Measuring torsional rotation with a video-based eye tracker

by

MINA MERCEDES GALIS

Donders' Institute of Neuroscience,
Radboud University

Supervisor: prof. dr. John van Opstal
Assistance: dr. Annemiek Barsingerhorn
& Jesse Heckman

January 30, 2019

Contents

List of Figures

1 Introduction

Eye movements have been studied for centuries. Franciscus Donders, a Dutch ophthalmologist, conducted a lot of research in 3 dimensional eye movement in the eighteen hundreds, before eye movement could even be recorded[1]. He had to base most of his research on psychophysical observations (the orientation of after images) and existing knowledge on human anatomy. In the years following his research many methods were devised to measure eye movement, the most important one being sceral search coils. Others include electrooculography, for which the potential difference between the cornea and fundus is measured and infra-red oculography, for which the amount of reflected intensity of infrared is measured. Scleral search coils make use of the principle of magnetic induction; a changing magnetic field induces a an alternating induction voltage in proportion to the flux of the external magnetic field. Scleral search coils are placed on the eyes as if they are contact lenses, while alternating magnetic fields of different frequencies are applied both horizontally and vertically. When the subject moves his or her eyes, the magnetic field perpendicular to the coil placed on the eye changes in the direction of motion. The induced alternating voltages in the search coils on the eye can be measured and the horizontal, vertical and even torsional components can be isolated.[2][3]

Scleral search coils have proven to be effective, but they do have their disadvantages. Because of lacrimal fluid, coils placed onto the eye can slightly shift during rapid movement. This phenomena usually results in a slight dynamic overshoot after a saccade. At first this was thought to be a characteristic of the human visual system, because rhesus monkeys did not seem to have the same kind of overshoot. Later it turned out to be a shift of the coil on top of the eye, an effect not seen in monkeys because the search coils were sewed onto the eye. Inserting scleral search coils is considered unpleasant and only done after local anaesthetic in humans. Moreover, coils are difficult to apply to patients, or to infants.

There is a need for a new, non-invasive method of 3D eye measurement. In the present time, image based methods have been shown efficient in determining the gaze direction with accuracy approaching the scleral search coils for eye cameras with a high sampling rate.[4]. However, measuring three dimensional eye movement is less trivial, as it requires the rotation around the line of sight to be measured. The position of the pupil can be measured by looking for dark intensity areas in the picture. However, the optical measurement of rotation around the line of sight, i.e. torsional rotation, requires tracking features in the iris. This can be difficult, especially for low resolution images for which less characteristics of the iris are visible. Also changes in illumination, or dark-brown eye colour, may prevent a reliable detection of iris features.

The aim of this thesis is to determine the accuracy and functionality of three dimensional eye tracking methods using video-based image tracking with a head-fixed video system provided by Pupil labs. The measured torsional rotation can

be compared with torsional rotation computed by applying known eye models to the acquired horizontal and vertical position of the eye. Comparing the measured and computed torsional rotation could give more insight in current eye models for head fixed eye movement. For the calculation of the torsional rotation in images, iris features will be tracked using a method called Maximally Stable Volumes[5]. This method was previously used for determining the amount of ocular torsion by Ong and Haslwanter (2010)[6]. The aim is to apply full three-dimensional eye-movement recordings with the video system during head-unrestrained eye-head gaze shifts and during complex vestibular stimulation in the Biophysics lab.

2 Three-dimensional eye movement

2.1 Definitions

The eye can be approximated as a spherical, rigid body fixed in the centre. This approximation neglects the parabolic shape of the cornea and the slightly rugby-like shape of the eye itself. However, these effects are rather insignificant, especially for small eye movement[7]. The combination of head and eye orientation is called the gaze orientation. The gaze direction is defined as the direction of the line of sight. When looking at a target, the line of sight connects the eyes and the target. Under normal circumstances humans move both head adn eye when orienting towards a target. Eye movement is inherently correlated to head movement, eye-in-head orientation changes during head movement.

Three dimensional eye movement is composed of horizontal and vertical rotation, but also rotation around the line of sight. The latter is called torsional rotation. We define the torsional rotation to be zero in the reference position, where eyes are looking straight ahead and the head is in the rest position. The torsional rotation stays zero for positions along the horizontal or vertical axis, called secondary positions. Only for both horizontal and vertical translation the orientation of the iris changes with respect to the primary position. This torsion is often called *pseudotorsion*, because it is dependent on the chosen coordinate system in comparison to actual rotation around the line of sight. Positions with both horizontal and vertical translation and solely pseudotorsion are called tertiary positions. Quaternary positions contain a real torsional component and they occur during sleep or during eye-head gaze shifts, during vergence eye movements or during vestibular stimulation.[8]

2.2 Vestibular system

The vestibular system takes care of balance and of stabilising gaze despite head movement, and it measures the 3D orientation of the head. It consists of three

different parts, the utricle, the saccule, and the 6 semicircular canals. The utricle and saccule detect linear head accelerations, including gravity and thus provide information on the vertical orientation of the head. The 6 semicircular canals, which detect rotational movement, are located near each ear. They are filled with a fluid that flows with a slight delay if the head is in motion. Because of Newton's second law, this delay causes the fluid to exert pressure on the sensory receptors that cause stereocilia on the hair cells to bend. In healthy individuals the canals should send a consistent signal. Individuals with malfunctioning vestibular organs experience vertigo.[9]

2.2.1 Vestibulo-ocular reflex

The vestibulo-ocular reflex controls eye movement during activation of the vestibular system. It ensures the world is seen as steady by controlling eye movement with respect to head movement. The eyes are corrected by rotating in the opposite direction of the head and thus minimising the motion of the image on the retina. The vestibulo-ocular reflex does not only occur for translations of head position, but also for acceleration.[10]

2.3 Three-dimensional rotations

The eye is both mechanically and neurally constrained. Mechanically, the eye is constrained by the extra-ocular muscles and surrounding tissues, but it is also constrained by neural control mechanisms in order to create a stable view of the world despite eye movements. When the head passively rotates about the horizontal interaural axis axis (pitch), the eyes rotate about the same horizontal axis in the opposite direction to exactly compensate for the rotation of the head and to stabilise the image on the retina. Head rotation about the frontal axis produces a compensatory counterroll of the eyes, which is called ocular torsion. Torsional compensatory eye movements are obtained when the head rotates passively about the frontal axis. The maximal amount of ocular torsion is approximately 20 degrees, due to the mechanical constraint imposed by the eye muscles and by the optic nerve leaving the back of the eye. The correlation between head and eye rotation about the line of sight is fixed, as for each position of the head there is a unique value of torsion with respect to the position of the eye. The steady fixation on the target minimises the motion on the image of the retina [11].

Rotations, unlike translations, do not commute. This property of rotations complicates calculations, because the final orientation of a spherical body like the eye depends on the order of the rotations. A rotation about the vertical axis followed by a rotation around the horizontal axis does not lead to the same final orientation as the other way around. The noncommutativity of rotations complicates calculations.

Because the boundary of the eye globe is a 2 dimensional manifold, we have to account for parallel transport. When moving a vector along a geometry, the vector can change its orientation. Since rotations do not commute, the orientation of a vector at a certain point on the geometry depends on the path taken by the vector. [13] As such, we expect to see that the orientation of the eye would also depend on its path, if it were to rotate about arbit axis. However, it has been repeatedly observed that the orientation of the eye, when directed at a specific point is independent of the previous position and path it took to get there. This indicates that the eye corrects for the torsion due to parallel transport, by rotating about the axis of the line of sight in the opposite direction. The eye does not actually produce an active counterroll movement to compensate for the result of parallel transport, but theoretically one could describe it in this way. Instead, the eye chooses a rotation axis in which the orientation of the eye does not accumulate torsion. This phenomenon is known as Donders' law and will be further explained in section Donders' law. First, we describe some general algebraic properties of 3D rotations.

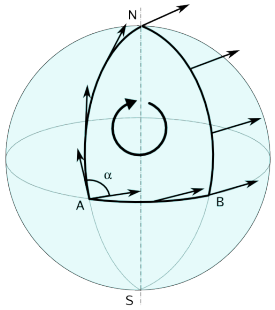

Figure 1: Intrinsic curvature of the sphere: if you parallel-transport a vector around the closed loop ANBA, the vector is shifted proportionally to the area delimited by the closed loop and to the curvature of the sphere[12].

2.3.1 Euler's rotation theorem

All possible orientations of a rigid, spherical body can be described by 3 Euler angles, namely θ, ϕ and ψ, that result from rotation in particular order around particular body-fixed axes. In a commonly used Euler system, these angles are defined in the right hand sense, and result from anticlockwise rotations about the first, the earth-fixed z-(vertical) axis, then about the rotated body-fixed y-(horizontal) axis and finally about the body-fixed, double-rotated x-(frontal) axis, respectively. Since the x-axis is defined to be along the body-fixed line of sight, rotation about ψ is referred to as torsional rotation. Any combination of horizontal and vertical rotations uniquely determines the gaze direction, but not the rotation about the line of sight.

Two frequently used gimbal systems show the difference in torsion of a fixed final body orientation, for different orders of the horizontal and vertical rotations. The above-described Euler system is the so-called Fick-gimbal (θ_F, ϕ_F, ψ_F). The Helmholtz-gimbal rotates first about the earth-fixed horizontal axis with ϕ_H, then about the body-fixed vertical axis with θ_H and finally about the line of sight, ψ_H[8]. In the *active notation*, the rotations are around the body-fixed

axes, as described above, and the overall rotation is described by:

$$R_{Fick} = R_z(\theta_F)R_y(\phi_F)R_x(\psi_F) \tag{1a}$$

$$R_{Helmholtz} = R_y(\phi_H)R_z(\theta_H)R_x(\psi_H) \tag{1b}$$

For equal horizontal and vertical rotation angles $\theta_F = \theta_H$, $\phi_F = \phi_H$, the final orientation of the sphere is equal if and only if $\psi_F = -\psi_H$[14]. Because the torsional angle ψ depends on the choice of coordinate system, it is often referred to as *pseudotorsion*, or *false torsion*. According to Euler's theorem all sequences of rotations of a body around a fixed point can be written as single-axis rotation around $\hat{n}$ through the fixed point with angle ρ. All points on this single axis are left invariant by the rotation, and therefore the axis is the eigenvector of the rotation matrix. The size of this eigenvector relates directly to the amount of rotation.

2.3.2 Donders' law

Fransiscus Donders was a famous ophthalmologist, to whom the Donders Institute of Nijmegen owes its name, and who laid the foundations of research on three-dimensional eye movement, together with his German contemporaries Hermann von Helmholtz, and Johann-Benedict Listing. In his time, all measurements had to be done by psychophysical observations. He devised a way to experience the rotation about the line of sight firsthand by using the phenomenon of *after images*. If a subject looks at a red cross, with its arms along the secondary directions (horizontal and vertical), for a prolonged amount of time, over-stimulating the rods and cones in the retina of the eye causes the signal to the brain to decrease. If the rods are no longer stimulated the rods and cones submit a negative signal to the brain. This leads to a 'afterimage' in the complementary color, in this case green. If the eye moves, the orientation of the cross changes with respect to the primary position, but the afterimage is imprinted on the retina and therefore does not rotate. Donders build his own eye model, called a '*phaenophthalmotrope*', which he used to explain the laws of eye movement to his students. This eye model could be rotated in three different directions, namely the three Euler angles, but it only had two degrees of freedom.[15]

For a stationary, upright head, Donders' law states that the orientation of the eye, i.e. the amount of torsion around the frontal, earth-fixed axis, is unique for each direction of the eye. This orientation is irrespective of the original position and of the path taken by the eye to get there. As we will see below, we can express the orientation of the eye by a single-axis rotation, which can be represented by a so-called 3D rotation vector, $\hat{n}$. The horizontal and vertical orientation of the eye's rotation axis, given by, r_y and r_z respectively, then uniquely determine the torsional component r_x.

$$r_x = f(r_y, r_z) \tag{2}$$

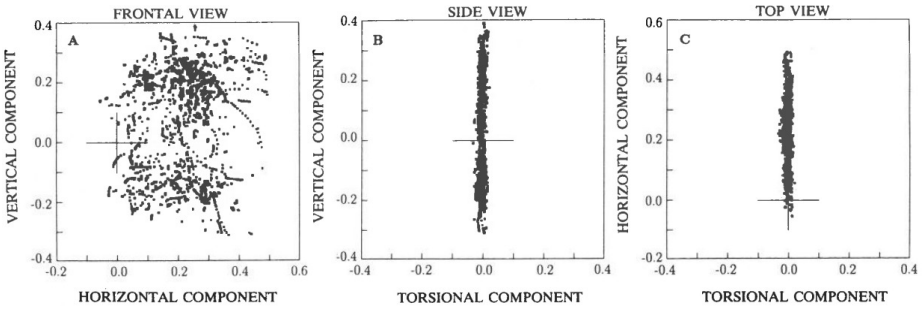

Figure 2: Experimental validation of Listing's law. Endpoints of eye positions are plotted for each component and the black line indicates the primary position. Measurements performed on primates using scleral search coils.[8]

Note that Donders' law constrains all possible eye orientations to a 2D surface. Donders' law also holds during active head movements. In that case the components r_i for $i = x, y, z$ of the eye direction have to be substituted by gaze direction. However, when determining the function f that specifies the relation between the components, head movement complicates the situation, and we will not discuss this topic in this thesis. It does not hold during vestibular-evoked passive head rotations, which allow the eye to use all three degrees of rotational freedom.

2.3.3 Listing's law

Listing's law is an extension on Donders' law, and is valid only for head-fixed eye movement between far targets, and specifies the amount of torsion. Which is according to Listing's law, the amount of rotation around the head fixed frontal axis. However, torsional rotation emerging from the choice of coordinate system - pseudotorsion - can be nonzero (see above). According to J. B. Listing, all rotation vectors which characterise eye orientations with respect to the reference position lie in a plane, now known as Listing's plane. If the (e_y, e_z)) plane of the reference coordinate system (the so-called primary reference frame) is chosen to align with Listing's plane, all rotation vectors have a zero torsional component by definition.[8]

$$r_x = q_x = 0 \tag{3}$$

The primary position is the vector perpendicular to Listing's plane. It is not to be confused with the reference position or gaze straight ahead position in the laboratory, for Listing's plane results to be tilted by a couple of degrees backward with respect to the head-fixed coordinate system[14]. Only by measuring Listing's plane for many eye orientations, the primary position can be determined. In this thesis report, we try to investigate the validity of Listing's law and Donders' law during head-restrained eye movements, by using a

video-based infrared eye-recording technique. To detect the amount of torsion in the camera's images, we needed to resort to a sophisticated computer-image technique, which is now first described.

3 Maximally Stable Volumes Method

3.1 Method

The main purpose of this thesis is to determine the reliability of the Maximally Stable Volumes (MSV) method in determining ocular torsional rotations from camera-based images of a single eye. Rotations around the line of sight can be measured by tracking characteristic features in the iris around the pupil. For a fixed head, the torsional rotation that is measured by a head-fixed camera, placed in front of the eyes, is related to pseudotorsion, as mentioned earlier, and does not directly represent the actual torsional rotation around the line of sight in the primary coordinate system of Listing. However, the amount of expected pseudotorsion according to Listing's law can be determined with respect to the position, once we know the particular geometrical relationships. Provided that Listing's law is valid (which we can safely assume for normal-sighted subjects), the measured amount of pseudotorsion can be compared with the expected amount of torsion.

3.1.1 Pupil detection

The pupil is relatively easy to detect, since both iris and pupil are darker than their surroundings. Pupils are found to be homogeneous dark intensities. By applying thresholds on pixel values, the pupil can be detected by standard software algorithms. Pupil detection is key to both determining the horizontal-vertical eye position and the amount of torsional rotation. Two of the most common gaze estimation methods are described below.

3.1.1.1 Ellipse fit

Ellipse fit gaze estimation is a shape-based method facilitated by the fact that a circle will appear elliptical when viewed under central projection[16]. By determining the pupil centre by intensity thresholds and the boundary of the pupil using edge detection techniques, the pupil can be approximated as an elliptical shape with a certain centre position, eccentricity and angle. Those three parameters combined with an eye model determine the gaze direction. Elliptical fits are often poor in eye corners due to smaller intensity differences between the pupil and the iris. Shape-based methods require high contrast images in order to determine the pupil ellipse fits in the eye corners. [17]

3.1.1.2 Feature-based estimation

Another approach to determine the gaze direction is to look at eye features, such as cornea reflections. Incoming light can be reflected on different parts of the eye, namely the cornea, sclera and the start and end of the lens. Each reflection enters the camera with a slightly different direction due to the curvature of the eyeball. The four reflections are called Purkinje images (see figure 3). Light that is reflected back from the cornea is called a glint. This manifests itself in a spot of very light intensity visible on the pupil (see the first Purkinje image in figure 3). The image of the pupil centre comes from within the lens (see the fourth Purkinje image in figure 3). By placing the infrared source slightly above or underneath the line of sight, there is a difference in direction between cornea and pupil reflection. Placing the source on the line of sight causes a so called *bright pupil*, the entire pupil appears to be lit up. Both dark and bright pupil phenomena can be used for eye tracking. For bright pupil tracking, the eye pupil appears to be lit up when the infrared source nearly lines up on the optical axis. Whenever a subject looks away from the source, the pupil darkens and a glint appears. For dark pupil tracking the gaze direction of a participant can be determined by calculating the vector formed by the angle between the cornea and pupil reflection and calibrating the mapping between that vector and the gaze direction. An artefact of the Purkinje-image method is that the eye lens vibrates immediately after a saccadic eye movement, which then shows up as oscillations in the recordings, immediately after the eye-movement that are not related to the a actual eye movement.

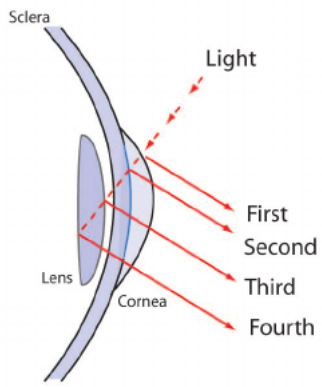

Figure 3: Light is reflected on the eye and results in various Purkinje images.[17]

3.1.2 Polar transform

Rotation of the eye is more difficult to measure than translation, because the iris features that are tracked will not only change in location, but also in orientation and thus in shape. Therefore the next step of the MSV method is to polar transform the iris of the eye. Basically, the iris is cut open on the right side ($\psi = 0$) and rolled out on the horizontal axis. This creates a plot of radius against the angle. The pupil ellipse fit, used to determine the gaze direction, is also used to extract the iris around the pupil, by cutting out a slightly enlarged ellipse around the centre of the pupil. The extracted elliptically shaped iris is then stretched into a circle such that the radius is equal for every value of the

angle, this is achieved by stretching the minor axis until it matches the size of the major axis. Subsequently, the radius in the image can be plotted against the angle, to form a polar plot of the image with the centre of the pupil as the origin. The number of pixels for the vertical axis of the polar plot (the radius) remains equal to the number of pixels between the centre of the pupil and the boundary of the iris in the original picture. However,the number of pixels on the horizontal axis (the polar angle) can be chosen, since the number of pixels in the actual picture depends on the radius. In the paper of Ong and Haslwanter (2010) the number of pixels was set to 1000, meaning that a translation of 250 pixels corresponds to a rotation of $\frac{\pi}{2}$ radians. Increasing the pixel density improves the accuracy of the amount of torsional rotation to some extent, but stretching the image too much results in blurry images.

3.1.3 Measuring torsional rotation

A polar plot, a function of polar coordinates with radius r as function of angle θ, allows an easier tracking of torsional movement, as it translated rotation around the line of sight into a translation of pixels parallel to the horizontal axis. In order to find the features in the image that would allow to determine the torsional rotation of the eye with respect to the primary position, a method called Maximally Stable Volumes (MSV) was used. This method, described in Appendix Maximally Stable Volumes, detects volumes of high intensity that change little in size between consecutive images. It is an extension on the Maximally Stable Extremal Regions (MSER) algorithm, which detects stable 2-dimensional regions in a single image. MSV is commonly used on video recording, where the third dimension of the volume is time t (or frame number) instead of a spatial dimension. This allows for tracking features in 2-dimensional images that continue over time. Torsional rotation is small in comparison to the large movements of the eye and the position of the iris changes the angle of the reflected IR light and its intensity. The tracking method should be stable under these conditions. MSV has shown to be robust under geometric transformations and these varying light conditions by Matas et al[18]. For detecting the maximally stable regions the VLFeat MSER function of Vedaldi and Fulkerson (2008) was used[19]. This function has the extension to higher dimensions built in, making it very suitable for tracking features in videos, necessary for determining torsion by tracking iris features.

3.1.3.1 Maximally stable volumes

The Maximally Stable Volumes (MSV) approach is an extension on the Maximally Stable Extremal Regions (MSER) devised by [18]. The method looks for extremal regions which stay relatively stable over time. See Appendix Maximally Stable Volumes for a more formal definition. The VLFeat toolbox has a maximally stable extremal regions programme with a built in MSV function[19]. This function regards time as the third dimension. For each frame, a polar plot

figure is entered into the programme and the centroids in time of the maximally stable volumes is given as output. Taking the difference between the centroids in time for each volume and taking the median over the differences of all volumes, gives a value for the torsional rotation in each frame. The more volumes (features) are found, the more accurate the end results will be. The median is taken instead of the mean to ensure that incorrect regions, such as eye-lashes, will not contribute with a large weighting factor in the end result.

3.2 Implementation of the method

Experiments are needed to ensure whether the programme written to determine torsion is accurate and reliable. In the preliminary study of implementing the MSV method, eye movement recordings during passive vestibular, sinusoidal head movement around the frontal head-fixed axis were used to test whether the method gave the expected results based on the literature. Eye movement during sinusoidal head movement was used, because the changes in torsion per time unit would be relatively large and the expected compensatory torsional values in time can be easily estimated by looking at the induced head movement by the vestibular chair. All eye movements in this thesis were recorded by a 120 Hz Pupil Labs eye tracker[20].

Eye movement is investigated for fixed head coordinates and therefore eye and gaze orientation are identical. Humans typically use both head and eye movements when looking at targets in peripheral space. One could move his or her eyes independently from head movement, but without an actual physical constraint on the head there will always be a small head movement, even for relatively small eye movements Measuring and describing three dimensional eye orientation in the head was the main objective for this thesis. In addition, it is necessary to determine the head position to determine the validity of the measurement.

3.2.1 Pupil labs eye tracker

The Biophysics department has two different models of Pupil labs eye trackers in their possession, three versions of an older models with a sampling rate of 120 Hz and one newer 200 Hz model. All the older versions use a 820 nm infrared source, but the newest tracker has a different headset for which the infrared source lies near the line of sight of the reference position. As discussed in section Bright pupil, placing the infrared source near the line of sight can cause a bright pupil and therefore bad ellipse fits. The 820 nm lies close to the visual spectrum and can cause a red glow when subjects are placed in the completely dark vestibular chair chamber. This glow could prevent a nystagmus from occurring, because the subject would have a reference point, even though it is fixed on the subject's head. The latter is irrelevant for experiments on torsional rotation.

The most recent Pupil lab eye tracker has a 900 nm infrared source, which lies further away from the visual spectrum. It reduces the red glow, but it further decreases the contrast between iris features. Low resolution images contain only roughly three grey scales, making it impossible to track iris features. An eye tracker with a 900 nm IR source is therefore not suitable for the MSV method in darkness, but may be used for experiments on the vestibulo-ocular reflex in the vestibular chair.

3.2.2 Pupil ellipse fit

The first part of the MSV method is the pupil ellipse fit. This ellipse fit is necessary, not only to determine the position of the eye, but also to extract the iris correctly. Applying the MSV method to the entire frame would be very memory extensive and in order to determine the amount of torsion the iris features would still have to be distinguishable from all other extremal volumes afterwards. Several pupil detection methods have been tested for accuracy, speed and the number of bad ellipse fits. The first method is an algorithm previously used by Ong[6]. This programme only works for large intensity differences between the pupil and the iris and is therefore more likely to give bad fits in eye corners due to inferior lighting conditions. Figure 4 gives an example of the Ong method applied to a single frame with the pupil in the downright corner. The Ong method causes the pupil fit to be slightly bigger and more elliptical than the actual pupil, therefore containing large parts of the iris.
Figure 4 shows an extreme case of an incorrect pupil fit due to the decrease in

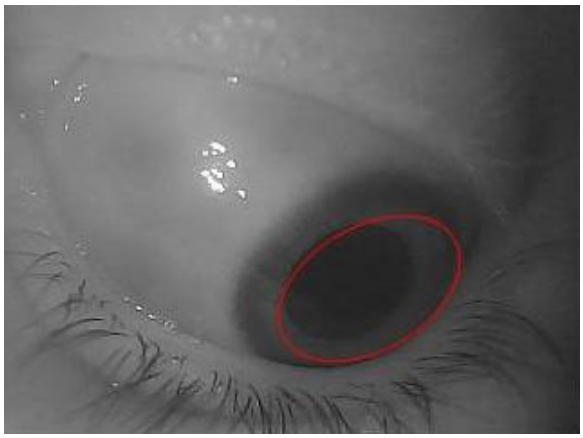

Figure 4: The pupil is over-fitted due to a decrease in intensity difference between pupil and iris in the eye corners.

iris intensity in the corners of the eye. Over-estimating the eccentricity of the pupil results in an over-estimation of the gaze direction with respect to the reference position. For we have used the Pupil labs gaze direction estimate, the Ong

pupil fit is irrelevant with regard to the direction. However, over-approximating the size and eccentricity of the ellipse give rise to a bad polar transformation of the iris. In figure 5 a wavy pattern of the iris is visible. Due to the incorrect

Figure 5: Over-estimated pupil due to corner position of eye shown in a polar plot causes the iris to display a sinusoidal pattern.

fit of the iris becomes a sinusoidal pattern. This does not only lead to irrelevant vertical differences, but extremal volumes also lose their stability and the amount of torsion becomes less accurate considering there are less maximally extremal volumes to track. Stretching out the horizontal axis over a sinusoidal pattern also gives rise to errors in the amount of torsion, because the angle of the features in the over-estimated pupil no longer corresponds to the angle of the features for the actual pupil.

Another possible method is to use the ellipse fits created by Pupil labs. One big preliminary advantage of this method is the fact that these fits are already determined and saved together with the eye position, therefore MSV method should increase in speed. Pupil labs also has a built in confidence measure that can be used to filter out bad ellipse fits. Only high confidence pupil labs ellipse fits are used, together with manual pupil ellipse fits from the James Ong method.

3.2.2.1 Bright pupil

Three 120 Hz Pupil labs eye trackers were tested for tracking eye movement. They did neither differ in camera, nor in settings, but rather in eye-tracker headsets. For one of the eye trackers, the infrared source of the headset is placed near the line of sight for the reference position, causing a partly bright pupil as described in Feature-based estimation. All experiments were performed using the other two 120 Hz eye trackers.

The right side of the pupil in figure 6 is not fitted properly due to the edge detection technique used. This could result in loss of eccentricity, thereby causing the calculated position to be slightly more off centre. Another possibility is that the pupil fit remains the eccentricity of the actual pupil, which would cut off parts of the horizontal boundary of the pupil, resulting in a sinusoidal pattern of the iris in a polar plot. In the eye recordings there is also a large glint visible, but that can be easily filtered out by size.

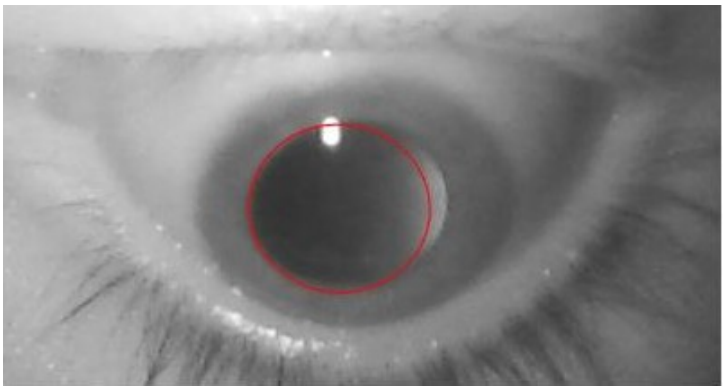

Figure 6: Partly bright pupil using a pupil labs eye tracker. The pupil fit fails due to the edge detection technique used.

3.2.3 Polar transformation

As shown in figure 7, there is a dark area on the top of the plot representing the pupil and two light arches on the bottom representing the eyelids. Polar transformations themselves are basic geometrical transformations. However, the preprocessing steps for the MSV can have a large influence on the end result. In order to ensure that the extremal regions of the pupil and eyelids do not contribute to the amount of torsion, they have to be cut out of the polar plot. Using edge detection on the polar plot, the pupil can be extracted quite easily. However, because of pupil dilation the edge of the pupil is not fixed and therefore fixed ranges are preferable. By using only a small portion near the top part of the plot of approximately 100 pixels, the MSV method increases in speed and the eyelids are not included.

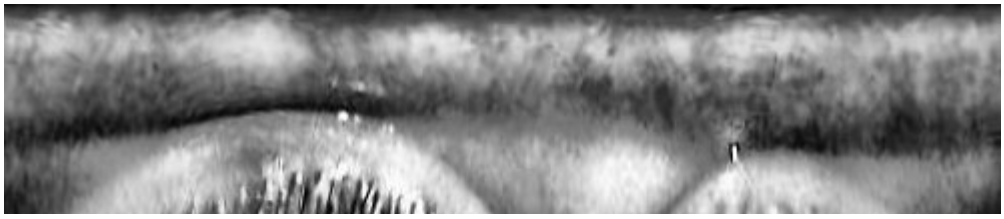

Figure 7: A polar plot made by extracting the enlarged ellipse surrounding the pupil and applying an affine transformation that transforms it into a circle. The largest part of the pupil on top is extracted and the number of pixels on the vertical axis is set to 500 (using linear interpolation). The entire horizontal axis corresponds to 2π

Histogram equalisation is applied to the cropped polar plot of the iris, which

increases the relative contrast in the picture. The fact that dark iris features near the pupil become indistinguishable from the pupil itself is irrelevant, as large maximally stable volumes are extracted by setting boundaries on the maximum size of the volumes. Similarly, light iris features near the glint do not contribute to the torsion estimate either. Increasing the contrast between the iris features, makes it easier to detect extremal regions.

3.2.4 Maximally stable volumes

An important aspect of improving the MSV method involves tuning the parameters. The speed and accuracy of the method relies on the sensitivity of the extremal regions, the maximum area and the number of pixels of the images.

Because the range of the polar plot used for the MSV method is fixed there could still be some dark regions near the top of the image during very large pupil dilation. Extremal regions near the top of the image should not contribute to the torsion estimate. Also extremal regions near the vertical boundaries should not be included, since regions on the boundary would be counted twice due to the cyclic boundary conditions. Because of these boundary issues the method is programmed to overlook regions close to the boundary, i.e. at least 5 pixels away.

The maximum size of the maximally stable volumes matters for both accuracy and speed. Larger maximally stable regions might not be single features, but e.g. sections with different shadings or glints. In order to improve the accuracy, the maximum size must be kept such that the maximally stable volumes contain single features. The minimal number of frames in which regions return can be lowered or raised to find more or less maximally stable volumes respectively. When lowering the minimal number some accuracy gets lost due to maximally stable volumes that represent shading instead of iris features. However, some accuracy is gained by the fact that more maximally stable volumes can be found over which to take the median.

The stability criterion is a measure of the stability of stable regions or volumes. By minimising the stability criterion the most stable volume can be found for a certain position, namely the maximally stable volume. Parameter Δ is a measure of the sensitivity of the stability of the maximally stable volumes (see equation (28)). Increasing the sensitivity Δ decreases the number of local minima and thereby the number of maximally stable volumes. In figure 9 the parameter Δ is decreased with respect to figure 8

In figure 8 polar plot images are shown for three different frames in time. For these regions a higher sensitivity was used, resulting in the detection of only a few maximally stable volumes. The large maximum area setting causes the number of volumes to decrease even further, which the figure illustrates. The volumes on the right side of the figure seem to display torsion, whereas the light

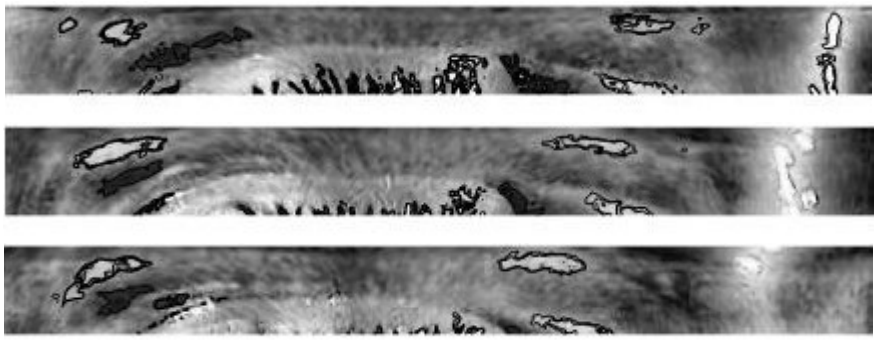

Figure 8: Maximally stable volumes for high maximum area and minimal duration parameters, displayed for three different frames that are less than a second apart. Some of the volumes seem to display a change in torsion. The horizontal axis corresponds to 2π.

intensities on the left do not. The dark features on the left do seem to vary with torsion. Because the median is taken over all volumes, the MSV method applied to these frames would result in an overall anticlockwise torsion.

It is clear from figure 8 that there are many iris features much smaller than the highlighted regions. Decreasing the maximum area of the regions by a factor ten gives regions that resemble the iris features more precisely. Having more maximally stable volumes results in a more stable method, as regions that represent other features, e.g. eye-lashes, eyelids, will have a lower weighting factor in the calculation of torsion. In figure 9 the same frames are plotted for a smaller maximum area and lower sensitivity.

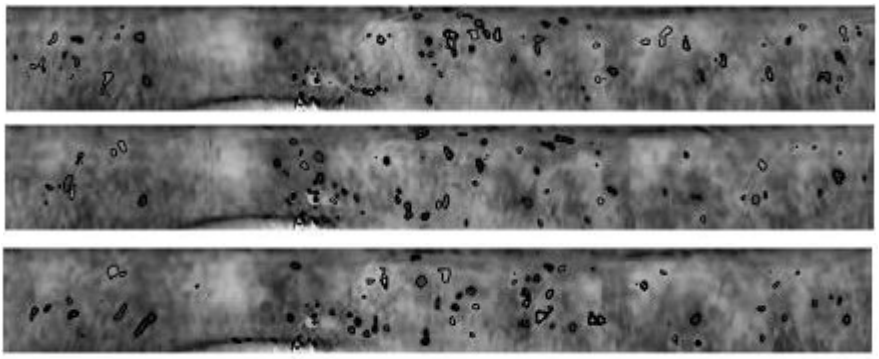

Figure 9: Maximally stable volumes for small maximum area and minimal duration parameters, shown for the same three frames as in figure 8

3.2.4.1 Sections

In order to detect iris features, high-resolution images are needed. However, applying MSV to images is rather memory extensive. The MSV method searches for volumes that occur throughout the entire video, hence running time increases exponentially for larger videos. Therefore the full video is cut into shorter pieces. To determine the values of torsional rotation either every section has to start with a frame of the same position of the eye each time or the sections have to overlap slightly. The first method is more efficient, but only possible for visual tasks with built-in calibration in controlled environments. The second one requires tracking all sections between the starting point and frame of interest. If the video is separated into small sections a bit of accuracy gets lost due to the fact that fewer maximally stable volumes can be found near the boundaries of the sections, but an immense amount of speed is gained. The errors on the boundaries can be slightly overcome by overlapping the sections by a couple of frames, e.g. 4 frames. This is not only a solution to the boundary issue, but it also allows each section to be determined more precisely with respect to its neighbouring sections. This technique was previously used by Ong and Haslwanter (2010)[6].

Using the most optimal values for the parameters can vary across individuals. Participants with highly noticeable iris features can be tracked more easily. However, sensitivity should be increased to give the prominent features a higher weighting factor. Size and shape of the eyes of the participant change the amount of influence of the eyelids in the polar plot and thereby the accuracy of the MSV method. The amount of pupil dilation in response to light intensity changes and the size of the iris differs for each participant, hence the range of the polar plot has to be tuned individually to prevent an influence of the pupil or eye lids. Ethnicity of the participant could also determine the amount of contrast between iris features. Although we have not tested this in this report, we expect that lighter eye colours can be more easily tracked using the MSV method than darker eye colours.

4 Validation of the method

The main goal of this thesis is to determine whether MSV is a robust and efficient method for tracking iris features. Experiments should point out the optimal image resolution and parameter values for this method and the accuracy with respect to the expected torsional values using Listing' Law. Two differed experiments were performed. The first experiment is dedicated for quantitative research on the MSV method and the second experiment is a intended to show the significance of being able to measure torsional rotation.

4.1 Sphere

It has previously been possible to measure torsional rotation by means of scleral search coils. However, due to the invasive property of scleral search coils, measured torsional rotation using the MSV method cannot simply be compared with torsion measured using coils. The only way to verify the measured torsion is to compare them with the expected value with respect to the position using Listings' law or with torsional nystagmus using torsional vestibular stimulation. Since the purpose of this thesis is to determine the accuracy of the MSV method in determining 3D spatial direction, the method should be able to measure torsion under different horizontal-vertical eye directions. Therefore, the sphere experiment is needed. This makes it harder to correctly determine the accuracy of the MSV, because it also depends on the accuracy and validity of the estimated eye direction. Although it can be calibrated, they could still contain small errors. It also puts constraints on the required experiment, because Listings' law only holds in specific conditions. The head has to be fixed, preferably mechanically, because eye and head are accustomed to move at once. If not mechanically, the position of the head needs to be measured to determine the validity of Listings' law on the acquired data.

4.1.1 Setup

Participants were instructed to sit in front of a half sphere with homogeneously placed LED's. First, the eye tracker had to be calibrated. Because the eye tracker consists of both an eye an world camera, this could be done either by head fixed calibration of multiple points in space or using a single point calibration where the eyes fixate on a point while moving the head. The latter calibration is easily executable in different places, because it does not require a large screen attached to a computer. For all participants a single point calibration was used, which Pupil labs immediately implements on the acquired data. Secondly, the position of the head needs to be calibrated to determine the movement of the head during the experiment. The calibration is done by asking the participant to fixate on the fixation LED in the centre of the sphere and move their head towards different LED's that were switched on at known positions. To ensure the head pointed directly at the target, the participant was instructed to keep the eyes fixed on the both the LED and the end of a lightweight aluminium bar fixed on the head tracker. The alternating voltage of the coil of the head tracker can be measured, to determine its horizontal and vertical components.

After both calibrations the actual experiment was performed. The experiment is similar to the head calibration, except that the participant was asked to look at the illuminated LED's while keeping the head still, thus moving only the eyes. The LED's were placed on at least a meter away from the eyes, therefore we can assume the eyes do not converge. In total 64 LED's appeared at different

azimuth and elevation locations and disappeared after the participant pressed the trigger button. Every LED was switched on once during the experiment. The measured torsional values can be compared with the expected torsional value of the mean of the position measurements according to Listing's law.

4.1.1.1 Fick Gimbal

The amount of pseudotorsion seen from a camera fixed in front of the eyes for a specific position is equal to the amount of torsion resulting from the Fick-gimbal system for said position. The torsion resulting from the Fick- and Helmholtz-gimbal system only differs in sign and therefore it is only a matter of definition which gimbal is chosen as the standard. Equation (1a) in section Euler's rotation theorem can be transformed into a matrix using the appropriate 3 dimensional rotation matrices for the three Fick angles θ_F, ϕ_F, ψ_F.

$$R_{Fick} = \begin{bmatrix} \cos\theta_F \cos\psi_F + & \cos\phi_F \sin\theta_F & \sin\psi_F \cos\theta_F - \\ \sin\theta_F \sin\phi_F \sin\psi_F & & \sin\theta_F \sin\phi_F \cos\psi_F \\ \\ \sin\theta_F \sin\phi_F \cos\phi_F - & \cos\phi_F \cos\theta_F & -\sin\psi_F \sin\theta_F - \\ \cos\psi_F \sin\theta_F & & \cos\theta_F \sin\phi_F \cos\psi_F \\ \\ -\sin\psi_F \cos\phi_F & \sin\phi_F & \cos\psi_F \cos\phi \end{bmatrix} \tag{4}$$

The result for R_{fick} in 4 has been obtained by Nakamaya (1974)[21]. Matrix 4 should be equal to the transformation matrix $R[\hat{n}, \theta]$ (see equation (27)) derived from the quaternion transformation in Appendix Quaternions. In order to determine the amount of pseudotorsion ψ_F either $R_{Fick,31}$ or $R_{Fick,33}$ can be used in comparison with the respective element of the quaternion transformation matrix $R[\hat{n}, \theta]$.

$$R_{31} = -\sin\psi_F \cos\phi_F \tag{5}$$

$$\psi_F = \arcsin \frac{R_{31}}{\cos\phi_F} \tag{6}$$

Note that $R_{31} \equiv R_{31}(\theta_F)$ and that $\theta \neq \theta_F$, since θ_F is the rotation about the z axis in the Fick gimbal system while θ is the rotation about axis $\hat{n}$. For full specifications see Appendix Quaternions.

The corresponding matrix element of the quaternion transformation is as follows:

$$R_{31} = 2(q_1 q_3 - q_0 q_2) \tag{7}$$

$$= 2\sin\theta(\sin\theta n_1 n_3 - \cos\theta n_2) \tag{8}$$

Using equations (8) above the amount of pseudotorsion can be determined with respect to the eye position.

From the Fick gimbal matrix (4) the rotation axis and angle θ can be determined using the following equations.

$$\cos\theta = \frac{1}{2}(R_{11} + R_{22} + R_{33} - 1) \tag{9}$$

$$\hat{n} = \frac{1}{2}\csc\theta \begin{bmatrix} R_{23} - R32 \\ R_{31} - R13 \\ R_{12} - R21 \end{bmatrix} \tag{10}$$

Similarly, the quaternion components for the single-axis rotation can be determined by comparing (4) and (27).

$$q_1 = \frac{1}{4}(1 + (R_{11} - R_{22} - R_{33})) \tag{11}$$

$$q_2 = \frac{1}{4}(1 + (R_{22} - R_{33} - R_{11})) \tag{12}$$

$$q_3 = \frac{1}{4}(1 + (R_{33} - R_{11} - R_{22})) \tag{13}$$

4.1.1.2 Listing's law

From the recorded eye movements both the gaze direction and the amount of torsion can be determined with the algorithms described above. It is important to remember that the acquired torsion is *pseudotorsion*, or torsion arising from use of the Fick gimbal system. Therefore it is not the real torsion that is used to determine Listing's plane. Since, the subject was instructed to move the eyes while keeping the head fixed, it is expected that there is no real torsional component. First the expected pseudotorsion should be calculated according to the Fick gimbal at the acquired position. Secondly, this calculated torsion should be subtracted by the amount of torsion determined by the MSV method. This leaves the real torsion about the line of sight, independent of the coordinate system, used for plotting Listing's plane. From the azimuth, elevation and torsion, one can determine the rotation quaternion q using (24) with respect to the reference position. The rotation vector $\hat{n}$ and single axis angle θ can be determined using (20). The components of the rotation vector can be plotted against eachother to determine Listing's plane and thus the primary position.

4.1.2 Results

During the first experiment both eye and head movement were recorded. The head movement was recorded to verify whether the participant kept the head immobile and upright, in order to comply with the requirements for Listing's law. From the eye movement both gaze direction and torsion were determined. It is important to remember that the acquired torsion is *pseudotorsion*, or torsion arising from use of the Fick gimbal system. Therefore it is not the real

torsion that is used to determine Listing's plane. The validity of Listing's law can be determined by measuring the amount of head movement during saccades. In figure 10 the gaze position is plotted against the eye position, with little to no variation. The first experiment consisting of 64 stimuli took about five min-

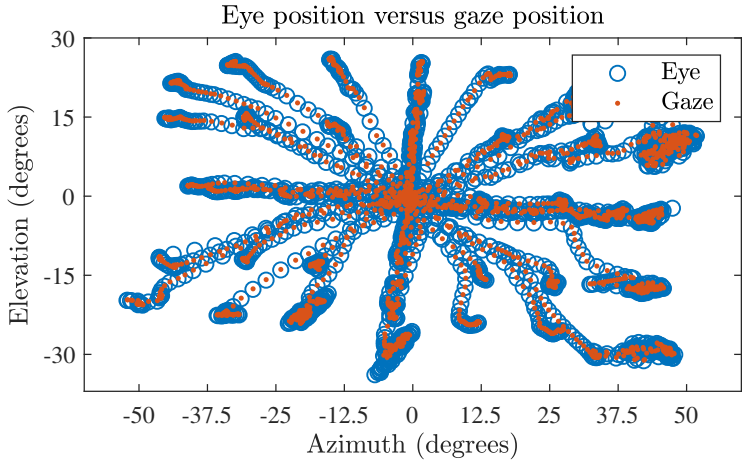

Figure 10: Eye position plotted against gaze position. Deviation of the gaze position caused by head movement are small, thus Listing's law can be applied

utes to complete. Considering 120 frames per second were recorded by the eye tracker, the total number of frames that had to be checked for maximally stable volumes was $120 \cdot 60 \cdot 5 = 36000$. Obviously, this is a rather memory extensive procedure and it requires a lot of time to search through the pixels of every frame. We have therefore divided the video into smaller sections of 90 frames, see section Sections.

4.1.2.1 Double trigger

The succes of the MSV method relied strongly on the quality of the pupil ellipse fit. Since the method relies on at least some of the maximally stable volumes to remain connected over time and the method is cumulative, bad ellipse fits during eye movement towards the corners or blinks will cause problems. To ensure that the torsional values can be compared with each other, the amount of torsion has to be *reset* to zero for the reference position several times during the entire video. This *reset* is used to maintain the amount of torsion relative to the reference position, therefore this can be done by instructing the participant to press a button every time she looks straight ahead. Every time the button is pressed while looking at the reference position after a saccade, the torsion can be reset. Several experiments have been conducted with a single button press, causing a trigger in the dataset,when a participant was looking at a LED in space. However, we have found that for accurate measurements the experiment

had to contain triggers not only for different positions in space, but also for the straight ahead position (torsional starting value). This way the amount of torsion can be reset to zero after each trigger, ensuring the amount of torsion for each stimuli is properly determined.

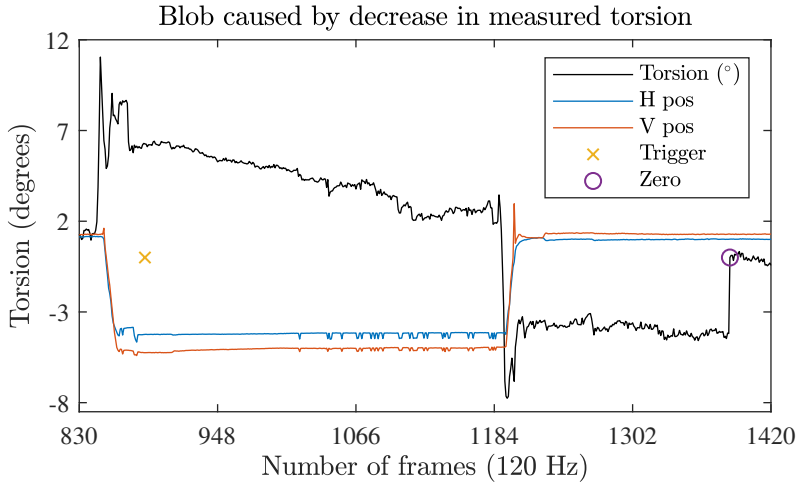

Figure 11: Measured torsion displaying a decrease in value, that does not correspond to the position measurement, followed by an overshoot with opposite sign. Horizontal and vertical position only plotted for reference, scale is not accurate.

Figure 11 shows how the amount of torsion has been set to zero after each zero trigger in the dataset, thus the moment the participant presses the trigger while looking at the reference position. Sometimes the determined amount of torsion while steadily looking at a stimuli drifts in time, resulting in an *overshoot* in the opposite direction after the participant looks back at the centre of the sphere. Because of the overshoot and the irregularity of this decrease, it seems to be an artefact of the method rather than a characteristic of torsional rotation. This overshoot is set to zero at the zero point trigger, which is slightly delayed from the actual moment the participant looks back. Because only the amount of torsion at a certain position in space is wanted, the torsion anywhere else is irrelevant for now. Besides overshoots due to a decrease of measured torsion, there were also *undershoots* due to an increase in measured negative torsion (see figure 12). This could indicate that the MSV method has a deviation to clockwise torsional rotation, represented by negative translation in the polar plot.

It is yet to be determined how the drift in the signal of the torsion arises from the MSV method and how this can be prevented. The overshoot might be irrelevant for our measurements, but that might not always be the case. It is also important to know which value of the torsion corresponds to the actual value over a number of 300 frames.

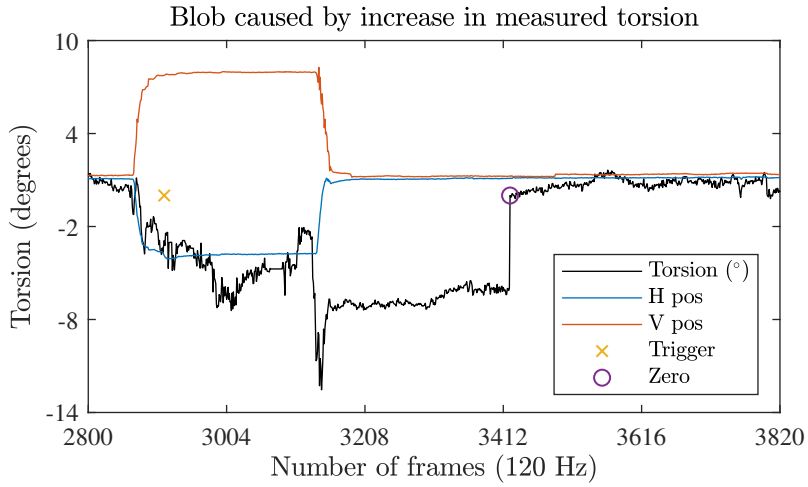

Figure 12: Measured torsion displaying a decrease in value, that does not correspond to the position measurement, followed by an undershoot with equal sign. Horizontal and vertical position only plotted for reference, scale is not accurate.

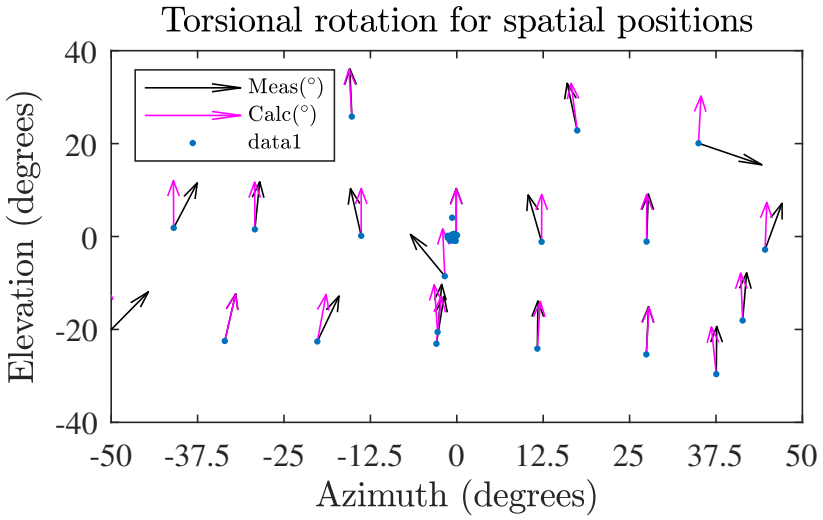

Figure 13: The measured using the MSV method, indicated by the black arrows, is plotted at the different stimulus positions. For each target the false torsion of the Fick gimbal system was also calculated according to Listing's law.

Using the amount of torsion for each stimuli corresponding to a certain position in space, a quiver plot (velocity plot) can be made of the amount of torsion against the gaze direction. As the amount of torsion measured is *pseudotorsion*, it can be determined by integrating the quaternion matrix into the Fick-gimbal

matrix. For all the points in space the calculated torsion is plotted at the corresponding target position. The torsion determined by the MSV method is plotted for reference.

In figure 13 most values of torsion correspond to the predicted values, at least in direction with respect to the centre of the plot. A few anomalies appear in the lower corners, which can be explained by bad ellipse fits. But anomalies on the azimuth or elevation axis could not have been caused by bad ellipse fits, but rather some other problem with either ellipse fit or MSV method.

4.2 Torsional nystagmus

Information from the vestibular system is directly connected to eye movement, as the eye movement can be controlled by passive rotation of the head. Since the vestibular system can only measure relative gravitational differences and accelerations, subjects can be 'fooled' into thinking they are standing still while they are rotating with constant speed. For tilted head movement, torsional rotation comes into play.

Stimulation of the semicircular canals due to acceleration manifests itself in smooth pursuit eye movements, which nearly all people are unable to achieve without either moving visual targets or active stimulation. Tracking stationary targets while in motion often results in alternation of smooth pursuits and saccades, called a nystagmus. This is not to be confused with the medical condition, which is involuntary eye movement. Approximately the first 20 seconds in the moving vestibular chair the subject displays smooth pursuits. Similarly, the subject experiences rotation about the first 20 seconds after the vestibular chair has stopped. Experiments on the VOR can be used to explain balance issues, but also gain insight in the time parameters of neural integration. For now, visual tasks in the vestibular chair involve only horizontal movement. This alone does not lead to any torsional rotation with respect to the primary position, but that changes when the chair is slightly titled. For a tilted chair and thereby a tilted head, the eyes rotate around the line of sight to correct for the rotation of the head. This result in a translation of the value of the torsional rotation for normal eye movement, but the vestibulo-ocular reflex causes the smooth pursuits and saccades to contain a torsional component.

4.2.1 Setup

The Biophysics lab owns a vestibular chair capable of rotating a subject around in all possible rotational axis in 3D. It is isolated from any light source such that participants do not have any reference points while being rotated. Participants can engage in visual or auditory tasks while being rotated in the chair, enabling the experimenter to measure both eye and head movement in response to these rotations. Without any reference points, the participant relies solely on

the vestibular organ for measuring the posture of the head in space. However, this can only measure acceleration and not velocity. The vestibular chair can fool the participants into thinking they are is standing still, while they are still moving at constant speed and even the other way around. When the chair stops moving at constant velocity, the participants will experience that they are rotating at the same speed in the opposite direction. The purpose of the vestibular setup is to study how the brain deals with moving in a moving world.[22] Eye movement is measured with a Pupil Labs eye tracker, calibrated using a single point calibration[20].

4.2.1.1 Torsional nystagmus

In order to measure the torsional nystagmus, an experiment has been conducted where a participant sits 45° tilted in the vestibular chair, after a step change in constant velocity signal of about $30°s^{-1}$. Initially, the participants were expected to display nystagmus eye movements until the fluid inside the vestibular organ balances out the rotation after approximately 20 seconds. After rotating for one minute, the vestibular chair stops and the participant should experience the vestibulo-ocular reflex and again display nystagmus eye movements. The

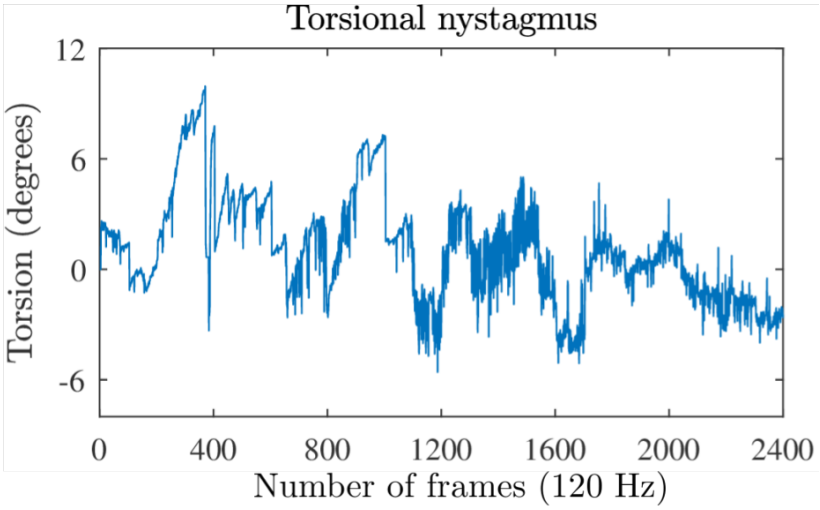

Figure 14: Vestibulo-ocular reflex induced by a 45 degree tilted vestibular chair driven by a step function signal. Torsional nystagmus occurs due to stimulation of the vestibular system after end of step function. Amount of torsion is shown after the step function signal has ended.

measured torsional rotation can be compared with the start and end points of the vestibular chair step function signal. The timeconstant of the neural integration can be determined by fitting the amplitudes with an exponential

function[23]. Since this has been done in previous experiments for the horizontal nystagmus, the time parameter is already determined to be approximately 20 seconds.

4.2.2 Results

Images from the Pupil labs eye tracker were used to determine the amount of torsion over time with respect to the step function signal of the vestibular chair. In figure 14 the amount of torsion is plotted against the time. The amplitude of the torsional nystagmus seems to display an exponential decrease. The amount of torsion is shown in the domain starting from the end point of the vestibular chair step function signal and ends 20 seconds later. The qualitative results of this experiment are not meant as prove for the functionality of the method, but as mere illustration of its potential utility.

5 Summary and conclusion

Three dimensional eye movement has only been measured using video-based techniques for a few decades. The aim of this thesis was to quantify the speed and accuracy of the Maximally Stable Volumes approach and assess whether the application of the method could be used both scientifically and medically. Preprocessing steps for the method, e.g. pupil ellipse fit, histogram equilisation and polar transformation, were also optimised for implementation.

5.1 Summary

Traditional scleral search coils are unpractical, especially in animal research. Additionally, the coils are invasive and can only be used for a limited amount of time. Imaging techniques provide the next best technique for measuring three-dimensional eye movement. Horizontal and vertical eye movement have been measured for many years using edge-detection and shape-based techniques. But torsion, rotation around the line of sight, requires tracking iris features. Manually tracking these features becomes more difficult for low resolution camera images and large geometrical changes in the shape of features due to three-dimensional eye movement.

Before iris features can be tracked, the iris has to be extracted and polar transformed. In the polar transformed image, iris features solely translate instead of rotate and thus change in shape. An important part of the process is the pupil ellipse fit, a method which fits the largest homogenous dark intensity to an ellipse in order to determine the eye direction. This ellipse fit is also used to extract the iris properly. Once the iris is cut out, the angle can be plotted against the radius resulting in a polar plot. Histogram equalisation is applied

in which the intensity of the pupil and glints are not included as these should not count towards the amount of torsion and the maximum area of the tracked features would eject them. This increases the contrast between the iris features, thus optimising the possibility to track them.

J. Ong and T. Haslwanter proposed to use a method called Maximally Stable Volumes (MSV) for tracking these iris features in 2010[6][5]. The MSV method is shown to be robust under geometric changes and imhomogenous lighting conditions by Matas, therefore it is potentially suited for tracking iris features. The method tracks many different features at once, causing the amount of torsion to be very precise. Low resolution camera images can be used, useful for experiments in dark rooms with infrared lighting.

There are several downsides to the MSV method. First of all, it is rather memory extensive and could therefore not be used as a build-in function on the eye tracker. Secondly, the amount of torsion is unreliable for large eye movement, i.e. the corners of the eye. Tracking eye features in corners can be problematic due to the fact that the eyelids cover a significant part of the iris and they decrease the intensity difference between the pupil and the iris. Lastly, the method strongly depends on the parameters which have to be tuned separately for each participant. Subjects with prominent iris feature should a higher sensitivity setting than subjects who have not. The MSV method most likely determines torsion more accurately for lighter eye colours than darker eye colours.

5.2 Conclusion

The MSV method is a promising method, which is robust for determining the amount of torsion during head or small eye movement. However, there are still many problems to overcome if the method will be implemented in scientific research. Since the method relies on frames being connected in time, incorrect measurements such as blinks or bad ellipse fits cause the torsional values to lose their relation with respect to other values. The zero point value has to be reset constantly to prevent deviation. Therefore torsion can only be measured in controlled environments with fixed zero point targets in space or a solution to the drift problem has to be found. To prevent these incorrect measurements the participant should either open his or her eyes wide open without blinking throughout small recordings (typically, one or two seconds) or some adjustments have to be made to the method. At the moment, the method is far from suitable for studying three dimensional eye movement in uncontrolled environments. It would be sufficient for determining torsion during solely head movement in a short period of time, because the full iris would be visible and there would be now problem ellipse fitting the pupil.

The biggest problem with the imaging technique does not lie with the MSV right now, but rather with the pupil ellipse fit method. Several methods have been put to the test, but none were able to track all pupils correctly in the eye corners.

Pupil labs ellipse fits were with certainty far superior to the others, but still a lot of problems with torsional values originated from the fits. In time pupil ellipse fits should be optimised, making it possible to track bright pupils by using techniques other than pixel intensity based ones. Tracking pupils correctly in eye corners with decreased light intensity could probably be improved by using either multiple light sources or multiple cameras. The MSV method itself also needs to be refined such that it does not track too many features in the same area, e.g. eyelashes, and that it tunes the parameters automatically based on the images. If smaller fractions of the iris are used, with higher resolution images the accuracy of the amount of torsion would most likely improve. Experiments with smaller eye movement would be better contestants to test the accuracy and validity of the MSV method, as it would not depend heavily on the pupil ellipse fit and polar transformation.

5.2.1 Discussion

The MSV method is even further away from medical implementation. This requires fast and easy methods, that would not have to be readjusted for every single participant. The method probably works better for light eye colours due to larger intensity differences. If the method were to be applied to the medical sector, it would raise some concerns as the method could not be applied to all ethnicities. Applying histogram equalisation would help to some extend, but the decrease in light intensity of the eye also decreases the amount of prominent iris features. Larger eyes also allow for better pupil ellipse fitting than smaller eyes and the more the iris is visible the more features can be tracked.

A Quaternions

Quaternions are useful for describing 3D rotations of a given angle θ about a single axis $\hat{n}$.

$$q = q_0 + q_1 i + q_2 j + q_3 k \tag{14}$$

$$q^{-1} = \|q\|^{-2}(q_0 - (q_1 i + q_2 j + q_3 k)) \tag{15}$$

$$q = \|q\|\hat{q} \tag{16}$$

We define $\hat{q}$ to be the unit quaternion, which satisfies $\|\hat{q}\| = 1$. Unit quaternions describe only rotation, whereas quaternions for which $\|q\| \neq 1$ represent

stretching in addition to rotation.

$$\vec{I} = (i, j, k) \tag{17}$$

$$\hat{n} = \frac{\vec{I} \cdot \vec{q}}{||\vec{q}||} \tag{18}$$

$$\hat{q} = exp(\frac{\theta}{2}\vec{I} \cdot \hat{n}) \tag{19}$$

$$\hat{q} = cos\frac{\theta}{2} + sin\frac{\theta}{2}\ \vec{I} \cdot \hat{n} \tag{20}$$

The $\vec{I}$ contains 3 unit vectors representing the Cartesian axes x, y, z respectively. The unit vectors are part of the non-abelian quaternion group Q_8 given by the following representation:

$$Q_8 = \langle \bar{e}, i, j, k \mid \bar{e}^2 = e,\ i^2 = j^2 = k^2 = ijk = \bar{e} \rangle$$

Where e is the identity element and $\bar{e}$ commutes with all other elements of the quaternion group Q_8.

The quaternions are basically an extension of complex numbers to 4 dimensions, 3 complex and 1 real. The cyclic properties of the elements of the quaternion group are well suited for describing rotations.

$$pq = p_0 q_0 - \vec{p} \cdot \vec{q} + p_0 \vec{q} \cdot \vec{I} + q_0 \vec{p} \cdot \vec{I} + \vec{p} \times \vec{q} \tag{21}$$

Vector rotation by matrix multiplication $\vec{v'} = R[\hat{n}, \theta]\vec{v}$ can be rewritten in terms of quaternion multiplication.

$$\begin{aligned}
\vec{v'} &= q\vec{v}q^{-1} \\
&= \|q\|^{-2}(q_0 + \vec{I} \cdot \vec{q})\vec{v}(q_0 - \vec{I} \cdot \vec{q}) \\
&= \|q\|^{-2}(\vec{q} \cdot \vec{v} + q_0\vec{v} \cdot \vec{I} + \vec{q} \times \vec{v})(q_0 - \vec{I} \cdot \vec{q}) \\
&= \|q\|^{-2}(-q_0\vec{q} \cdot \vec{v} - q_0\vec{v} \cdot \vec{q} - \vec{q}(\vec{q} \times \vec{v}) + \vec{q}(\vec{q} \cdot \vec{v}) \\
&\quad + q_0^2\vec{v} + q_0\vec{q} \times \vec{v} - (q_0\vec{v} + \vec{q} \times \vec{v}) \times \vec{q}) \\
&= \|q\|^{-2}(\vec{q}(\vec{q} \cdot \vec{v}) + q_0^2\vec{v} + 2q_0(\vec{q} \times \vec{v}) - \vec{q} \times \vec{v} \times \vec{q})
\end{aligned}$$

In order to rewrite quaternion multiplication as a rotation matrix, $\vec{v}$ needs to be separated from the equation above. A cross product matrix A_x allows a cross product $a \times b$ to be written as a matrix multiplication $A_x b$.

$$a \times b = A_x b \tag{22}$$

$$A_x = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix}$$

And its square A_x^2 needed for the term $\vec{q} \times \vec{v} \times \vec{q}$:

$$A_x^2 = \begin{bmatrix} -a_2^2 - a_3^2 & a_2a_1 & a_3a_1 \\ a_1a_2 & -a_1^2 - a_3^2 & a_3a_2 \\ a_1a_3 & a_2a_3 & -a_1^2 - a_2^2 \end{bmatrix}$$

The term $\vec{q}(\vec{q}\cdot\vec{v})$ can be rewritten into an outer product between the $\vec{q}$'s multiplied by $\vec{v}$. An outer product $xx^\intercal$ results in a matrix.

$$xx^\intercal = \begin{bmatrix} x_1^2 & x_2x_1 & x_3x_1 \\ x_1x_2 & x_2^2 & x_3x_2 \\ x_1x_3 & x_2x_3 & x_3^2 \end{bmatrix}, \ \forall i = 1,2,3$$

Now $\vec{v}$ can be separated from the equation.

$$\vec{v'} = q\vec{v}q^{-1} \tag{23}$$
$$= \|q\|^{-2}(q\vec{q}^T + \vec{q}_0^2\hat{\mathbb{1}} + 2q_0q_x + q_x^2)\vec{v} \tag{24}$$
$$= R[\hat{n},\theta]\vec{v} \tag{25}$$

The rotation matrix can now be represented in terms of quaternion vector components. $\|q\|^{-2}$ is taken to be 1 for simplicity.

$$R[\hat{n},\theta] = \begin{bmatrix} q_0^2+q_1^2-q_2^2-q_3^2 & 2(q_2q_1-q_0q_3) & 2(q_3q_1+q_0q_2) \\ 2(q_1q_2+q_0q_3) & q_0^2-q_1^2+q_2^2-q_3^2 & 2(q_3q_2-q_0q_1) \\ 2(q_1q_3-q_0q_2) & 2(q_2q_3+q_0q_1) & q_0^2-q_1^2-q_2^2+q_3^2 \end{bmatrix} \tag{26}$$
$$= \begin{bmatrix} 1-2(q_2^2+q_3^2) & 2(q_2q_1-q_0q_3) & 2(q_3q_1+q_0q_2) \\ 2(q_1q_2+q_0q_3) & 1-2(q_1^2+q_3^2) & 2(q_3q_2-q_0q_1) \\ 2(q_1q_3-q_0q_2) & 2(q_2q_3+q_0q_1) & 1-2(q_1^2+q_2^2) \end{bmatrix} \tag{27}$$

B Maximally Stable Volumes

Maximally Stable Volumes (MSV) is an extension on Maximally Stable Extremal Regions (MSER) in 3 dimensions. This method is used for detecting certain regions in images. It was first proposed by Matas et al[18], for establishing correspondence between different images. This technique searches for a *region* of connected pixels for which all of the pixels have a higher intensity than on the *boundary*, the pixels adjacent to the ones in the region. These regions are called *extremal regions*. For such a region to be stable, it has to satisfy a stability criterion fixed by the rate of change in size for a certain threshold intensity. The formal definition of Matas is given by:

Image I is a mapping $I : D \subset \mathbb{Z}^2 \to S$. Extremal regions are well defined on images if:

1. S is totally ordered (total, antisymmetric and transitive binary relations $\leq$ exist).
2. An adjacency relation $A \subset D \times D$ is defined.

Region Q is a contiguous subset of D. (For each $p, q \in Q$ there is a sequence $p, a_1, a_2, .., a_n, q$ and $pAa_1, a_iAa_{i+1}, a_nAq$.)

(Outer) Region Boundary $\partial Q = \{q \in D \setminus Q : \exists p \in Q : qAp\}$, which means the boundary ∂Q of Q is the set of pixels adjacent to at least one pixel of Q but not belonging to Q.

Extremal Region $Q \subset D$ is a region such that either for all $p \in Q, q \in \partial Q : I(p) > I(q)$ (maximum intensity region) or for all $p \in Q, q \in \partial Q : I(p) \leq I(q)$ (minimum intensity region).

Maximally Stable Extremal Region Let $Q_1, .., Q_{i-1}, Q_i, ...$ be a sequence of nested extremal regions ($Q_i \subset Q_{i+1}$). Extremal region Q_{i*} is maximally stable if and only if $q(i) = |Q_{i+\Delta} \setminus Q_{i-\Delta}|/|Q_i|$ has a local minimum at $i*$. (Here $|\cdot|$ denotes cardinality.) $\Delta \in S$ is a parameter of the method.

We can translate the stability criterion $q(i)$ mentioned above for the number of pixels N_i inside an extremal region.

$$q(i) = |N_{i+\Delta} - N_{i-\Delta}|/N_i \tag{28}$$

MSER can be extended into MSV by means of replacing *regions* by *volumes*, a group of connected pixels in a 3 dimensional space. The stability criterion still holds for MSV. The number of pixels N_i within an extremal volume is product of the number of pixels of the extremal regions connected in time.

.

6 Acknowledgement

This research was supported in part by the Donders Centre for Neuroscience, Radboud University. I would like to thank my supervisor, Professor John van Opstal for his advice and counsel. Furthermore I would like to thank my daily assistant dr. Annemiek Barsingerhorn for her guidance and support throughout this project and technical assistant Jesse Heckman for his work on the setup and help with the experiments.

References

[1] Fransiscus C. Donders. Elfde jaarlijksch verslag betrekkelijk de verpleging en het onderwijs in het nederlandsch gasthuis voor ooglijders [the 11th yearly report of the netherlands hospital for necessitous eye patients]. pages 1–195, 1870. (in Dutch).

[2] David A Robinson. A method of measuring eye movemnent using a scieral search coil in a magnetic field. *IEEE Transactions on bio-medical electronics*, 10(4):137–145, 1963.

[3] H Collewijn, F Van der Mark, and TC Jansen. Precise recording of human eye movements. *Vision research*, 1975.

[4] Josef N van der Geest and Maarten A Frens. Recording eye movements with video-oculography and scleral search coils: a direct comparison of two methods. *Journal of neuroscience methods*, 114(2):185–195, 2002.

[5] M. Donoser and H. Bischof. 3d segmentation by maximally stable volumes (msvs). In *18th International Conference on Pattern Recognition (ICPR'06)*, volume 1, pages 63–66, Aug 2006.

[6] James KY Ong and Thomas Haslwanter. Measuring torsional eye movements by tracking stable iris features. *Journal of neuroscience methods*, 192(2):261–267, 2010.

[7] John Van Opstal. *The auditory system and human sound-localization behavior*. Academic Press, 2016.

[8] J. Opstal. Representation of eye position in three dimensions. In A. Berthoz, editor, *Multisensory control of movement*, Oxford science publications. Oxford University Press, 1993.

[9] MoR Dix and CS Hallpike. Lxxviii the pathology, symptomatology and diagnosis of certain common disorders of the vestibular system. *Annals of Otology, Rhinology & Laryngology*, 61(4):987–1016, 1952.

[10] Th Raphan, V Matsuo, and B Cohen. Velocity storage in the vestibulo-ocular reflex arc (vor). *Experimental Brain Research*, 35(2):229–248, 1979.

[11] Tony Pansell, Hermann D. Schworm, and Jan Ygge. Torsional and vertical eye movements during head tilt dynamic characteristics. *Investigative Ophthalmology & Visual Science*, 44(7):2986, 2003.

[12] Wikimedia Commons. Parallel transport, 2014.

[13] Sijbrand De Jong. Tensors and applications in physics (version 3.1), 2018.

[14] Thomas Haslwanter. Mathematics of three-dimensional eye rotations. *Vision research*, 35(12):1727–1740, 1995.

[15] A John Van Opstal. 200 years franciscus cornelis donders, 2018.

[16] Charles G Moore. To view an ellipse in perspective. *The College Mathematics Journal*, 20(2):134–136, 1989.

[17] Dan Witzner Hansen and Qiang Ji. In the eye of the beholder: A survey of models for eyes and gaze. *IEEE transactions on pattern analysis and machine intelligence*, 32(3):478–500, 2010.

[18] Jiri Matas, Ondrej Chum, Martin Urban, and Tomás Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and vision computing*, 22(10):761–767, 2004.

[19] Andrea Vedaldi and Brian Fulkerson. Vlfeat: An open and portable library of computer vision algorithms. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 1469–1472. ACM, 2010.

[20] Moritz Kassner, William Patera, and Andreas Bulling. Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Adjunct Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '14 Adjunct, pages 1151–1160, New York, NY, USA, 2014. ACM.

[21] Ken Nakayama. Photographic determination of the rotational state of the eye using matrices. *Optometry and Vision Science*, 51(10):736–742, 1974.

[22] Cognition Donders Institute for Brain and Behaviour. The machine: New not-so-merry-go-round for science.

[23] Richard V Abadi. Mechanisms underlying nystagmus, 2002.