

The mean point of vergence is biased under projection

Xi Wang
TU Berlin
Germany

Kenneth Holmqvist
Regensburg University, Germany
Nikolaus-Kopernikus University
Poland

Marc Alexa
TU Berlin
Germany

The point of interest in three-dimensional space in eye tracking is often computed based on intersecting the lines of sight with geometry, or finding the point closest to the two lines of sight. We first start by theoretical analysis with synthetic simulations. We show that the mean point of vergence is generally biased for centrally symmetric errors and that the bias depends on the horizontal vs. vertical noise distribution of the tracked eye positions. Our analysis continues with an evaluation on real experimental data. The estimated mean vergence points seem to contain different errors among individuals but they generally show the same bias towards the observer. And it tends to be larger with an increased viewing distance. We also provided a recipe to minimize the bias, which applies to general computations of gaze estimation under projection. These findings not only have implications for choosing the calibration method in eye tracking experiments and interpreting the observed eye movements data; but also suggest to us that we shall consider the mathematical models of calibration as part of the experiment.


Keywords: eye movement, eye tracking, fixations, gaze, vergence, bias, calibration

Introduction

Humans tend to direct both of their eyes at roughly the same point in 3D space. Binocular saccades and smooth pursuit between objects in a 3D scene often exhibit vergence, which means that two eyes move in opposite directions (Cassin et al., 1984) for fixation to coincide with the intended object. In other words, vergence is the movement of both eyes towards or away from each other, depending on the relative change from the previous to the current target. It is often assumed that the fixation points of the two eyes are perfectly aligned but it has been shown that the eyes first diverge before they converge at the gaze point during fixations (Collewijn et al.,

1995, 1997). Studies on binocular coordination of eye movements during reading show that fixation points of two eyes vary during reading and disparity in both horizontal and vertical directions were observed (Liversedge et al., 2006b; Nuthmann and Kliegl, 2009). There is considerable variation among participants ability to fixate the same point in depth, depending on their eye dominance and squinting, or even strabismus (when the weak eye is off-target). In addition, measurements of pupil positions with video-based eye-trackers are very sensitive to variations in pupil dilation (Hooge et al., 2019), which leads to uncertainty over measurements of vergence.

Many approaches have been proposed to estimate the direction of gaze for each eye in physical space, based on recorded pupil positions by eye-tracking devices. Using these gaze vectors, it is possible to reconstruct the gaze point on real three-dimensional stimuli by intersecting one or both rays with the fixated object in space, assuming its geometry is known (Hammer et al., 2013; Maurus et al., 2014; Wang et al., 2017b). Alternatively, we can

Received December 19, 2018; Published September 9, 2019.
Citation: Wang, X., Holmqvist, K. & Alexa, M. (2019). The mean point of vergence is biased under projection. *Journal of Eye Movement Research*, 12(4):2.
Digital Object Identifier: 10.16910/jemr.12.4.2
ISSN: 1995-8692
This article is licensed under a [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). 

attempt to find the point where the two vectors intersect with each other in space (Hennessey and Lawrence, 2009; Maggia et al., 2013; Pfeiffer and Renner, 2014; Gutierrez Mlot et al., 2016; Pfeiffer et al., 2016), but in 3D space two gaze vectors typically do not intersect.

However, even if the observer experiences looking at a point in space with both eyes, the eye rays provided by the eye-tracker contain error, for a variety of reasons:

- data from eye-trackers have an inaccuracy (systematic error) and introduce imprecision (variable error, or noise) onto the signal (Holmqvist et al., 2011; Wang et al., 2017a; Niehorster et al., 2018).
- the inaccuracy is not constant, but varies with pupil size (Drewes, 2014) and quantization of the CR in the eye camera (Holmqvist et al., 2019).
- ideally, human gaze direction is controlled to bring the object into the fovea centrals (Atchison and Smith, 2000), which has a non-negligible extent of 1.5–2°.
- it is well-known that in binocular vision, many observers have a dominant eye which is more accurately directed towards the target (in about 70 % of the cases, the right eye) and a weaker eye, which may be considerably off-target (Erkelens et al., 1989; Collewijn et al., 1995, 1997), which is called binocular disparity (strabismus in extreme cases).
- the resulting unknown and likely non-linear function mapping from tracked pupil and CR centers in the eye video to lines of sight is approximated using low order polynomials (Essig et al., 2006; Holmqvist et al., 2011; Cerrolaza et al., 2012).

For these reasons, the two projected eye rays produced by eye-trackers are generally skewed and have no common intersection point in space. In order to calculate a point that approximates the expected intersection of the rays, the most natural and commonly employed solution is to compute the point that has the smallest distance to both rays in 3D. Here we call this point the vergence point. We derive the necessary equations for this computation next and then use it for simulating the reconstruction of vergence points in the presence of systematic (accuracy) and variable (precision) errors. We then develop the mathematical description of inaccuracy and imprecision of gaze vectors, that are used to simulate the effect on estimated intersection point of inaccuracy (offsets) and imprecision (noise). Hooge et al. (2019) present human data of vergence error as an effect of inaccuracy from changes in pupil dilation. In this work, we focus on the effect of imprecision and present recorded human vergence data that validate the simulation on noise. Finally, we present a method to better estimate the intersection

of eye rays and reconstruct the position of the fixated object in 3D, given noisy vergence data. Note that the whole analysis applies to eye tracking not only in space but also on flat surface, as long as the underlying projective mapping is used in the model.

Part 1: Mathematical model and simulation

When recording binocular gaze in 3D, the two gaze vectors can be thought of as originating in the centers of the two eyes of the observer. Two vectors in three-dimensional space are generally skew, i.e. they have no common (intersection) point. For the reasons mentioned in the introduction, the two gaze vectors are commonly far from intersecting. In order to assign a point of interest given two gaze vectors that have no intersection point, the dominant strategy is to compute the point in 3D space that has the smallest sum of squared distances to the two gaze lines. Here we show how to compute distances of a point to a line by using the formulation of projector, and then how to find the point of interest.

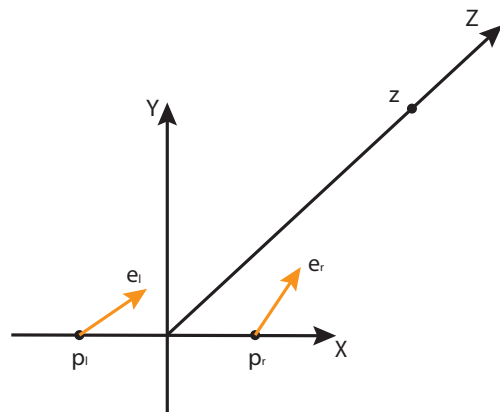


Figure 1 Geometric setup. \mathbf{p}_l and \mathbf{p}_r are the centers of the left and right eyes, and \mathbf{e}_l and \mathbf{e}_r are gaze vectors of unit length, i.e. the eye rays. We want to calculate the point \mathbf{z} with the minimal distance to both rays.

Choose the coordinate system such that centers of the eyes are displaced symmetrically from the origin along the first coordinate direction. The up direction defines the second coordinate direction and target is placed on the third direction pointing away from the observer. This means the centers of the left and right eyes are

$$\begin{pmatrix} -a \\ 0 \\ 0 \end{pmatrix} = p_l, \begin{pmatrix} a \\ 0 \\ 0 \end{pmatrix} = p_r. \quad (1)$$

Here, and in the following, boldface lower-case characters denote column vectors in Euclidean space. Define an up direction as $\mathbf{u} = (0, 1, 0)^T$. We refer to the first coordinate axis as horizontal and the second one (i.e. along the up direction) as vertical. Objects of interest located at $\mathbf{z} \in \mathbb{R}^3$ are displaced from the eyes mostly along the third coordinate axis, i.e. $\mathbf{z} = (\approx 0, \approx 0, z)$. Then the normalized (unit-length) vectors from eye to interest point are

$$e_l = \frac{\mathbf{z} - p_l}{\|\mathbf{z} - p_l\|}, e_r = \frac{\mathbf{z} - p_r}{\|\mathbf{z} - p_r\|}, \quad (2)$$

where the subscripts l and r refer to the left and right eye. Normalization ensures that the vectors have unit length:

$$1 = \|e_{l,r}\| = \|e_{l,r}\|^2 = \langle e_{l,r}, e_{l,r} \rangle = e_{l,r}^T e_{l,r}. \quad (3)$$

Note that latter notation for the inner product follows from $\mathbf{e}_{l,r}$ being a column vector and the usual conventions for matrix multiplication; we use this notation in what follows.

Given only the positions of the eyes and unit gaze vectors, we want to compute the point closest to both eye rays. One way to do so is measuring the squared distances to the rays and finding the point that minimizes them. Let's first consider a ray through the origin. We define it by specifying a unit direction vector $\mathbf{v} \in \mathbb{R}^3, \|\mathbf{v}\| = \mathbf{v}^T \mathbf{v} = 1$. Then the points on the ray in the direction \mathbf{v} are given by $\lambda \mathbf{v}$, where $\lambda \in \mathbb{R}$ is a scalar parameterizing the ray.

Consider the (symmetric) matrix $\mathbf{V} = \mathbf{I} - \mathbf{v}^T \mathbf{v}$. Here \mathbf{I} denotes the 3×3 identity matrix (we generally use uppercase bold-face letter for matrices) and $\mathbf{v}^T \mathbf{v}$ is an outer product, following directly from the common rules for matrix multiplication:

Notation	Meaning
$\mathbf{p}_{l,r} \in \mathbb{R}^3$	the position of left/right eye
$a \in \mathbb{R}$	half of the distance between two eyes
$\mathbf{u} \in \mathbb{R}^3$	the up direction vector
$\mathbf{z} \in \mathbb{R}^3$	object of interest
$z \in \mathbb{R}$	third element of the object of interest \mathbf{z}
$\mathbf{e}_{l,r} \in \mathbb{R}^3$	normalized eye ray directions
$\mathbf{E}_{l,r} \in \mathbb{R}^{3 \times 3}$	projectors of the gaze vectors
$d \in \mathbb{R}$	Euclidean distance between two points
$\mathbf{e}_{l,r} \in \mathbb{R}^3$	variable error vector of each eye
$\mathbf{e}'_{l,r} \in \mathbb{R}^3$	noisy eye ray directions
$\eta_{l,r} \in \mathbb{R}$	horizontal noise
$\nu_{l,r} \in \mathbb{R}$	vertical noise
p	Gaussian distribution
$\Sigma_{l,r} \in \mathbb{R}^{3 \times 3}$	Covariance matrix
$\sigma \in \mathbb{R}$	standard deviation

$$\mathbf{V} = \mathbf{I} - \begin{pmatrix} v_0 \\ v_1 \\ v_2 \end{pmatrix} (v_0 \ v_1 \ v_2) = \begin{pmatrix} 1 - v_0^2 & -v_0 v_1 & -v_0 v_2 \\ -v_0 v_1 & 1 - v_1^2 & -v_1 v_2 \\ -v_0 v_2 & -v_1 v_2 & 1 - v_2^2 \end{pmatrix}. \quad (4)$$

Multiplication of this matrix with any point $\lambda \mathbf{v}$ on the ray yields

$$\mathbf{V} \lambda \mathbf{v} = (\mathbf{I} - \mathbf{v} \mathbf{v}^T) \lambda \mathbf{v} = \lambda (\mathbf{v} - \mathbf{v} \mathbf{v}^T \mathbf{v}) = \lambda (\mathbf{v} - \mathbf{v}) = \mathbf{0}, \quad (5)$$

while multiplying with a vector $\mathbf{w} \in \mathbb{R}^3$ of arbitrary length but orthogonal to \mathbf{v} , i.e. $\mathbf{v}^T \mathbf{w} = \mathbf{0}$, yields

$$\mathbf{V} \mathbf{w} = (\mathbf{I} - \mathbf{v} \mathbf{v}^T) \mathbf{w} = \mathbf{w} - \mathbf{v} \mathbf{v}^T \mathbf{w} = \mathbf{w}. \quad (6)$$

So the matrix V annihilates components in direction v and leaves directions orthogonal to v unchanged. It is commonly called a *projector* for the direction v . Similarly define the projectors for the gaze vectors E_l, E_r .

Multiplying a point $x \in \mathbb{R}^3$ from the left and the right, i.e. $x^T V x$, results in taking the inner product of the component orthogonal to the vector or, in other words, measuring the squared distance of x to

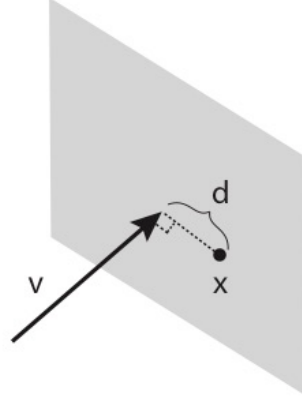


Figure 2 Scalar variable d represents the distance from point x to vector v that lies in the plane perpendicular to v .

the line along v through the origin.

If the line is not through the origin, all we need to know is a point p on the line. Then we translate everything so that p is in the origin, meaning we get the squared distance of x to the line along v through p as

$$(x - p)^T V (x - p). \quad (7)$$

With this way of measuring the distance to a ray, the sum of the squared distances to the eye rays for any point x in space can be written as:

$$d^2(x) = (x - p_l)^T E_l (x - p_l) + (x - p_r)^T E_r (x - p_r). \quad (8)$$

To find the point in space that minimizes this sum of squared distances compute the gradient of this function (with respect to x)

$$\nabla d^2 = 2E_l(x - p_l) + 2E_r(x - p_r), \quad (9)$$

and set it to zero:

$$(E_l + E_r)x = E_l p_l + E_r p_r = (E_r - E_l)p_r. \quad (10)$$

In this way the point of interest x is defined as the solution of a 3×3 linear system. The system has a unique solution as long as the sum $E_l + E_r$ is non-singular. Each of the two matrices $E_{l,r}$ has a one-dimensional kernel: the ray direction $e_{l,r}$ is an eigenvector with zero eigenvalue. If the two gaze vectors are parallel, then the

projectors are identical and $E_l + E_r = 2E_l = 2E_r$ is singular. This is quite intuitive, as there is no unique point with smallest distance to two parallel lines.

If the eye rays are not parallel, however, the sum $E_l + E_r$ is non-singular. This is also geometrically intuitive as there is a unique point minimizing the squared distances to the two lines; and this fact can be proven rigorously (Tian and Styan, 2001, Corollary 2.5). Thus, the point of interest is defined as

$$e_l \neq \lambda e_r, \lambda \in \mathbb{R} \Rightarrow x = (E_l + E_r)^{-1}(E_r - E_l)p_r. \quad (11)$$

Eye ray errors

Firstly, we introduce variable error (imprecision, noise) ϵ_l, ϵ_r into the eye rays, with separate horizontal (η) and vertical (v) noise as well as separate noise for left and right eyes:

$$\epsilon_l = \begin{pmatrix} \eta_l \\ v_l \\ 0 \end{pmatrix}, \epsilon_r = \begin{pmatrix} \eta_r \\ v_r \\ 0 \end{pmatrix}. \quad (12)$$

While errors are usually represented in terms of angular deviation (i.e. radians), for small enough values the linear approximation $\sin \varphi \approx \varphi$ is very good and adding the error vectors to the eye ray vectors has the same effect as rotating the eye rays. Including renormalization this yields:

$$e'_l = \frac{e_l + \epsilon_l}{\|e_l + \epsilon_l\|}, \quad e'_r = \frac{e_r + \epsilon_r}{\|e_r + \epsilon_r\|}. \quad (13)$$

With this model we simulate how noise affects the computation of the vergence point. Following the results in Niehorster et al. (in prep) (which show that noise in eye trackers are mostly Gaussian distributed), we use a zero-mean Gaussian distribution, i.e.

$$p(\epsilon_{l,r}) = \left| 2\pi \Sigma_{l,r} \right|^{-1/2} \exp\left(-\frac{1}{2} \epsilon_{l,r}^T \Sigma_{l,r}^{-1} \epsilon_{l,r}\right). \quad (14)$$

with the horizontal and vertical deviations being uncorrelated (meaning noise distribution in each direction varies independently)

$$\Sigma_{l,r} = \begin{pmatrix} \sigma_{\eta_{l,r}}^2 & 0 & 0 \\ 0 & \sigma_{v_{l,r}}^2 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (15)$$

In this setup, error vectors can be generated by simply drawing the components $\eta_{l,r}, v_{l,r}$ independently from univariate normal (Gaussian) distributions with zero mean and standard deviation $\sigma_{\eta_{l,r}}, \sigma_{v_{l,r}}$. univariate normal (Gaussian) distributions with zero mean and standard deviation $\sigma_{\eta_{l,r}}, \sigma_{v_{l,r}}$.

A point of interest \mathbf{z} defines the unbiased eye rays $\mathbf{e}_{l,r}$, which align with the lines of sight. To generate points of vergence in the presence of precision error, we draw error vectors $\boldsymbol{\varepsilon}_{l,r}$ (i.e. the added noise), modify the eye rays accordingly and reconstruct the vergence point using the linear system above. The python script given in the appendix does exactly this. It outputs the mean vergence and offers graphical output, such as the one shown in Figure 3. We chose three standard deviations of 0.16° , 0.5° and 1.5° for the Gaussian distributions of noise (which are motivated by experimental data described later). 0.5° is the commonly used threshold of the average calibration accuracy of each marker, 0.16° corresponds to the average precision of fixations and 1.5° corresponds to difficult eye tracking situations for example when observers wear glasses. The simulation shows a large error in direction of depth – this is quite intuitive given the short base line relative to the distance to the object.

More severely, the mean vergence point appears to be biased systematically towards or away from the observer. This is clearly visible when the noise distribution has different variance in horizontal and vertical directions as shown in Figure 3: if variance is larger in the horizontal direction, the mean vergence points shifts away from the viewer; if variance is larger in the vertical direction, the mean vergence point shifts towards the viewer. Table 1 shows detailed simulation results when the distance between observer and the fixation target is considered as another variable. We perform the simulation when the fixation target is placed at three different distances, namely 50 cm, 70 cm and 110 cm. The further away the target is, the larger the bias of the mean vergence point contains. Note that this bias is non-linear due to the underlining projective relation. Within each condition, the

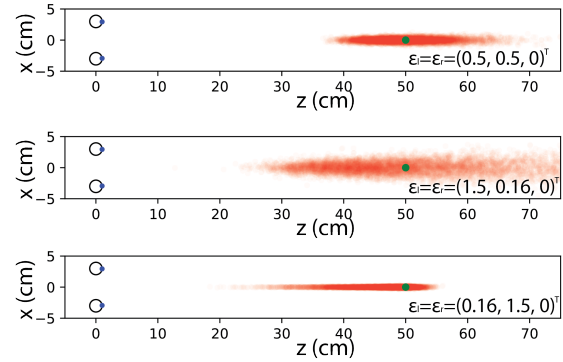


Figure 3 Vergence points from eye rays distorted with zero-mean normally distributed errors (i.e. noise). Distance between two eyes is 6 cm and the object is at the distance of 50cm. Top row: the noise distribution has standard deviation of 0.5 degrees in both horizontal and vertical direction. Despite the relatively small error, the variation in depth is quite large. Middle row: standard deviation in vertical direction is only 0.16, but in horizontal direction it is 1.5. The noise in horizontal direction gets larger and the mean estimated point of vergence is shifted towards larger depths. Bottom row: standard deviation is 1.5 degrees in vertical direction and 0.16 in horizontal direction. Horizontal noise is small, but the mean estimated point of vergence shifts significantly towards the viewer. Same number of samples is drawn in each simulation. 10,000 points are current shown in the figure.

bias is mainly in depth but large variation in horizontal direction leads to large bias while the same amount of variation in vertical direction corresponds to smaller bias.

Secondly, although it is not the focus of this paper, our noise formulation could also be used to investigate the effect of systematic errors (inaccuracy, offset) by introducing constant offset for the eye positions in Equation 12. As shown in the simulation results in Figure 4, unsurprisingly, the resulting errors, i.e. the distances between the estimated mean vergence point and the true target point, are larger when inaccuracy is introduced compared to when only noise is added to data. Furthermore, systematic offsets in the horizontal direction shift the mean of the estimation no matter whether the offsets in the two eyes converge or diverge. Meanwhile, vertical systematic offsets in the same direction lead to larger estimation errors without shifting the distribution mean much. Again, the estimated mean is biased towards the observer when systematic offsets in opposite vertical directions are introduced, similar to the bias we observed when only noise was introduced where the estimation of

Table 1. Simulation errors with target placed at three different distance 50 cm, 70 cm and 110 cm. σ_x and σ_y represents the standard deviation in horizontal and vertical directions measured in degrees. \bar{p}_z is the mean position of estimated vergence points in depth. Averaged distance errors and standard deviations are reported in cm with x represents the horizontal direction, y the vertical direction, and z in depth. Notice that the small baseline of distance between the eyes leads to large bias when the target point is further away. On average, the errors of the estimations are 27% (50 cm), 35% (70 cm), and 49% (110 cm) in percentage of the corresponding distances.

Target distance (cm)	σ_x (deg)	σ_y (deg)	\bar{p}_z (cm)	\overline{error} (cm)	\overline{std} (cm)	\overline{error}_x (cm)	\overline{error}_y (cm)	\overline{error}_z (cm)	\overline{std}_x (cm)	\overline{std}_y (cm)	\overline{std}_z (cm)
50	1.5	1.5	38.6	14.0	8.1	0.6	0.6	13.9	0.5	0.5	8.2
50	1.5	0.16	41.3	12.8	8.6	0.6	0.6	12.8	0.5	0.6	8.7
50	0.16	1.5	36.8	13.2	2.5	0.06	0.5	13.2	0.04	0.4	2.6
70	1.5	1.5	49.5	25.4	17.1	0.7	0.7	25.4	0.7	0.7	17.1
70	1.5	0.16	56.1	23.9	28.6	0.8	0.09	23.9	0.9	0.1	28.7
70	0.16	1.5	45.8	24.2	4.7	0.1	0.7	24.2	0.05	0.5	4.7
110	1.5	1.5	63.5	54.7	44.8	1.0	1.0	54.6	1.1	1.1	44.8
110	1.5	0.16	82.9	58.4	101.4	1.3	0.1	58.4	2.1	0.2	101.4
110	0.16	1.5	59.0	51.0	9.1	0.09	0.9	51.0	0.07	0.7	9.1

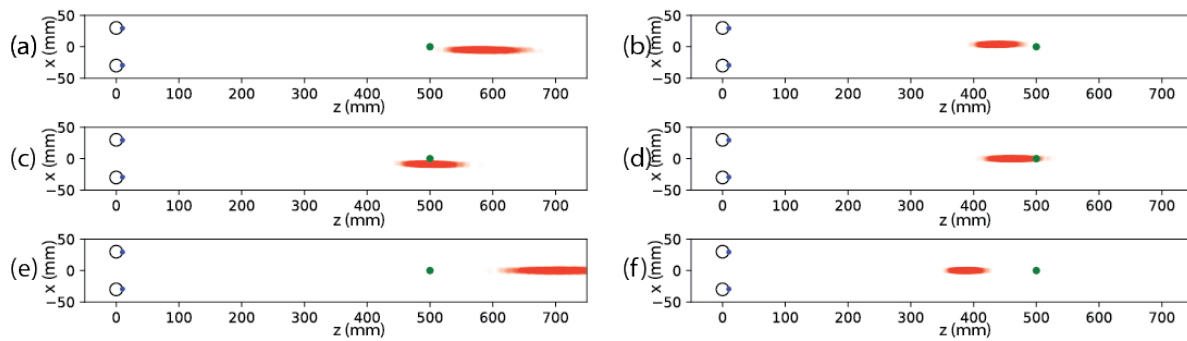


Figure 4 Effects of systematic offsets (inaccuracy) on the distribution of the vergence estimations. The distance between the two eyes is 6 cm and the object is at a distance of 50 cm. Each eye ray is distorted with zero-mean Gaussian noise with standard deviation of 0.16 degree. In the plots, we see the effects when a systematic offset (inaccuracy) of 1 degree is applied, not an uncommon effect size from a small change in pupil dilation. (a) Left eye is horizontally rotated by one degree outwards. (b) Left eye is horizontally rotated towards the right eye direction by one degree (similar distributions are obtained when right eye is horizontally rotated left by one degree). (c) Both left and right eyes are rotated in the same direction by one degree. Figures (d) and (e) show the distribution of vergence points when both eyes are systematically rotated inwards and outwards by one degree respectively. Systematic offsets in vertical direction lead to larger estimation errors without obvious shifts of the distribution mean, except when the offsets in opposite vertical directions are applied as shown in (f). Similar to the previous simulation of variable error (noise), vertical offsets lead to a bias towards the observer, i.e. the estimated mean vergence point is closer to the observer than the target.

the mean vergence point is always closer to the observer than the true target point.

Qualitative analytical analysis of bias

In the section above (and examples shown in Figure 3), we use a simple numerical simulation to visualize the distribution of computed points of interest given an assumption on the probability distribution of the noise. This simulation suggested that the mean vergence point is biased, shifted away from the true target point, depending on the standard deviation of the noise distribution in horizontal vs. vertical directions. We wish to analyze this behavior analytically. For this we attempt to compute the mean (or expected) vergence point for a given probability distribution of noise. As we will see, the general problem is difficult to approach. Yet, by assuming the geometric situation exhibits symmetry (see Appendix A for details) we are able to show that the trend we have observed in the numerical simulation holds qualitatively for wide classes of practically relevant scenarios.

The expected value for a discrete noise distribution would be the sum of values multiplied with the respective probabilities for the input parameters. In the continuous case, this sum turns into an integral over the product of the computed value and the probability distribution for the input parameters.

The projectors $E'_{l,r}(\varepsilon_{l,r})$ are generated from the noisy eye rays $e'_{l,r}$. Assuming a probability distribution $p'_{l,r}(\varepsilon_{l,r})$ the expected intersection point is:

$$x = \iint (E'_l(\varepsilon_l) + E'_r(\varepsilon_r))^{-1} (E'_l(\varepsilon_l)p_l + E'_r(\varepsilon_r)p_r) p_l(\varepsilon_l)p_r(\varepsilon_r) d\varepsilon_l d\varepsilon_r \quad (16)$$

This integral, in general, cannot be treated analytically. However, note that the resulting mean is a linear function of the individual points of vergence. This means we can generate the mean of several instances of the distribution and then integrate over these local means. We will make use of this property in the following.

From the perspective of applications, we are mostly interested in understanding the bias with respect to different distances in depth. This allows us to concentrate on geometric configurations that exhibit symmetry. First, we assume the fixated object is on the symmetry line and at unit distance, i.e. $\mathbf{z} = (0, 0, 1)^T$. Note that by setting the interocular distance $\mathbf{p}_r - \mathbf{p}_l = 2\mathbf{a}$ this still considers

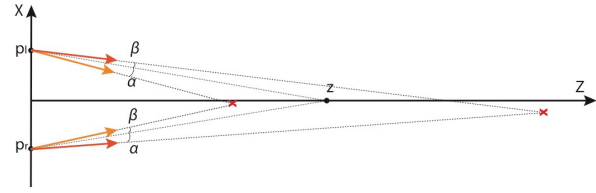


Figure 5 Illustration of two intersection points with inverted pair of errors α, β . Inverted pair of errors leads to an intersection point in the opposite directions; however, the mean vergence point is biased to be further away from the target point \mathbf{z} . In this illustration, $\alpha = 5^\circ$ and $\beta = -3^\circ$, where positive angle corresponds to the clockwise direction.

arbitrary distances because geometrically it makes no difference if we change the distance of the target or the distance between the eyes. Second, we assume the probability distributions for left and right eyes are identical. We denote this by dropping the subscript: $p = p_l = p_r$. Third, we may additionally assume the probability distributions are symmetric around the origin. Point symmetry of the distribution means the probabilities of the variable errors $+\varepsilon$ and $-\varepsilon$ are the same: $p(+\varepsilon) = p(-\varepsilon)$. This would allow us to consider pairs of points based on the error vectors $\varepsilon_l, \varepsilon_r$ and $-\varepsilon_l, -\varepsilon_r$, which are symmetric around the line $(0, 0, 1)^T$ (i.e. one error vector is the reflection of the other by the line). Since their probabilities are the same, their mean is on this line. Pairing all instances of variable errors in this way shows that the mean would be on the symmetry line, for all probability distributions with point symmetry.

Consider the pair of error vectors $\varepsilon_l, \varepsilon_r$ and the reversed pair $\varepsilon_l, \varepsilon_r$. By our assumptions, these instances have the same probability. The two pairs give rise to two points of interest. Figure 5 illustrates the case of horizontal error only. In this case, one of the two intersection points is closer to the observer than the target, while the other one further away. Their mean will never be closer than the real target, indicating that the mean over all instances with horizontal noise only will be biased to be further away than the target – as observed in the numerical simulations. We now prove that this intuitive reasoning is correct.

To compute the expected depth value, we start by analytically solving the least squares estimation (LSE) in Eq. 10 for the last component. This can be done by elementary computations. In this way we can express the depth for a certain pair of error vectors $\varepsilon_l, \varepsilon_r$ as well as

for the reversed pair $\varepsilon_l, \varepsilon_r$. Then we take the mean of the two depth values. All computations can be carried out by a computer algebra system such as Maple or Mathematica. The resulting expression is lengthy and unsuitable for direct inspection. However, considering the case of horizontal noise only leads to the following simple expression:

$$\frac{z'}{z} = \frac{4a^2}{4a^2 - (\eta_l - \eta_r)^2} \quad (17)$$

As defined in Equation 1 and 12, $2a$ equals to the interocular distance and $\eta_{l/r}$ represents the horizontal component of variable error $\varepsilon_{l/r}$. It makes sense to assume that the distance between the displacements $\eta_{l/r}$ is smaller than the interocular distance $2a$ (otherwise the intersection of the eye rays would be behind the observer). In this case the denominator is positive but never larger than the numerator, so the depth is larger. Because this is true for any pair of instances (see more in Appendix A), this is also true for the mean resulting from an arbitrary probability distribution (as long as they are the same for both eyes). This confirms our intuitive geometric conclusion from Figure 5.

Next we consider only vertical noise $v_{l/r}$. In this case it is convenient to consider the pair of error vectors $\varepsilon_l, \varepsilon_r$ and $-\varepsilon_l, -\varepsilon_r$. Note that compared to the former case we now also reverse the signs, which is admissible if the probability distribution is symmetric. Using computer algebra as before, we find that the relative mean depth for the two intersection points is

$$\frac{z'}{z} = \frac{4a^2 + a^2(v_l + v_r)^2}{(v_l - v_r)^2 + 4a^2 + a^2(v_l + v_r)^2}. \quad (18)$$

Comparing numerator and denominator reveals that $z' \leq z$ with equality only for $v_l = v_r$ regardless of the probability distribution and without any restrictions on the vertical noise.

Together, these two results confirm our observations that noise in horizontal direction biases the mean depth towards larger values, while noise in vertical direction biases it towards the observer. This result holds for all probability distributions, as long as they are identical for both eyes (and exhibit symmetry in the vertical direction).

In any realistic scenario, however, noise errors have horizontal and vertical components. The resulting bias will depend on the relative magnitude of these errors. If the magnitude of horizontal and vertical noise error is equal, then the bias is toward smaller depth and the magnitude of this effect is dominated by the squared difference $(v_l - v_r)^2$. In general, however, the variance along the horizontal versus vertical axes in the probability distribution determines the bias in depth. If the variance is higher in horizontal direction, the depth will be biased towards greater values; if the variance is higher in vertical direction, depth will be biased closer to the viewer. As before, this result holds for probability distributions as long as they are identical for both eyes and the point closest to any two eye rays in the distributions is in front of the viewer (not behind).

Part 2: Human data

The numerical simulations above show that variable and systematic errors in eye rays have a significant influence on the estimated mean point of vergence under projection. In this section, we collect human vergence data to study the real noise distributions when using a video-based eye-tracker. Eye-trackers primarily output mapped gaze positions, which in 2D are points on a screen, described in pixel coordinates. This mapping is established using a calibration routine, where participants are asked to look at several targets on the screen while features of their eyes are tracked in the camera image. These pre-calibrated features - the pupil centre and the centre of the reflection in the cornea - are mapped to gaze positions on the monitor during calibration. With an established mapping, any eye positions can be mapped into the target space, which corresponds to the gaze point.

Several sources of errors are known to interfere in this procedure, in particular when both eyes are being calibrated. Despite existing research on vergence eye movements (Collewijn et al., 1995, 1997; Erkelens et al., 1989), there is no established consensus on how to calibrate binocularly. Nuthmann and Kliegl (2009) have brought up the question whether binocular calibration (i.e. calibrating both eyes at the same time) or monocular calibration (i.e. calibrate one eye at one time while the view of the other eye is occluded) is better suited for binocular eye movements study. Besides the intrinsic properties of eye movements in binocular viewing, the

estimated mapping function in calibration also introduces errors in the estimated point of interest. The estimation of parameters in the mapping function is a minimization procedure. In practice, low order polynomials are often used to map pupil and corneal reflection positions onto screen coordinates, and the mapping parameters are approximated through an optimization procedure (e.g. least squares), which inevitably contains modeling errors.

Therefore, we designed a data collection including both binocular and monocular viewing conditions and analyzed the detected raw pupil and corneal reflection positions without mapping them to the target space. Two depth variations were included and we used symmetric presentation of stimuli to counterbalance any potential behavioral differences due to spatial dependency.

Participants

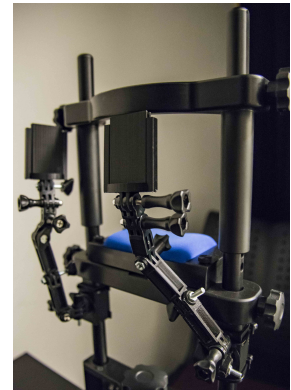
25 participants from TU Berlin (students and staff) joined our experiment (mean age = 27, SD = 7, 4 females). They all had normal or corrected to normal vision and provided informed consent. Participants with glasses were excluded from the experiment due to concerns over eye tracking accuracy. Five of them had previous experience with eye tracking experiments. Their time was compensated. Additionally we performed a monocular eye examination for each participant and the averaged acuity is 0.93 (SD=0.26) measured in the decimal system. We also measured the eye dominance of observers following the standard sighting eye dominance test, as our experiment does not involve any interocular conflict. Observers were asked to look at a distant point through a small hole placed at arm length, and then to close their eye one after another. Only the dominant eye supposes to see the point while viewing monocularly. Among all 25 observers, only 7 of them has a left dominant eye while 18 of them reported to have a right dominant eye. We discuss on the eye dominance test in the discussion section.

Apparatus and recording setup

The data collection was conducted in a quiet and dark room. We used an EyeLink 1000 desktop mount system in the experiment and binocular eye movements were tracked at a sampling rate of 1000 Hz. A chin-forehead rest was used to stabilize observers' head positions. A 24-inch display (0.52 m × 0.32 m, 1920 × 1200 pixels) was placed at two distances of 0.7 m and 1.1 m from the eyes. Stimulus presentation was controlled using

PyLink provided by SR Research. Note that any eye trackers can be used for the recordings as our analysis is based on detected eye positions in the camera frames.

Two custom-built eye covers fabricated by 3D printing were mounted on the chin-forehead rest (see Figure on the right). Each of them can be rotated by 180° to open or block the view of one eye.



Design and procedure

In order to familiarize observers with fixational eye movements, and to verify the setup that observers' eyes were clearly tracked, we ran a default 9-point calibration routine before the data collection started. A calibration is followed by a validation to verify its accuracy. Experiment was continued only when the achieved validation accuracy is on average smaller than 1.0° of visual angle. Otherwise we either repeated the calibration and validation routine immediately or after a short break. However, the resulting gaze data are not used. All further analyses are purely based on the pre-calibrated pupil minus CR (corneal reflection) data, to minimize influence from the calibration mathematics of the EyeLink.

The data collection consisted of two distances and each distance had five repetitions. In each repetition, we presented three viewing conditions, namely monocular viewing with left eye, monocular viewing with right eye, and binocular viewing. Eye covers were rotated by the instructor to occlude/reveal the eyes in different viewing conditions. Meanwhile, the EyeLink was continuously tracking in binocular mode throughout.

For each distance, repetition and condition, the participant was asked to fixate each one out of 12 targets in the form of a ring with an addition central point (Figure 6). Participants were instructed to follow the marker and fixate the white dot at the center as accurately as possible. Black circles with their center marked by a white dot were used as the target. The radius of the ring corresponded to 7° of visual angle. Each marker was presented for 1.5 s and a beep sound was used to signal the start. Between each fixation trial, participants refixated the center point. In total, there were 12 targets fixations on

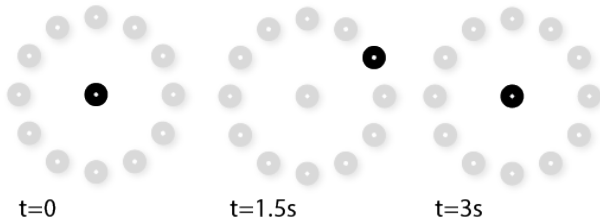


Figure 6 An exemplary trial. The center was always presented at the beginning of each trial. A randomly selected marker on the ring was then presented for 1.5s followed by the presentation of the center marker for another 1.5s. Then another random selected marker on the ring was presented. The presentation continues in such way until all markers on the ring were presented once. Gray markers are drawn here for illustration purpose and were invisible during experiment.

the periphery targets and 12 at the center target.

Each block of five repetitions for a distance took about 20 minutes and participants were asked to take a break before the second block. In total it lasted about one hour including a simple trial run at the beginning.

Analysis Methods and Results

Our analysis was based on raw samples of detected eye positions, i.e., the pupil minus CR positions, represented as pixel coordinates in the eye camera. The only use of calibrated gaze was our use of the built-in velocity-based algorithm of the EyeLink software to detect fixations. In order to find fixations on targets, we filtered out all short fixations that are less than 100ms in a preprocessing step as short fixations very often are the result of undershoots (Holmqvist and Andersson 2017, p.222-223) and are quickly followed by a correction saccade onto targets. Conservatively, among fixations on the same target, we considered those that are two standard deviations away from the cluster center as outliers. The remaining fixations were from then on only processed as uncalibrated pupil minus CR.

For each dataset of one observer, we aggregated all fixations (pupil-CR) in the five repetitions for each target in the same viewing condition. On average there are 56 fixations in each such block (SD = 20), and each target has 2.3 fixations as we have 24 fixation targets in each repetition. Here fixations from individual eye are considered independently. The averaged duration of fixations is 821ms (SD = 494ms). There is no significant difference among different viewing conditions at two distances.

Comparing sets of covariance matrices

In the previous simulation (see Section “Ray Errors”), fixation positions have independent zero-mean Gaussian distributions for variable errors in horizontal and vertical directions, which can be represented by a 2×2 diagonal covariance matrix Σ . Diagonal elements in Σ describe the variances in horizontal and vertical directions separately and off-diagonal elements show the correlation between them. Therefore, Σ is a positive semi-definite symmetric matrix, and it can be visualized as an error ellipse with its axes pointing into the directions of its eigenvectors. The lengths of the semi-axes are proportional to the square roots of the corresponding eigenvalues λ_i . We used $\sqrt{5.991} \sqrt{\lambda_i}$ as the semi-axis length (derived from a Chi-Square distribution), which gave us the ellipse that covers a 95% confidence interval. We continued using covariance matrices to analyze the error distributions and visualized them as ellipses in the following. Essentially covariance matrix measures the precision in two dimensions, as we are interested in its special distribution.

In our data collection, we had a set of markers on screen and a covariance matrix represented fixation distribution at each marker position. To model the variability among different covariance matrices and to compare the differences among sets of covariance matrices, we computed distances between all pairs of covariance matrices in a set and then compared the resulting distributions. Despite the raising amount of applications of analyzing the variance among covariance matrices, there is still no consensus on how to analyze the covariance structures (Rencher, 2003). We settled on a logarithm-based distance estimator (Förstner and Moonen, 2003), a Riemannian metric, defined as

$$d(\Sigma_1, \Sigma_2) = \left\| \log(\Sigma_1^{-\frac{1}{2}} \Sigma_2 \Sigma_1^{-\frac{1}{2}}) \right\|, \quad (19)$$

where the logarithm is given by $\log \Sigma = \mathbf{U} \log(\mathbf{S}) \mathbf{V}$ and $\mathbf{U}, \mathbf{S}, \mathbf{V}$ can be factorized from singular-value decomposition (SVD) as $\Sigma = \mathbf{U} \mathbf{S} \mathbf{V}$. \mathbf{S} is a diagonal matrix of singular values of Σ . $\|\mathbf{X}\|$ is the Euclidean norm (also called Frobenius norm) and it can be computed by the matrix trace $\|\mathbf{X}\| = \sqrt{\text{Tr}(\mathbf{X}^T \mathbf{X})}$. In case of covariance matrix, the

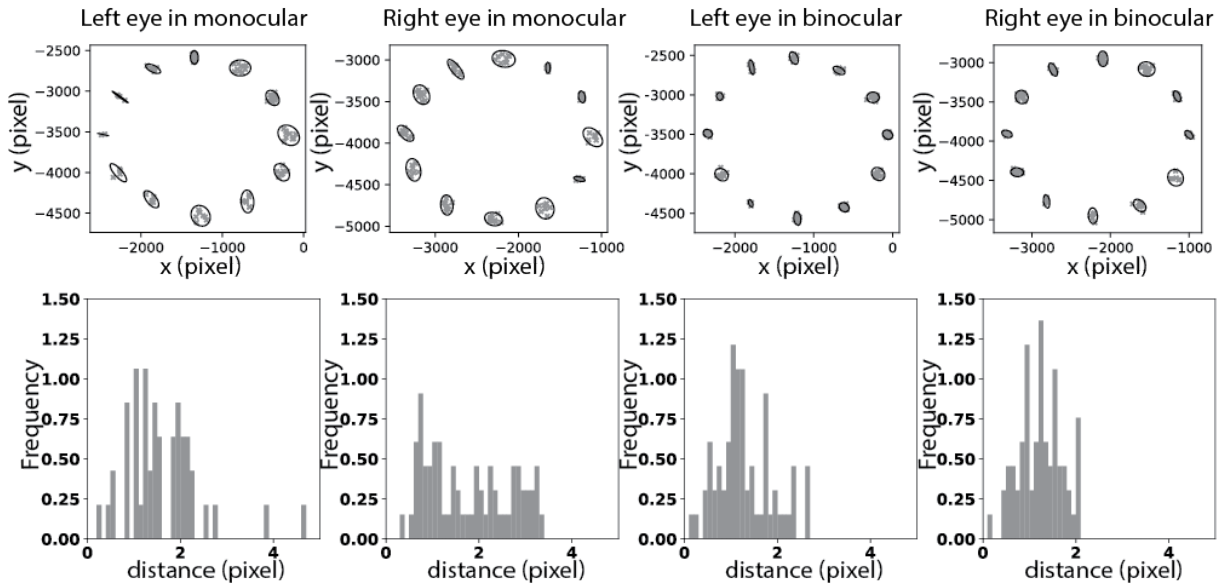


Figure 7 Example of one-dataset samples at the distance of 70 cm. In the first row, fixations at each marker position collected over five repeats are scattered and corresponding covariance matrices are visualized as ellipses. Second row shows the histograms of distances between all pairs of ellipses in each condition (summed over five repeats). Observer of this dataset has right dominant eye and eye acuity is 0.9 for left eye and 1.0 for right eye.

trace measures the total variation in each dimension without considering the correlations among variables. As shown in Förstner and Moonen (2003), this distance measure is symmetric and non-negative, and it is invariant under affine transformation and inversion.

Intuitively speaking, by multiplying with $\Sigma_1^{-\frac{1}{2}}$ in bilinear form, we transform Σ_2 into a new basis $\Sigma_1^{-\frac{1}{2}} \Sigma_2 \Sigma_1^{-\frac{1}{2}}$ is a perfect circle. The distance measures the relative ratio of eigenvalues in the new basis and the largest eigenvalue of $\Sigma_1^{-\frac{1}{2}} \Sigma_2 \Sigma_1^{-\frac{1}{2}}$ corresponds to the ratio of maximum variance between two groups (Rencher, 2003). To aggregate a set of covariance matrices, we used the arithmetic average as the mean covariance matrix. Dryden et al. (2009) compared many covariance distance measures especially in the context of shape interpolation; however, it is not clear which one suits best in our case. Here we only want to be able to compare sets of covariance matrices.

In the next step, the distance distributions of two sets of covariance matrices were compared using the Kolmogorov-Smirnov test (KS-test) (Sachs, 2012) and a p-value was computed to determine whether the two distributions differ significantly. The KS-test computes the vertical distance between two cumulative fraction functions that

are used to represent two distributions and takes the largest distance as the statistic. Therefore, it is robust with respect to variance types of distributions.

Results

Below, we examined whether the bias was present also in the human data. We first compared the distributions of variable errors among all individual observers. Then we aggregated all datasets to examine whether the distributions has any spatial dependency. In the last step, we used the averaged variable error distribution to sample eye positions following the procedure in the previous simulation (see Section Ray Error) to investigate whether there is a bias in the direction of the observer in human data as there was in simulated data.

Eye dominance and acuity do not seem to matter

Each dataset (of one observer) had five repetitions of each viewing condition. Each eye had two viewing conditions, namely monocular viewing and binocular viewing. Targets were presented at two different distances. We first aligned the five repetitions in the same condition and computed the covariance matrix of fixations at each marker position. Following Equation 18, each covariance

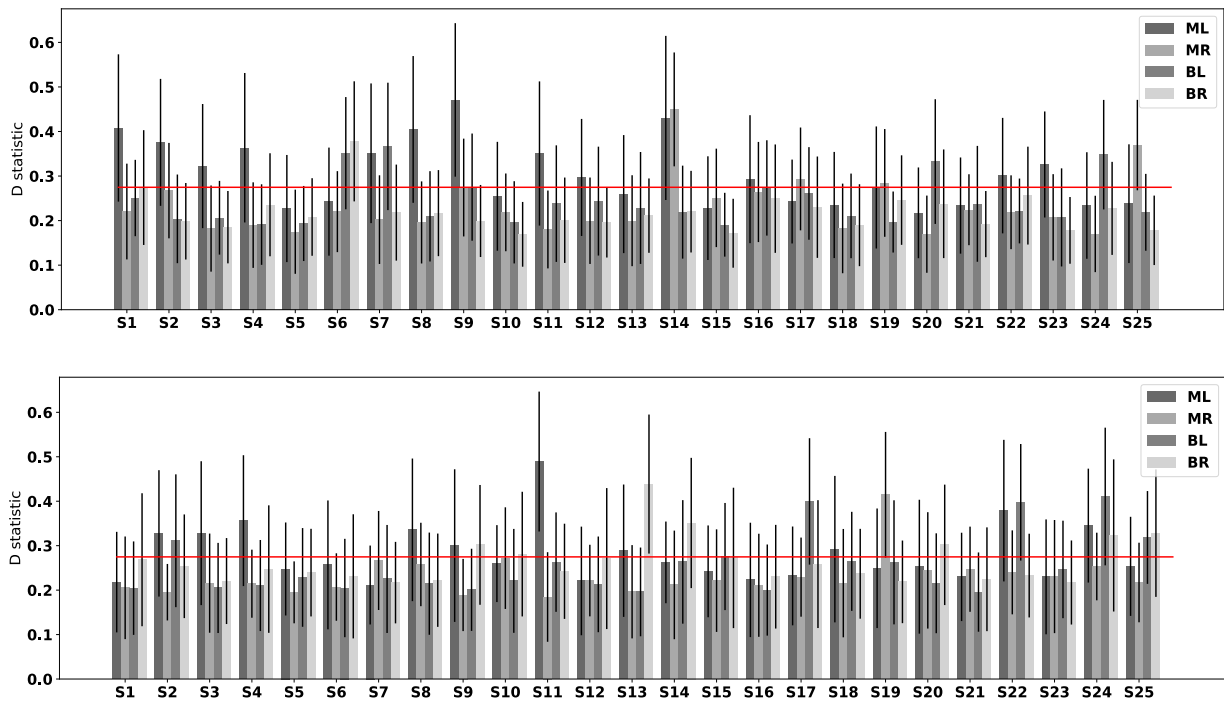


Figure 8 Statistics of KS-test on individual differences. Upper plot shows the statistics of data with a display placed at a distance of 70 cm and the lower one shows the result when the display was at a distance of 110 cm. Red lines mark the 0.05 significance level. ML stands for the condition of monocular viewing of left eye, MR monocular viewing of right eye, BL binocular viewing of left eye and BR binocular viewing of right eye.

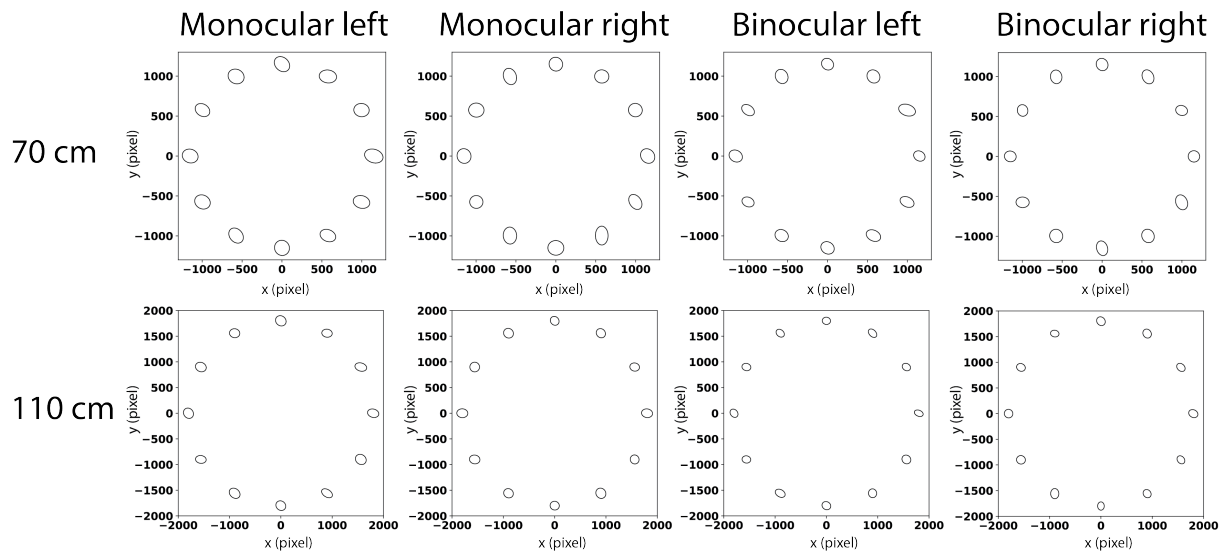


Figure 9 Visualization of mean covariance matrices at each marker position. First row shows error distributions of data collected at a distance of 70 cm and second row shows data collected at a distance of 110 cm. In each viewing condition at one distance, data from all observers were aggregated together and covariances at each marker position are visualized as ellipses. For instance, a left-tilted ellipse means that vertical noise is larger than horizontal noise. The size of the ellipse thus corresponds to the total variation while orientation indicates the correlation between horizontal and vertical directions.

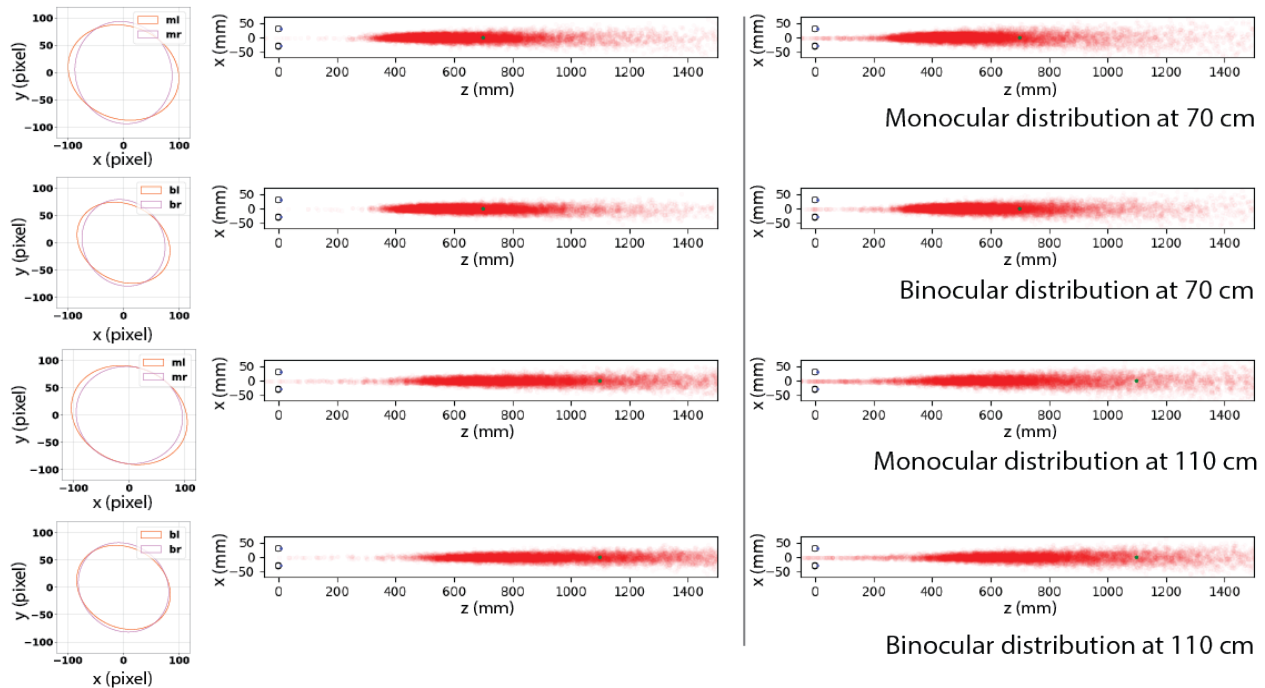


Figure 10 Estimation of mean vergence point based on distributions formed from real human data. Left column shows the mean covariance matrices of left eye (l) and right eye (r) in monocular viewing (m) and binocular viewing (b). First two rows show results when target is placed at a distance of 70 cm and last two rows show results at a distance of 110 cm. Right column shows the distribution of estimated vergence points in space where variance (visualized in the left column) was converted to 1° and 2° of visual angle respectively. See Table 2 for more detailed statistics.

matrix was compared to the other eleven covariance matrices. Note that we discarded all fixations at the center marker. One exemplar dataset is shown in Figure 7. Covariance matrices are visualized as ellipses and distributions of all pair-wise distances are plotted as histograms.

Using the KS-test, we compared histograms of the distance distributions of each individual dataset to the others and results are shown in Figure 8. Red line marks the significance level of 0.05. Note that any statistic value larger than the level (i.e. above the line) corresponds to a significantly different distribution. Considering each eye separately, we had four viewing conditions in one dataset, namely monocular viewing of left eye (ML), monocular viewing of right eye (MR), binocular viewing of left eye (BL) and binocular viewing of right eye (BR). At the distance of 70 cm, 13 out of 25 datasets have significantly different fixation distributions in ML, 4 in MR, 5 in BL and 1 in BR. According to the categorization of eye dominance, we have 8 significantly different distributions of dominant eye in monocular viewing and 12 significantly different distributions of nondominant eye. In

binocular viewing condition, 2 distributions of dominant eye significantly differ from the others and 6 of nondominant eye. At the distance of 110 cm, the numbers of datasets, which are significantly different from the others, are 10 (ML), 1 (MR), 6 (BL), and 7 (BR). In monocular viewing, 3 and 9 datasets are significantly different for distributions of dominant eye and nondominant eye respectively. In binocular viewing condition, the number of different datasets is 9 for dominant eye and 6 for nondominant eye. Note that this counting is based on the mean statistic value. And we do observe large variances in each dataset as shown in Figure 8.

In conclusion, the differences between dominant and non-dominant eye are so small and varied that we cannot conclude that eye dominance matters. Neither did we observe any correlation between the distribution differences and eye acuity, which indicates that eye acuity does not contribute to the differences in distributions. It is likely that other factors, such as eye colour, may explain part of the noise.

Noise does not vary depending on where participants look

It is known that noise varies across the measurement space (Holmqvist and Andersson, 2017, p.182). To test whether there is any such spatial dependency of noise distributions of fixations, we computed the distance distribution at each marker position by comparing each pair of covariance matrices from 25 datasets. Similarly, we applied the KS-test to compare the distance distributions. Only one distribution from monocular viewing of right eye when screen was at a distance of 70 cm is significantly different. To visualize the variance at each marker position, we computed the mean of 25 covariance matrices for each marker in one viewing condition and plotted corresponding ellipses in Figure 9. Note that comparing two distances, covariance ellipses are similar in both sizes and orientations. But ellipses in binocular viewing condition have smaller sizes but still similar orientations. In conclusion, the amount of noise does not seem to differ between the conditions, nor between positions. There is however a tendency that vertical noise is larger than horizontal noise.

But there is bias in the human data

After we have established that noise does not vary across eye dominance, acuity and fixation position, we can now aggregate the covariance matrices at all marker positions, which gives us an overall representation of noise distribution of fixations in each condition. Using the distributions - a covariance matrix for each of the 12 fixation point in each condition - rather than the data themselves introduces no bias, but is computationally easier, since it allows us to calculate the mean easily without being biased by unevenly taken data (i.e. uneven contribution of individual observers because of fixations of differing lengths).

We used the arithmetic mean of the 12 covariance matrices (of 12 targets) to represent the fixation distribution in each viewing condition and obtained two covariance matrices, one for each eye in either monocular or binocular viewing condition at one distance. As shown in Figure 10, an ellipse is used to represent a covariance matrix. Errors seem to be larger when targets move from 70 cm to 110 cm with increased sizes of ellipses. Noise in binocular viewing condition is smaller comparing to that in monocular viewing condition, evidenced by smaller

ellipses in most positions in all conditions. The radius of the ring corresponds to 7° of visual angle and opposite sample positions span a visual angle of 14° . Following this ratio, we converted sample units (measured in pixel) into degrees of visual angle and applied the analysis framework used in the simulation above. Variances of error distributions shown in Figure 10 approximate to 1° of visual angle, which is commonly used as a calibration threshold in 2D eye tracking experiment. Additionally we also experiment with 2° of visual angle, which is equivalent to the state-of-art eye tracking accuracy in 3D space (Pfeiffer and Renner, 2014; Gutierrez Mlot et al., 2016; Wang et al., 2017b). Simulated results based on sampling from real error distributions are plotted in Figure 10. On the left side we have spatial distributions of vergence points when variance was converted to 1° of visual angle and on the right side we see the results after converting to 2° of visual angle. Detailed statistic results are given in Table 2. Although the error in estimated mean vergence point is acceptable when everything is perfect within 1° of visual angle, however, achieving such accuracy in 3D is rather challenging. Even with an acceptable accuracy of 2° , there exists a strong bias towards the viewer in the estimated mean point of vergence.

Minimizing the uncertainty in vergence point estimation

When researching vergence using a video-based pupil and corneal reflection eye-tracker such as the Eyelink, what can you do to minimize biases and errors? We assume that the human participant has a negligible difference in gaze directions between left and right eye, that luminance conditions are fixed and no other effects on pupil dilation are present, and that the only remaining issue is to minimize the bias from the noise in the signal.

As this bias is an effect of the projective mapping, the non-linear mapping that is commonly used to estimate the vergence point in space, there is not much you can do if you use the calibration routine shipped with the eye-tracker. The single fixation per calibration point will have an inaccuracy and noise that increases the bias. However, it is possible to instead record multiple fixations on each *point in your own set of calibration targets*. Then take the average of the data samples in the several fixations per fixation target and use that average to calibrate the eye-

Table 2. Vergence errors sampled from real noise distributions. σ represents converted standard deviation in degree of visual angle. \bar{p}_z represents the mean point of vergence in depth measured in cm. Averaged errors and standard deviations are reported in cm with x represents the horizontal direction, y the vertical direction, and z in depth. On average, the errors in percentage of the corresponding distance are 30% ($\sigma = 1$) and 46% ($\sigma = 2$) when target distance is 70 cm, and 41% ($\sigma = 1$) and 50% ($\sigma = 2$) when target distance is 110 cm.

target distance (cm)	σ (deg)	condition	\bar{p}_z (cm)	\overline{error} (cm)	\overline{std} (cm)	\overline{error}_x (cm)	\overline{error}_y (cm)	\overline{error}_z (cm)	\overline{std}_x (cm)	\overline{std}_y (cm)	\overline{std}_z (cm)
70	1.0	monocular	69.1	22.4	32.0	0.88	0.86	22.3	0.88	0.90	32.0
		binocular	70.4	18.9	27.8	0.76	0.74	18.8	0.79	0.77	27.8
	2.0	monocular	60.0	35.1	82.6	1.2	1.2	34.9	2.8	2.1	82.6
		binocular	64.3	28.6	53.7	1.1	1.0	28.5	1.3	1.3	53.8
110	1.0	monocular	98.5	49.9	102.3	1.2	1.1	49.8	1.5	1.7	102.3
		binocular	104.4	41.6	88.8	9.9	9.5	41.5	1.4	1.2	88.8
	2.0	monocular	73.3	70.1	141.2	1.5	1.3	70.0	2.5	2.8	141.2
		binocular	86.7	60.2	171.7	1.3	1.3	60.1	2.3	2.7	171.7

tracker. The key is to use many fixations per calibration point. The same principle applies to real experiment after calibration: collect many fixations if possible. For each point of interest, taking the average of fixation samples in the camera space leads to better estimation in space. In practice, this could mean a repetition of the same experimental condition, for example, where observers are asked to refixate on the same targets in a vergence study.

The reason this method works is that for geometrical reasons, it is better to average in the calibration data than in the depth data of the intersection points. Assume that noise in the eye samples are Gaussian distributed, taking the average in the calibration data leads to better approximation to the noise-free eye samples. However, due to the non-linearity of the mapping, averaging in the depth data only leads to a bias as we see before. Figure 11 shows for simulated data how the offset in depth decreases with an increasing number of fixations per calibration point.

This solution works on the DPI, the EyeLink and the SMI eye-trackers, which all provide access to the center point of the corneal reflection and the pupil (for EyeLink

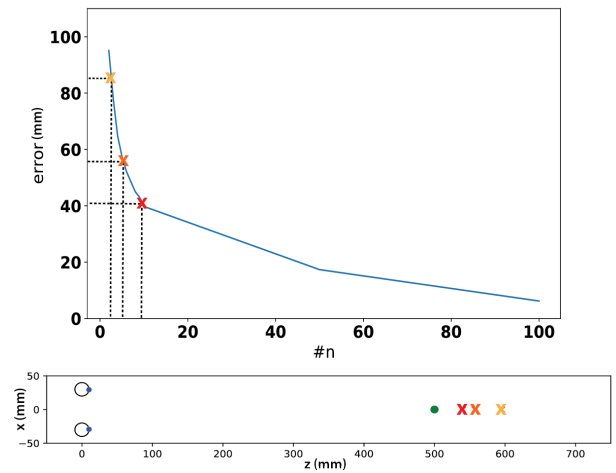


Figure 11 More fixations per target leads to better estimation of mean vergence point. At the top of the figure, offset error is plotted as a function of number of fixations per calibration point. Three color-coded crosses mark the offset errors when number of fixations equal to 2, 5, and 10. These three offsets are visualized in a top view plot at the bottom. We used standard deviation of 1.5 in horizontal direction and 0.16 in vertical direction.

and SMI) or the 4th Purkinje (for the DPI) in the data. Note that the recipe works with raw positions of the eyes and does not depend on specific calibration model given by any eye trackers.

Discussion

The major finding in this paper is that vergence data from eye-trackers exhibit a bias, depending on the noise level in horizontal vs. vertical directions. This applies to the estimation of the point of vergence in three-dimensional space, as well as to the vergence estimation on planar surface in two-dimensional space. As long as a projective model is used in the mapping, the estimated mean vergence point will be biased.

Reading research and calibration algorithms

The bias we have found is in line with Nuthmann and Kliegl (2009), who reported that the fixations during reading were almost always crossed, peaking 2.6 cm in front of the plane of text, which is very much in line with the error reported in Table 2. However, instead of resulting in a discussion about potential biases from the measurement technology itself, the subsequent papers instead investigated whether monocular vs. binocular calibration could cause the crossing of fixations.

Liversedge et al. (2006a) had found that “When the points of fixation were disparate, the lines of gaze were generally diverged (uncrossed) relative to the text (93% of fixations),..., but occasionally converged (crossed) (7% of fixations).” using a monocular calibration. Later study by Kirkby et al. (2013) replicated the finding that - after monocular calibration - the majority of fixations are uncrossed, both with the EyeLink and the DPI.

Our results using the collected human data show a significant bias towards observers for both monocular and binocular calibration, corresponding to crossed fixations in reading. This bias must have existed also for the studies by Liversedge et al. (2006a) and Kirkby et al. (2013), so it is surprising that they found crossing result in the opposite direction of the bias.

Švede et al. (2015) argued that monocular calibration is the only physiologically correct form in the sense that it preserves the difference in gaze direction between the two eyes. It could be the case that the average gaze differences between left and right eye after monocular cali-

bration are so large that they consume the whole bias and nevertheless can remain uncrossed.

Despite the fact that the bias in the estimated gaze positions in binocular viewing is smaller than the bias in monocular viewing condition, the ratio in depth between target position and the estimated mean vergence point is around 1.2.

Eye dominance, interpupillary distance and fixation disparity

The collected human data show a large variance among individual observers. However, neither eye dominance nor acuity leads to significantly different noise distribution and where participants look also doesn't seem to matter. Nevertheless we should be aware that eye dominance information is based on observer's self-report. Moreover, the eye dominance is determined by the so-called sighting eye dominance test. A recent study (Ding et al. 2018) brought up the question whether one type of eye dominance exists and their results indicate disagreements among different eye dominance test methods, especially the difference between sighting eye dominance test and binocular rivalry based test. Even though our experiment does not involve any interocular conflict, the actual dominant eye during the experiment might still be different from the measured one, which might further explain the observed no-impact findings.

Interpupillary distances may also influence the bias, as it changes the geometry, i.e. the short edge of the triangle formed by the two eyes and a target. How the bias is correlated to the interpupillary distance is not studied in this work.

It is not clear so far how fixation disparity may contribute to our observations. Liversedge et al. (2006b) reported that fixation disparity decreased over time during reading, and it is tightly linked to vergence eye movements (Saladin JJ. 1986, Jaschinski et al. 2008, Švede et al. 2011) as well as binocular vision (Jaschinski et al. 2008). Additionally, vision training may improve stereo perception and eliminate fixation disparity (Dalziel CC. 1981). We neither performed any fixation disparity test nor measured the binocular version in our study, although how to accurately measure fixation disparity seems to be an on-going effort (de Meij et al. 2017). Future studies should include measurements such as the near point of convergence, positive and negative fusional range, dissociated phoria at near and far, stereopsis and

amplitude of accommodation. It would also be interesting to test whether there exists a correlation between fixation disparity and the bias in individual's dataset.

Limitations of our study

Independent Gaussian distributions were used for each dimension. It would be interesting to study the dependency between noise in horizontal and vertical directions and to see how a multivariate distribution influences the bias.

In our experiment targets were displayed on a flat plane without the need of focus change. But noise distribution might be different due to the dynamic changes of focus in 3D. For example we might need to take into account the changes of pupil size. How to collect enough fixation data in 3D while maintaining the same precision is another practical but challenging problem.

To minimize the bias from the signal noise, we suggest to collect many fixations per calibration target, and then use the mean fixation point to calculate the mapping. Future work should validate this proposal with real human data. Note that this proposed recipe does not account for fixation disparity, i.e. the alignment difference between the dominant eye and the weak eye.

Suggestion for future experiments

There seems to be no good reason to prefer one viewing condition to the other in calibration. In monocular calibration, each individual eye is forced to fixate on targets without the interference of binocular fusion and eye dominance. However, visual acuity is also limited in monocular viewing and possibly decreases over distances. Precision of eye movements in binocular viewing condition seems to be higher.

We believe it is important to be aware that the propagation of noise may lead to a bias in the estimated mean vergence point in space. It is also very important for future eye movement studies, especially for vergence studies, to provide the validation error of calibration not just as a single scalar, but also in the form of a spatial distribution (i.e. an error ellipse). It provides a confidence level of the observed data and puts their interpretation into perspective. It is not even clear if the error distribution would be roughly similar for all types of tracking devices and experiments, or if the distribution changed with the type of experimental task. If so, resulting mean vergence depth would vary with experimental setup.

Moreover, our results suggest that, for researchers using eye-tracking devices, it is good to think about the procedure, instead of being only concerned about the data after calibration. The mathematical models behind calibration might provide additional information as being part of the experiment.

Conclusions

The estimated point of interest from intersection of eye rays has large error in depth. The mean of the linearly estimated points of vergence is biased and depends on the horizontal vs. vertical noise distribution of the fixation positions. It is generally hard to interpret results for depth from binocular vision and our suggestion is to take the average of fixations of the same target to minimize the uncertainty, in both calibration and experiment phase.

Ethics and Conflict of Interest

The authors declare that the contents of the article are in agreement with the ethics described in <http://biblio.unibe.ch/portale/elibrary/BOP/jemr/ethics.html> and that there is no conflict of interest regarding the publication of this paper.

Acknowledgements

We would like to thank Marianne Maertens and Albert Chern for valuable advice. This work has been partially supported by the ERC through grant ERC-2010-StG 259550 (XSHAPE). We acknowledge support by the German Research Foundation and the Open Access Publication Fund of TU Berlin.

References

- Atchison DA, Smith G (2000). Optics of the human eye. *Butterworth-Heinemann Oxford*.
- Cassin B, Rubin ML, Solomon S (1984). Dictionary of eye terminology. *Triad Publishing Company Gainesville*.

- Cerrolaza JJ, Villanueva A, Cabeza R (2012). Study of polynomial mapping functions in video-oculography eye trackers. *ACM Trans. Comput.-Hum. Interact.* 19(2):10, doi:10.1145/2240156.2240158.
- Collewijn H, Erkelens CJ, Steinman RM (1995). Voluntary binocular gaze-shifts in the plane of regard: Dynamics of version and vergence. *Vision Research* 35(23), 3335-3358, doi:10.1016/0042-6989(95)00082-P.
- Collewijn H, Erkelens CJ, Steinman RM (1997). Trajectories of the human binocular fixation point during conjugate and non-conjugate gaze-shifts. *Vision Research* 37(8), 1049-1069, doi:10.1016/S0042-6989(96)00245-3.
- Dalziel, C. C. (1981). Effect of vision training on patients who fail Sheard's criterion. *American journal of optometry and physiological optics*, 58(1), 21-23.
- Drewes J, Zhu W, Hu Y, Hu X (2014). Smaller is better: drift in gaze measurements due to pupil dynamics. *PLoS ONE* 9(10), 1-6, doi:10.1371/journal.pone.0111197.
- de Meij L., Telleman MAJ, Luijten MRC, Polling JR, Gutter M. (2017). An Optimal Measurement of Fixation Disparity Using Ogle's Apparatus. *Strabismus*, doi:10.1080/09273972.2017.1349813.
- Ding Y, Naber M, Gayet S, Van der Stigchel S, Paffen CL (2018). Assessing the generalizability of eye dominance across binocular rivalry, onset rivalry, and continuous flash suppression. *Journal of Vision*, 18(6): 6, 1-13, doi:10.1167/18.6.6.
- Dryden IL, Koloydenko A, Zhou D (2009). Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *The annals of Applied Statistics* 3(3), 1102-1123, doi:10.1214/09-AOAS249.
- Erkelens CJ, Steinman RM, Collewijn H (1989). Ocular vergence under natural conditions. II. Gaze shifts between real targets differing in distance and direction. *Proceedings of the Royal Society of London Series B, Biological Sciences* 236(1285), 441-465, doi:10.1098/rspb.1989.0031.
- Essig K, Pomplun M, Ritter H (2006). A neural network for 3D gaze recording with binocular eye trackers. *International Journal of Parallel, Emergent and Distributed Systems* 21(2), 79-95, doi:10.1080/17445760500354440.
- Förstner W, Moonen B (2003). A metric for covariance matrices. *Geodesy-The Challenge of the 3rd Millennium*, Springer, pp. 299-309.
- Gutierrez Mlot E, Bahmani H, Wahl S, Kasneci E (2016). 3D gaze estimation using eye vergence. In: *Proceedings of the International Joint Conference on Biomedical Engineering Systems and Technologies, SCITEPRESS - Science and Technology Publications, Lda, Portugal, BIOSTEC 2016*, pp. 125-131, doi:10.5220/0005821201250131.
- Hammer JH, Maurus M, Beyerer J (2013). Real-time 3D gaze analysis in mobile applications. In: *Proceedings of the 2013 Conference on Eye Tracking South Africa, ACM, New York, NY, USA, ETSA '13*, pp. 75-78, doi:10.1145/2509315.2509333.
- Hennessey C, Lawrence P (2009). Noncontact binocular eye-gaze tracking for point-of-gaze estimation in three dimensions. *IEEE Transactions on Biomedical Engineering* 56(3), 790-799, doi:10.1109/TBME.2008.2005943.
- Holmqvist K, Andersson R (2017). Eye Tracking. A comprehensive guide to methods, paradigms, and measures. *Lund Eye-tracking Research Institute*.
- Holmqvist K, Nyström M, Andersson R, Dewhurst R, Jarodzka H, Van de Weijer J (2011). Eye tracking: A comprehensive guide to methods and measures. *OUP Oxford*.
- Holmqvist K, Blignaut P, Niehorster D (2019). Small eye movements cannot be reliably measured by current video-based eye-trackers. *Under preparation for Behavior Research Methods*.
- Hooge IT, Hessels RS, Nyström M (2019). Do pupil-based binocular video eye trackers reliably measure vergence? *Vision Research*, 156, 1-9, doi:10.1016/j.visres.2019.01.004.
- Jaschinski, W., Schroth, V. (2008). Ocular prevalence: difference between crossed and uncrossed disparities of stereo objects. *Strabismus*, 16(4), 159-164, doi:10.1080/09273970802451059.
- Jaschinski, W., Švede, A., Jainta, S. (2008). Relation between fixation disparity and the asymmetry between convergent and divergent disparity step responses. *Vision Research*, 48(2), 253-263, doi:10.1016/j.visres.2007.11.004.

- Kirkby JA, Blythe HI, Drieghe D, Benson V, Liversedge SP (2013). Investigating eye movement acquisition and analysis technologies as a causal factor in differential prevalence of crossed and uncrossed fixation disparity during reading and dot scanning. *Behavior Research Methods* 45(3), 664-678, doi:10.3758/s13428-012-0301-2.
- Liversedge SP, Rayner K, White SJ, Findlay JM, McSorley E (2006a). Binocular coordination of the eyes during reading. *Current Biology* 16(17), 1726-1729.
- Liversedge SP, White SJ, Findlay JM, Rayner K (2006b). Binocular coordination of eye movements during reading. *Vision Research* 46(15), 2363-2374, doi:10.1016/j.visres.2006.01.013.
- Maggia C, Guyader N, Guérin-Dugué A (2013). Using natural versus artificial stimuli to perform calibration for 3D gaze tracking. In: *SPIE 8651, Human Vision and Electronic Imaging XVIII, Bellingham, CA, United States*, doi:10.1117/12.2004683.
- Maurus M, Hammer JH, Beyerer J (2014). Realistic heatmap visualization for interactive analysis of 3D gaze data. In: *Proceedings of the Symposium on Eye Tracking Research and Applications, ACM, New York, NY, USA, ETRA '14*, 295-298, doi:10.1145/2578153.2578204.
- Niehorster DC, Cornelissen THW, Holmqvist K, Hooge ITC, Hessels RS (2018). What to expect from your remote eye-tracker when participants are unrestrained. *Behavior Research Methods* 50(1), 213-227, doi:10.3758/s13428-017-0863-0.
- Niehorster DC, Zemblys R, Hein O, Beelders T, Holmqvist K (in prep). Noise in eye-tracker data: what it is and what shapes it. *Behavior Research Methods*.
- Nuthmann A, Kliegl R (2009). An examination of binocular reading fixations based on sentence corpus data. *Journal of Vision* 9(5), 31-31, doi:10.1167/9.5.31.
- Pfeiffer T, Renner P (2014). Eyesee3d: A low-cost approach for analyzing mobile 3d eye tracking data using computer vision and augmented reality technology. In: *Proceedings of the Symposium on Eye Tracking Research and Applications, ACM*, 369-376.
- Pfeiffer T, Renner P, Pfeiffer-Lessmann N (2016). Eye-See3D 2.0: Model-based real-time analysis of mobile eye-tracking in static and dynamic three-dimensional scenes. In: *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications, ACM, New York, NY, USA, ETRA '16*, 189-196, DOI 10.1145/2857491.2857532.
- Rencher AC, Schimek MG (1997). Methods of multivariate analysis. *Computational Statistics*, 12(4), 442-442.
- Sachs L (2012). Applied statistics: a handbook of techniques. *Springer Science & Business Media*.
- Saladin, J. J. (1986). Convergence insufficiency, fixation disparity, and control systems analysis. *American journal of optometry and physiological optics*, 63(8), 645-653.
- Švede, A., Hoormann, J., Jainta, S., & Jaschinski, W. (2011). Subjective fixation disparity affected by dynamic asymmetry, resting vergence, and nonius bias. *Investigative ophthalmology & visual science*, 52(7), 4356-4361, doi:10.1167/iovs.10-6499.
- Švede A, Treija E, Jaschinski W, Krūmina G (2015). Monocular versus binocular calibrations in evaluating fixation disparity with a video-based eye-tracker. *Perception* 44(8-9), 1110-1128, doi:10.1177/0301006615596886.
- Tian Y, Styanc GP (2001). Rank equalities for idempotent and involutory matrices. *Linear Algebra and its Applications* 335(1), 101-117, doi:10.1016/S0024-3795(01)00297-X.
- Wang D, Mulvey FB, Pelz JB, Holmqvist K (2017a). A study of artificial eyes for the measurement of precision in eye-trackers. *Behavior Research Methods* 49(3), 947-959, doi:10.3758/s13428-016-0755-8.
- Wang X, Lindlbauer D, Lessig C, Alexa M (2017b). Accuracy of monocular gaze tracking on 3d geometry. In: *Burch M, Chuang L, Fisher B, Schmidt A, Weiskopf D (eds) Eye Tracking and Visualization, Springer International Publishing, Cham*, pp. 169-184.

Appendix A: Analytical analysis of bias

Our variable errors contain independent horizontal and vertical components. Each of them follows the zero-mean Gaussian distribution, which is a symmetric function around 0, i.e. $p(a) = p(-a)$.

To simplify the scenario, we set vertical errors to zero and only consider errors in the horizontal direction. Assume left eye and right eye have identical Gaussian distributions, we have $p(\eta_l, \eta_r) = p(\eta_r, \eta_l)$. The probability of observing one pair of errors is the same when it corresponds to left and right eyes or vice versa.

Based on previous two symmetries, we have the same probability for 8 error pairs as shown in Figure 12. As η_l and η_r follow two independent Gaussian distribution of the same variance, we have a joint normal distribution. Each circle centered at origin is a contour line where samples have the same probability. Given an error vector (η_l, η_r) , we can find the other seven pairs by the following operations using symmetry:

- $p(\eta_l, \eta_r) = p(-\eta_l, \eta_r)$ by changing the sign of left eye,
- $p(\eta_l, \eta_r) = p(\eta_l, -\eta_r)$ by changing the sign of right eye,
- $p(\eta_l, \eta_r) = p(-\eta_l, -\eta_r)$ by changing the sign of both eyes,
- $p(\eta_l, \eta_r) = p(\eta_r, \eta_l)$ by swapping between left and right eyes,
- $p(\eta_l, \eta_r) = p(-\eta_r, \eta_l)$ by swapping and changing one sign,
- $p(\eta_l, \eta_r) = p(\eta_r, -\eta_l)$ by swapping and changing one sign,
- $p(\eta_l, \eta_r) = p(-\eta_r, -\eta_l)$ by swapping and changing both signs.

Eight samples are symmetric across four lines which are $\eta_l = 0, \eta_r = 0, \eta_l = \eta_r, \eta_l = -\eta_r$. Together these four lines divide the whole variable domain into eight domains $\{D_i, i = 0, \dots, 7\}$. We define function Φ_i that maps samples from D_0 to D_1 . For example $\Phi_1(\eta_l, \eta_r) := (\eta_r, \eta_l)$ which maps sample in D_1 to D_0 as shown in Figure 12. As Φ_i is a linear map, we know $\det(d\Phi_i) = \det(\Phi_i) = 1$.

Given a pair of horizontal errors (η_l, η_r) , we know the intersection point x can be computed as

$$x_{\eta_l, \eta_r} = (E'_l(\eta_l) + E'_r(\eta_r))^{-1} (E'_l(\eta_l)p_l + E'_r(\eta_r)p_r) \quad (20)$$

and its probability is $p(\eta_l, \eta_r)$. Let us define a function f that $f(\eta_l, \eta_r) := x_{\eta_l, \eta_r}, (\eta_l, \eta_r) \in D_i$ and $f_i = f \circ \Phi_i$. The expected value of x in the whole domain is

$$x = \iint f(\eta_l, \eta_r) p(\eta_l, \eta_r) d\eta_l d\eta_r$$

$$\begin{aligned} &= \sum_{i=0}^7 \iint_{D_i} f(\eta_l, \eta_r) p(\eta_l, \eta_r) d\eta_l d\eta_r \\ &= \sum_{i=0}^7 \iint_{\Phi_i(D_0)} f(\eta_l, \eta_r) p(\eta_l, \eta_r) d\eta_l d\eta_r \\ &= \sum_{i=0}^7 \iint_{D_0} (f \circ \Phi_i)(p \circ \Phi_i) \det(d\Phi_i) d\eta_l d\eta_r \\ &= \sum_{i=0}^7 \iint_{D_0} f_i p d\eta_l d\eta_r \\ &= 8 \iint_{D_0} \sum_{i=0}^7 \frac{1}{8} f_i p d\eta_l d\eta_r \end{aligned} \quad (21)$$

Since probability on the same contour line is the same, $p \circ \Phi_i = p$. The expected intersection point x can be computed by computing the average of f_i . Therefore, we only show the simplified computation of one pair of error vectors in Section Qualitative analytical analysis of bias.

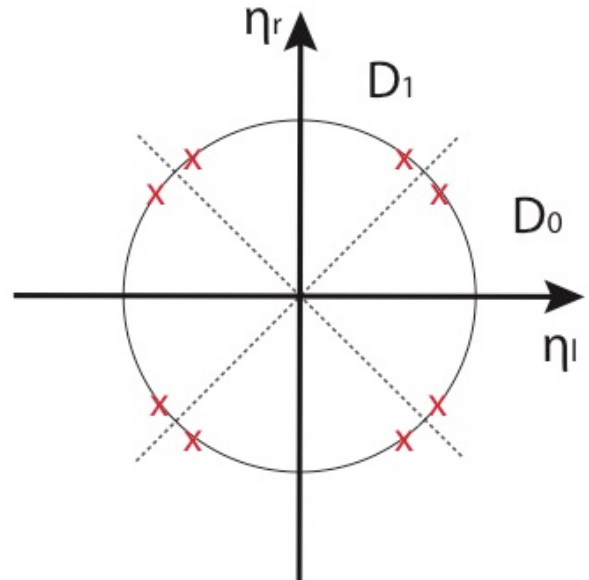


Figure 12 Symmetry of error pairs. x and y axes correspond to the horizontal errors of left eye η_l and right eye η_r . Eight crosses on the circle have the same probability.

Appendix B: Python script for simulation

```
import math

import numpy as np

# little helper to generate normed vectors

# not safe to use with vanishing vectors!

def normed(v):

    l = np.linalg.norm(v)

    return v/l

def closestpoint(p0,p1,ray0,ray1):

    # Generate projectors from rays

    R0 = np.identity(3) - np.outer(ray0,ray0)

    R1 = np.identity(3) - np.outer(ray1,ray1)

    # Solve linear system

    return np.linalg.solve(R0+R1,np.dot(R0,p0)+np.dot(R1,p1))

# graphical output requires matplotlib

graphical_output = True

# we assume units are mm

# positions of left and right eye in the plane

eye_l = np.array ((0.0 , -30.0 ,0.0))

eye_r = np.array ((0.0 ,30.0 ,0.0))

# position of object

lookat = np.array ((500.0 ,0.0 ,0.0))
```

```
# direction between eyes

eye_base = normed(eye_r - eye_l)

# up direction

# should be orthogonal to eye_base

up = np. array ((0.0 ,0.0 ,1.0))

# normalized rays from eyes to object

ray_l = normed(lookat - eye_l)

ray_r = normed( lookat - eye_r )

# normalized horizontal and vertical direction per eye

hor_l = normed(np.cross(ray_l ,up))

ver_l = np.cross(hor_l,ray_l)

hor_r = normed(np.cross(ray_r ,up))

ver_r = np.cross(hor_r,ray_r)

# variance of the angular deviation

# we sample angular deviation on a tangent plane

# and then project back to the sphere

variance_l = np. radians ([1.5 ,0.16])

variance_r = np. radians ([1.5 ,0.16])

#print eyes and object

print("Eyes at ",eye_l ," , ",eye_r)

# containers to store x and y positions of

# intersection of left and right eye ray

if graphical_output :

    x=[]

    y=[]
```

```
min_psi=[]

max_psi = []

max_s = 0.0

min_s = np.linalg.norm(0.5*eye_r + 0.5*eye_l-lookat)

# number of random measurements

trials = 10000

# center of interest , i .e. average position

coi = np.zeros(3)

for i in range(trials):

    # change of angles

    # using normal distribution

    psi_l = np.random.normal(0.0,variance_l)

    psi_r = np.random.normal(0.0,variance_r)

    # using uniform distribution

    #psi_l = np.random.uniform(-2.0*variance ,2.0*variance)

    #psi_r = np.random.uniform(-2.0*variance ,2.0*variance)

    # using laplace distribution

    #psi_l = np.random.laplace(0.0,variance)

    #psi_r = np.random.laplace(0.0,variance)

    # generate the slightly rotated eye rays

    # simply by adding the components along eye and up direction

    vray_l = normed(ray_l + psi_l[0] * hor_l + psi_l[1] * ver_l)

    vray_r = normed(ray_r + psi_r[0] * hor_r + psi_r[1] * ver_r)

    v = closestpoint(eye_l ,eye_r ,vray_l ,vray_r)

    # keep track of centroid of estimated object positions
```

```
coi += v

# keep track of smallest and largest distance

# we simply judge distance by length of s

ls = np.linalg.norm(0.5*eye_r + 0.5*eye_l - v)

if (ls>max_s):

    max_s = ls

    max_psi = [psi_l ,psi_r]

if (ls<min_s):

    min_s = ls

    min_psi=[psi_l,psi_r]

# add to container for drawing

# projection is onto the plane spanned by the first two comp.

if graphical_output :

    x . append ( v [ 0 ] )

    y . append ( v [ 1 ] )

print("Largest distance:", max_s)

print("resulting from angle variations:", np.degrees(max_psi))

print("Smallest distance:", min_s)

print("resulting from angle variations:", np.degrees(min_psi))

coi /= float(trials)

print("Centroid of estimated object positions: ", coi)

print ("Compared to object at " , lookat )

if graphical_output :

    import matplotlib . pyplot as plt

    # draw eyes and object
```

```
circle = plt.Circle(eye_l , 10.0, fill=False)

plt.gca().add_patch(circle)

iris_l = eye_l + 10.0 * ray_l / np.sqrt(ray_l.dot(ray_l))

circle = plt.Circle(iris_l , 3.0, fill=True, color='b')

plt.gca().add_patch(circle)

circle = plt.Circle(eye_r , 10.0, fill=False)

plt.gca().add_patch(circle)

iris_r = eye_r + 10.0 * ray_r / np.sqrt(ray_r.dot(ray_r))

circle = plt.Circle(iris_r , 3.0, fill=True, color='b')

plt.gca().add_patch(circle)

circle = plt.Circle(lookat , 5.0, fill=True, color='g')

plt.gca().add_patch(circle)

plt . scatter (x,y,marker='.' ,color='r' ,alpha=0.02)

plt.axis('scaled')

plt .xlim([-50,750])

plt .ylim([-50,50])

plt . margins (0.2 ,0.2)

plt.savefig('vergence.pdf' , bbox_inches='tight')

plt .show()
```


Appendix B: Mathematica notebook for analytical analysis

(*Horizontal errors only*)

$p_l = \{-a, 0, 0\};$

$p_r = \{a, 0, 0\};$

$target = \{0, 0, 1\};$

$e_l = target - p_l;$

$e_r = target - p_r;$

$epsilon_l = \{h_l, 0, 0\};$

$epsilon_r = \{h_r, 0, 0\};$

$e_{l1} = (e_l + epsilon_l) / \text{Sqrt}[1 + (a + h_l)^2];$

$e_{r1} = (e_r + epsilon_r) / \text{Sqrt}[1 + (-a + h_r)^2];$

$e_{l2} = (e_l + epsilon_r) / \text{Sqrt}[1 + (a + h_r)^2];$

$e_{r2} = (e_r + epsilon_l) / \text{Sqrt}[1 + (-a + h_l)^2];$

$E_{l1} = \text{IdentityMatrix}[3] - \text{Transpose}[\{e_{l1}\}].\{e_{l1}\};$

$E_{r1} = \text{IdentityMatrix}[3] - \text{Transpose}[\{e_{r1}\}].\{e_{r1}\};$

$E_{l2} = \text{IdentityMatrix}[3] - \text{Transpose}[\{e_{l2}\}].\{e_{l2}\};$

$E_{r2} = \text{IdentityMatrix}[3] - \text{Transpose}[\{e_{r2}\}].\{e_{r2}\};$

$p_1 = \text{Inverse}[E_{l1} + E_{r1}].(E_{r1} - E_{l1}).p_r;$

$p_2 = \text{Inverse}[E_{l2} + E_{r2}].(E_{r2} - E_{l2}).p_r;$

FullSimplify [0.5*(p1+p2), **Assumptions** -> {}][[-1]]

Out[1]= $(4. a^2) / ((2. a + h_l - 1. h_r) (2. a - 1. h_l + h_r))$

(*Vertical errors only*)

$p_l = \{-a, 0, 0\};$

$p_r = \{a, 0, 0\};$

$target = \{0, 0, 1\};$

```
e1 = target - pl;  
  
er = target - pr;  
  
epsilon1 = {0, vl, 0};  
  
epsilon2 = {0, vr, 0};  
  
e11 = (e1 + epsilon1)/Sqrt[1+a^2+vl^2];  
  
er1 = (er + epsilon2) / Sqrt[1+a^2+vr^2];  
  
e12 = (e1 - epsilon2) / Sqrt[1+a^2+vr^2];  
  
er2 = (er - epsilon1)/Sqrt[1+a^2+vl^2];  
  
E11 = IdentityMatrix[3]-Transpose[{e11}].{e11};  
  
Er1 = IdentityMatrix[3]-Transpose[{er1}].{er1};  
  
E12 = IdentityMatrix[3]-Transpose[{e12}].{e12};  
  
Er2 = IdentityMatrix[3]-Transpose[{er2}].{er2};  
  
p1 = Inverse[E11+Er1].(Er1-E11).pr;  
  
p2 = Inverse[E12+Er2].(Er2-E12).pr;  
  
FullSimplify[0.5*(p1+p2), Assumptions -> {}][[-1]]  
  
Out[2]= (1. a^2 (4. +vl^2+2. vl vr+vr^2))/(vl^2-2. vl vr+vr^2+a^2 (4. +vl^2+2. vl vr+vr^2))
```