Helmholtz 1924].

Because the visual axis goes through the center of curvature of the cornea, $c$, the PoG, $g$, in the WCS is given by:

$$g = c + \mu v(\alpha, \beta) = c + \mu R v_{ECS}(\alpha, \beta)$$

where $\mu$ is a line parameter, proportional to the distance from the eye to the monitor.

### 2.1 Estimation of angle kappa for a single time sample

As was shown in [Guestrin and Eizenman 2008], under the assumption that the subject fixates on a calibration target, the horizontal, $\alpha$, and vertical, $\beta$, components of angle kappa (the angle between the optical and visual axes) can be estimated as follows.

Let’s denote, $\textbf{z}$, the position of the calibration target in 3-D. Then, the unit vector in the direction of the line that connects the center of curvature of the cornea, $c$, and $\textbf{z}$ is given by

$$\textbf{d} = \frac{\textbf{z} - c}{\|\textbf{z} - c\|}$$

If one assumes that the target is being looked at, according to (3), the visual axis is, in fact, aligned with $\textbf{d}$:

$$v(\alpha, \beta) = \textbf{d} = \frac{\textbf{z} - c}{\|\textbf{z} - c\|}$$

Using (2), the visual axis in the ECS is given by

$$v_{ECS}(\alpha, \beta) = R^T v(\alpha, \beta) = R^T \frac{\textbf{z} - c}{\|\textbf{z} - c\|}$$

Since $v_{ECS}(\alpha, \beta)$ can also be expressed by Eq. (1):

$$\begin{bmatrix} -\sin(\alpha) \cos(\beta) \\ \sin(\beta) \\ \cos(\alpha) \cos(\beta) \end{bmatrix} = R^T \frac{\textbf{z} - c}{\|\textbf{z} - c\|}$$

The angles $\alpha$ and $\beta$ can be reconstructed directly from (7) as follows

$$\alpha = \arctan \left( \frac{\textbf{R}^T \frac{\textbf{z} - c}{\|\textbf{z} - c\|} \cdot \hat{x}}{\textbf{R}^T \frac{\textbf{z} - c}{\|\textbf{z} - c\|} \cdot \hat{z}} \right)$$

$$\beta = \arcsin \left( \frac{\textbf{R}^T \frac{\textbf{z} - c}{\|\textbf{z} - c\|} \cdot \hat{y}}{\|\textbf{R}^T \frac{\textbf{z} - c}{\|\textbf{z} - c\|} \cdot \hat{y}\|} \right)$$
where $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ are unit vectors in the directions of $X_w$, $Y_w$ and $Z_w$ respectively, and $\cdot \cdot$ denotes dot product.

### 2.2 Estimation of angle kappa with uncooperative subjects

As shown in the previous section, the horizontal, $\alpha$, and vertical, $\beta$, components of angle kappa (the angle between the optical and visual axes) can be easily estimated if the subject fixates on a calibration target. With infants, unfortunately, it is impossible to control when and for how long they will look at the calibration target. However, based on the natural tendency of humans to look at visual stimuli with distinctive features [Fantz et al. 1975; Koch and Ullman 1985], one can assume that there is higher likelihood that infants will fixate on the stimulus rather than anywhere else on the computer monitor.

Based on the above assumption, the following strategy can be used to estimate angle kappa. A sequence of visual stimuli is presented at different locations on a computer screen (Figure 2 (a)). The angles $\alpha_t$ and $\beta_t$ for each time instant $t$, are calculated under the assumption that the infant is fixating on the presented stimulus using equations (8) and (9). Then, a histogram of $\alpha_t$, $\beta_t$, $t=1..N$ is created. Example of such a histogram for a sequence of 9 stimuli (50 estimates/stimuli for a total of $N=450$ estimates) that were presented to an 8 months old infant is shown in Figure 2(b).

If the subject fixates exclusively on the presented stimuli, similar values of $\alpha_t$ and $\beta_t$ should be computed for all $N$ measurements (i.e., all $\alpha_t$ or $\beta_t$ should fall in a single bin of the histogram). If, on the other hand, the subject fixates on points that are uncorrelated with the presented stimuli, the distribution will become more uniform and at the limit the histogram will resemble a uniform distribution. In Figure 2(b), the histograms of $\alpha_t$ and $\beta_t$, $t=1..N$ is created. Example of such a histogram for a sequence of 9 stimuli (50 estimates/stimuli for a total of $N=450$ estimates) that were presented to an 8 months old infant is shown in Figure 2(b).

Estimation of angle kappa in infants is summarized in Algorithm 1.

#### Algorithm 1

**Estimation of angle kappa**

1. Present a sequence of visual stimuli with limited spatial extent at pseudo-random locations on a computer monitor.
2. For each time instant $t$, calculate $\alpha_t$ and $\beta_t$ using equations (8) and (9), respectively, under the assumption that the subject is looking at the presented stimuli.
3. Find time instances for inliers in $\alpha$, $\beta$:
   i. Partition all $\{\alpha_t | -10^\circ < \alpha_t < 10^\circ, t=1..N\}$, (angle kappa is known to be within $\pm 5^\circ$ for adult subjects [Carpenter 1977; Hashemi et al. 2010], up to $\pm 8^\circ$ in infants [Schaeffel 2002]) into bins of width $W$. Estimates of $\alpha_t$, outside the range of $\pm 10^\circ$ are inconsistent with the physiological properties of the human eye and can be discarded.
   ii. Select the bin with the largest count of data points.
   iii. Calculate the average value, $\bar{\alpha}$, of all the data points in the selected bin and two adjacent bins.
   iv. Inliers are $\alpha$-s that satisfy $|\alpha_t - \bar{\alpha}| < W$, i.e., $t_\alpha = \{t | |\alpha_t - \bar{\alpha}| < W \}$
4. Find all the time instances for inliers in $\beta$, $t_\beta$ using the procedure described in Step 3.
5. The time instances for inliers in both $\alpha$ and $\beta$ are given by: $t_{in} = t_\alpha \cap t_\beta$ and the number of all inliers is $N_{in}$.
6. The estimated $\alpha$ and $\beta$ are given by
   
   $\bar{\alpha} = \frac{1}{N_{in}} \sum_{i \in \alpha_{in}} \alpha_i$  \hspace{1cm} (10)
   
   $\bar{\beta} = \frac{1}{N_{in}} \sum_{i \in \beta_{in}} \beta_i$  \hspace{1cm} (11)

#### 2.3 Confidence measure for the outcome of the M-E procedure

Algorithm 1 analyzes the distribution of angular offsets between $\omega$ (the optical axis of the eye) and a line $d$ (a line that connect $c$ and the center of the target). When the subject looks at the target, line $d$ is aligned with the visual axis of the eye and the offset is equal to angle Kappa. Algorithm 1 selects an offset ($\bar{\alpha}$ and $\bar{\beta}$) that results in the maximum number of inlier samples, $N_{in}$. In this section we estimate the probability that Algorithm 1 will select/estimate $\bar{\alpha}$ and $\bar{\beta}$ from the recorded data points that are not associated with fixations on the presented targets (i.e., the probability of false detection).

In Algorithm 1, the angular offsets, $\alpha_t$, and $\beta_t$, between $\omega$ and the target being presented, $d_t$, are measured for each time instant $t=1..N$. However, $\alpha_t$, $t=1..N$ (as well as $\beta_t$, $t=1..N$) are not independent because of temporal correlation between consecutive $\omega_t$ and possible spatial correlation if a subject re-visits the same salient point several times. To remove such correlations, let’s define a cluster, $\Gamma$, as a set of $\alpha_t$ and $\beta_t$ that are within $\pm \Phi$ of the
center of such cluster during the presentation of the \( k \)-th target (\( t = T_1 \ldots (T_{k-1} + 1) \)), where \( T_k \) is the time instant when the \( k \)-th target was presented. Under this definition, if the subject fixes the same point during several consecutive time samples or/and revisits the same point several times during the presentation of the \( k \)-th target, this will still constitute a single cluster. If \( \Phi \) is large enough (e.g., equal to the half of the angular extent of the target presented), the positions of the clusters during the presentation of the \( k \)-th target can be assumed to be independent. Furthermore, since the positions of the targets are randomized (\( \text{d}_i \), \( \text{i} = 1 \ldots \), are independent random variables), \( \Gamma_{\text{m}} \), \( m = 1 \ldots M \), are also independent across all targets. Even if a subject keeps fixating the same point on the screen, i.e., \( \text{o}_i \) is constant, the angular offsets between \( \text{o}_i \) and \( \text{d}_i \) are not, because \( \text{d}_i \) are random and independent of \( \text{o}_i \).

Let’s assume that if a subject does not pay attention to the stimuli, \( \Gamma_{\text{m}} \) is equally likely to fall anywhere within \( \pm \Omega \) around \( \text{d}_i \). That is, \( \Gamma_{\text{m}} \), \( m = 1 \ldots M \), are independent uniformly distributed random variables within \( \pm \Omega \). If during the experiment, there were \( M \) clusters in total with \( M_{\text{n}} \) ‘inlier’ clusters (such clusters that correspond to inlier samples as identified by the Algorithm 1), the probability of such scenario (to have \( M \) overlapping clusters out of \( M \)) under the assumption that the subject does not pay attention to the stimuli can be estimated using Monte-Carlo simulations. In other words, the purpose of this simulation is to estimate the probability of having \( M_{\text{n}} \) overlapping clusters, when the subject’s gaze is independent of the position of the presented stimuli.

### Table 1: Probability of having \( M_{\text{n}} \) ‘inlier’ clusters out of \( M \) clusters, under the assumption of a uniform distribution of clusters within \( \pm 10^\circ \) of \( \text{d} \).

<table>
<thead>
<tr>
<th>( M_{\text{in}} )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>0.966</td>
<td>0.034</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>0.888</td>
<td>0.111</td>
<td>0.001</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>0.802</td>
<td>0.195</td>
<td>0.003</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>0.683</td>
<td>0.311</td>
<td>0.006</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>0.569</td>
<td>0.416</td>
<td>0.014</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>0.444</td>
<td>0.528</td>
<td>0.028</td>
<td>0.001</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>0.343</td>
<td>0.617</td>
<td>0.039</td>
<td>0.001</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>0.251</td>
<td>0.695</td>
<td>0.054</td>
<td>0.001</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0.175</td>
<td>0.746</td>
<td>0.078</td>
<td>0.002</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>0.119</td>
<td>0.772</td>
<td>0.105</td>
<td>0.004</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>0.081</td>
<td>0.787</td>
<td>0.128</td>
<td>0.005</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>0.043</td>
<td>0.788</td>
<td>0.160</td>
<td>0.008</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>0.028</td>
<td>0.773</td>
<td>0.189</td>
<td>0.011</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>0.015</td>
<td>0.730</td>
<td>0.242</td>
<td>0.013</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>0.008</td>
<td>0.702</td>
<td>0.271</td>
<td>0.018</td>
<td>0.001</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>0.005</td>
<td>0.651</td>
<td>0.320</td>
<td>0.024</td>
<td>0.001</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>0.002</td>
<td>0.611</td>
<td>0.356</td>
<td>0.029</td>
<td>0.001</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>0.001</td>
<td>0.356</td>
<td>0.406</td>
<td>0.035</td>
<td>0.002</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>0</td>
<td>0</td>
<td>0.300</td>
<td>0.451</td>
<td>0.046</td>
<td>0.003</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 1 provides the results of the Monte-Carlo simulation using parameters that are similar to those used in the experiments de-

### 3 Experiments

The experiments were carried out with the stereo-camera system shown in Figure 3. The study was carried out with 5 healthy infants (6, 8, 9, 15 and 16 months old). During the experiments, infants were seated on their parents’ lap, with their heads supported by the parents’ hands. Infants’ head was approximately 85 cm from the computer monitor.
Nine 2cm x 2cm animated images (that is equivalent to 1.35° x 1.35° of visual field at a distance of 85 cm) were presented sequentially at different positions on the monitor and 450 estimates were recorded at a rate of 30 estimates per second. The measurement was repeated twice, to assess repeatability [Bland and Altman 1986].

The average difference (± Standard Deviation, SD) between two independent measurements of $\alpha_L$, $\alpha_R$, $\beta_L$ and $\beta_R$ was $0.04 \pm 0.31^\circ$ and the 95% limits of agreement for repeated measurements were ±0.61° (see Figure 4).

To test the methodology relative to a known benchmark (fixation on a calibration target), we have performed a one-point calibration procedure with 22 adult subjects. The adults were instructed to look at one of the animated images that were used during the experiments with the infants and 50 estimates of the direction of the optical axis were obtained. The angles $\alpha_L$, $\alpha_R$, $\beta_L$ and $\beta_R$ were calculated as described in Algorithm 1. The procedure was repeated twice for each subject. The average difference (± Standard Deviation, SD) between two independent measurements of $\alpha_L$, $\alpha_R$, $\beta_L$ and $\beta_R$ in adults was $0.03 \pm 0.44^\circ$ and the 95% limits of agreement for repeated measurements were ±0.87° (note that the size of the calibration target is 1.35° x 1.35° and subjects can fixate anywhere within the target). The slightly better repeatability with infants can be explained by the fact that by using more calibration targets with infants, differences in fixation positions within targets between repeated trials were averaged over more targets (i.e., the standard deviation of the differences between repeated estimates was reduced) which improved slightly the repeatability of the estimation procedure.

In the experiments with 5 infants, the average percentage of time that infants looked at the stimuli was 47% (range from 26% to 70%). The average difference (± Standard Deviation, SD) between two independent measurements of $\alpha_L$, $\alpha_R$, $\beta_L$ and $\beta_R$ in infants was $0.04 \pm 0.31^\circ$. The average difference (± Standard Deviation, SD) between two independent measurements of $\alpha_L$, $\alpha_R$, $\beta_L$ and $\beta_R$ in adults with one-point calibration procedure was $0.03 \pm 0.44^\circ$. That is, the accuracy of M-E procedure is similar to the accuracy of one point calibration.

Even though the M-E procedure does not guarantee that angle Kappa can be estimated (i.e., when subjects do not look at the targets), the probability that the M-E procedure will provide a false estimate (an estimate that is associated with fixations that are not correlated with the calibration targets) can be assessed using the confidence measure derived in Section 2.3.

**Acknowledgements**

This work was supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) and Canadian Institute of Health Research (CIHR) and in part by scholarships from NSERC and the Vision Science Research Program Award (Toronto Western Research Institute, University Health Network, Toronto, ON, Canada).

**References**


