Combining Eye- and Head-Tracking Signals for Improved Event Detection

Andrea Schwaller

September 2014

LUND UNIVERSITY

Master's Thesis

Faculty of Engineering, LTH
Department of Biomedical Engineering

Advisor: Martin Stridh
Co-Advisor: Linnéa Larsson

Acknowledgements

I would like to express my great appreciation to my two supervisors, Assoc. Prof. Dr. Martin Stridh and Lic. Eng. Linnéa Larsson, for their valuable and constructive suggestions during the planning and development of this master's thesis.

My grateful thanks are also extended to Prof. Dr. Hans-Andrea Loeliger for making this master's thesis possible across half of Europe.

Further, I would like to thank everyone in the eye-tracking group at the Humanities Laboratory, especially Kenneth Holmqvist, for the possibilities of recording the eye-tracking data and the opportunity of participating at the LETA workshop.

Finally, I wish to thank my family and my friends for their continuous support and encouragement throughout this project.

Abstract

Analysing eye movements recorded with mobile eye-tracking devices is difficult since the eye-tracking signals are severely affected by simultaneous head and body movements. The automatic analysis methods developed for static eye-tracking systems do not take this into account and are therefore not suitable for application to data which also contains head and body movements. As a result, data recorded using mobile eye trackers are often analysed manually.

The goal of the present master's thesis is to develop a method that can robustly detect the three most common types of eye movements from an eye-tracking signal recorded with mobile eye-tracking glasses. Furthermore, the method should compensate for head movements, which are simultaneously recorded using an inertial measurement unit.

A model for eye-in-space motion estimation is proposed which combines eye-tracking signals and head-tracking signals. In addition, a new enhanced event-detection algorithm for the classification of saccades, fixations, and smooth-pursuit movements is developed. In order to test the method, a pilot study is conducted. Moreover, the classification performance of the algorithm is evaluated by comparing the detected events to manual annotations and to the detected events of two existing algorithms.

The results show that by compensating for head movements, the proposed algorithm is able to accurately perform ternary classification of eye movements based on mobile eye-tracking data. With sensitivities and specificities of over 95 % for both a developmental and validation database, the proposed algorithm exhibits a considerably better detection performance than the two existing algorithms used for comparison.

Contents

Chapter 1

Introduction

Eye tracking is a technique that enables to estimate where a person is looking. It is a well-established research tool which can be used to investigate different types of eye movements and their relationship to the underlying processes in the brain. Measurements of eye movements are important for basic research in visual attention, perception and cognition, in psychology and linguistics, but also in applied fields such as product design.

The development of lighter, cheaper, and smaller electronics has miniaturised eye-tracking equipment, transforming it from a large box only available in the laboratory to a pair of glasses. This makes it possible to perform eye tracking in everyday environments such as driving a car or shopping in a supermarket, as well as virtual reality environments. Although mobile eye-tracking glasses allow for greater freedom and more potential applications, they also present a variety of challenges, mainly related to the fact that nothing is static anymore. Subjects are able to freely move their head and body and interact in a natural way with a changing environment. All of these factors cause the eye-tracking signal to not only include eye movements but also head movements, which in turns influences the classification of different types of eye movements in the eye-tracking signal. In order to draw the correct conclusion about the underlying processes in the brain, therefore, it is important to compensate for head motion. Since the tools for analysing eye-movement signals are mainly developed for data recorded with static eye trackers, a new set of algorithms and methods is needed, specifically geared towards mobile eye-tracking data.

To date, there is no commercial event-detection algorithm that is able to perform ternary classification of eye movements when head movement is present. Researchers are forced to perform tedious manual encoding to enable analysis of the recorded signals.

The objective of the present thesis is to develop a method that can robustly detect the three most common types of eye movements from an eye-tracking signal recorded using eye-tracking glasses, while compensating for head movements which are simultaneously recorded using an inertial measurement unit (IMU).

In order to achieve this goal, the project is divided into two main parts: *Gaze Estimation* and *Event Detection.* The general approach is visualised in Figure 1.1 and briefly described thereafter.

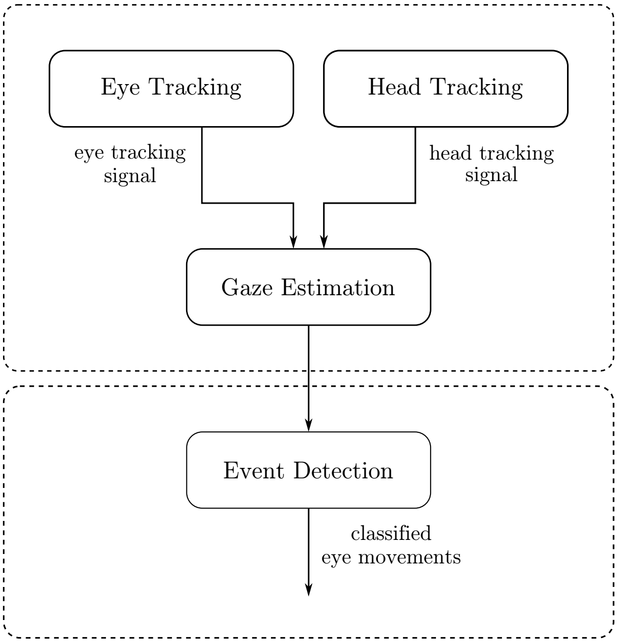


Figure 1.1: Visualisation of the general approach of the thesis.

During *Gaze Estimation*, the head-tracking signal and eye-tracking signal are combined in order to generate a new signal that is as free of head movement as possible. The goal of the subsequent *Event Detection* is to develop and implement a new enhanced event-detection algorithm for the detection of the three most common types of eye movements: saccades, fixations and smooth pursuits.

The thesis is outlined as follows: A description of the eye-tracking field including the relevant standards and technologies applicable to this thesis are summarised in Chapter 2. The gaze-estimation method, including aspects of its implementation, and the proposed event-detection algorithm are presented in Chapter 3. The results are detailed in Chapter 4 and, finally, the results as well as suggestions for future work are discussed and summarised in Chapters 5 and 6, respectively.

Chapter 2

Background

This chapter contains an introduction to the field of eye tracking. Sections 2.1 and 2.2 provide descriptions of the anatomy and physiology of the eye. Section 2.3 contains an overview of the principles of eye- and head-tracking systems. Finally, Section 2.4 outlines the current status of event-detection algorithms.

2.1 Anatomy of the Eye

The eye is the basic organ of sight and is often referred to as one of the most complex parts in the human body. As part of the visual system, it contributes to the processing of visual information. The eyeball is a spherical structure located in a protective framework of bones and connective tissue. It is composed of three layers (the fibrous tunic, the vascular tunic and the retina) and divided into two cavities (the anterior cavity and the vitreous chamber) [1]. This structure is shown in Figure 2.1. The outermost layer, the fibrous tunic, consists of the sclera and the cornea. The sclera is the white part of the eye and gives the eyeball its shape and structural stability. Furthermore, it protects the inner, more sensitive parts of the eye. The cornea is a thin, transparent protective structure at the front of the eye. It covers the iris, pupil, and anterior chamber and permits light rays to enter the eye. The second or middle layer, the vascular tunic, is composed of the choroid, the ciliary body and the iris. The choroid is located at the back of the eyeball and consists of a network of blood vessels that nourish the retina. At the front of the eyeball, the choroid is specialised into the ciliary body and the iris. The ciliary body contains the muscles that determine the shape

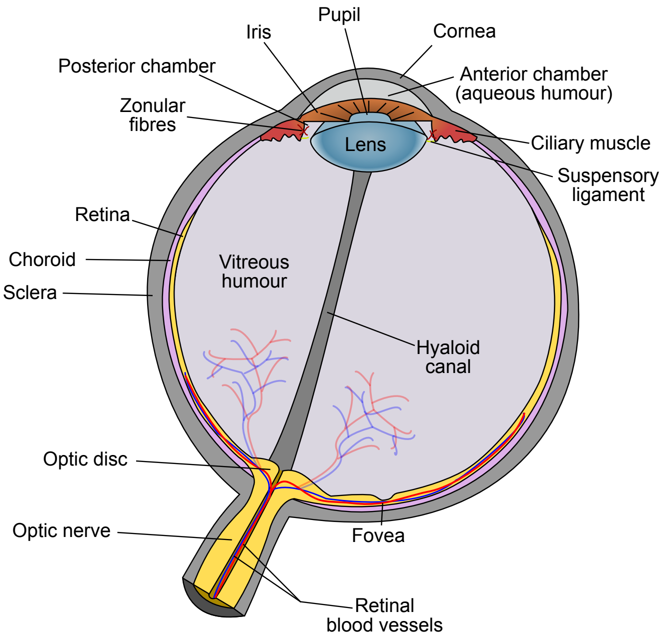


Figure 2.1: Structure of the human eye, from [2].

of the lens, whereas the iris is the coloured part of the eye. The iris regulates the amount of light that enters the eye by adjusting the size of the pupil, the black hole in the centre of the iris. The lens is located behind the iris and the pupil. It focuses light on the retina, which is the innermost layer [3]. The retina is a light-sensitive membrane, which is responsible for converting visual signals into nerve impulses and subsequently transmitting them via the optic nerve to the brain. It contains about 200 million photoreceptive cells which are functionally classified into two different types, rods and cones. As rods are sensitive to dim and achromatic light, they are important for night vision. Conversely, cones respond to bright and chromatic light and are thus responsible for daylight and colour vision. The fovea is located at the centre of the retina. The fovea has the highest concentration of cones, which means that images focused there are seen with the highest visual acuity or resolution. Towards the periphery of the retina, the concentration of cones decreases whereas the concentration of rods increases [4]. The retinal periphery can detect new objects of interest. In order to be able to clearly see a new object, the eyes need to be redirected towards it so that the image of the new object is focused on the fovea [5].

2.2 Eye Movements

During visual perception, the eyes are constantly moving. The purpose of eye movements is to focus an object of interest on the centre of the fovea and/or keep it there in order to see a clear image of the object. Three antagonistic pairs of muscles are responsible for controlling the movements of the eye: the superior and inferior rectus, the medial and lateral rectus, and the superior and inferior oblique. They are depicted in Figure 2.2. These muscles allow the eye to move vertically, horizontally and torsionally within its orbit and are thus responsible for the three-dimensional orientation of the eye inside the head [5].

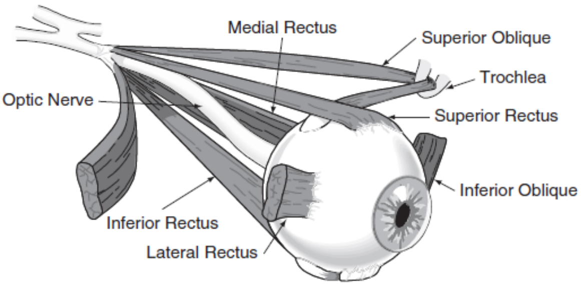


Figure 2.2: Extraocular muscles of the right eye responsible for horizontal, vertical, and torsional eye movements, from [5].

On the basis of their function, eye movements can be divided into two main categories - those that stabilise the gaze and those that shift the gaze. In [6], these two categories are further divided into seven main functional classes. The classes and respective main functions are summarised in Table 2.1. While vestibular, visual fixation, and optokinetic systems belong to the first category, in that they hold images steady on the retina, saccades, nystagmus quick phase, and smooth-pursuit movements belong to the second category, in that they redirect the line of sight to a new object of interest. Vergence movements have gaze-holding as well as gaze-shifting properties.

Out of these seven functional classes, the three most common types of eye movements are: fixations, saccades and smooth pursuits. Examples are shown in Figures 2.3 and 2.4. These three types of eye movements are sufficient to gain insight into the localisation of visual attention [7]. In addition, another type of eye movement will be important in this thesis, namely the vestibulo-

Table 2.1: Functional classes of human eye movements, from [6].

Class of Eye Movement	Main Function
Vestibular	Holds images of the seen world steady on the retina during brief head rotations or translations
Visual Fixation	Holds the image of a stationary object on the fovea by minimising ocular drifts
Optokinetic	Holds images of the seen world steady on the retina during sustained head rotation
Smooth Pursuit	Holds the image of a small moving target on the fovea; or holds the image of a small near target on the retina during linear self-motion; with optokinetic responses, aids gaze stabilisation during sustained head rotation
Nystagmus Quick Phase	Reset the eyes during prolonged rotation and direct gaze towards the oncoming visual scene
Saccades	Bring images of objects of interest onto the fovea
Vergence	Moves the eyes in opposite directions so that images of a single object are placed or held simultaneously on the fovea of each eye

ocular reflex (VOR) which belongs to the class of vestibular eye movements. Its purpose is to stabilise the gaze during head movements. In addition, other types of eye movements such as postsaccadic oscillations (PSO) are also reported and discussed in the literature. PSO are oscillatory movements that may occur at the end of a saccade [8]. They will not be investigated in this thesis, however, because they usually occur at such a high frequency that they would not be visible in the low-speed data which will be recorded.

2.2.1 Fixations

Fixations are the short time periods when the eye remains more or less still (cf. Figure 2.3). For example, this is the case during reading when the eye

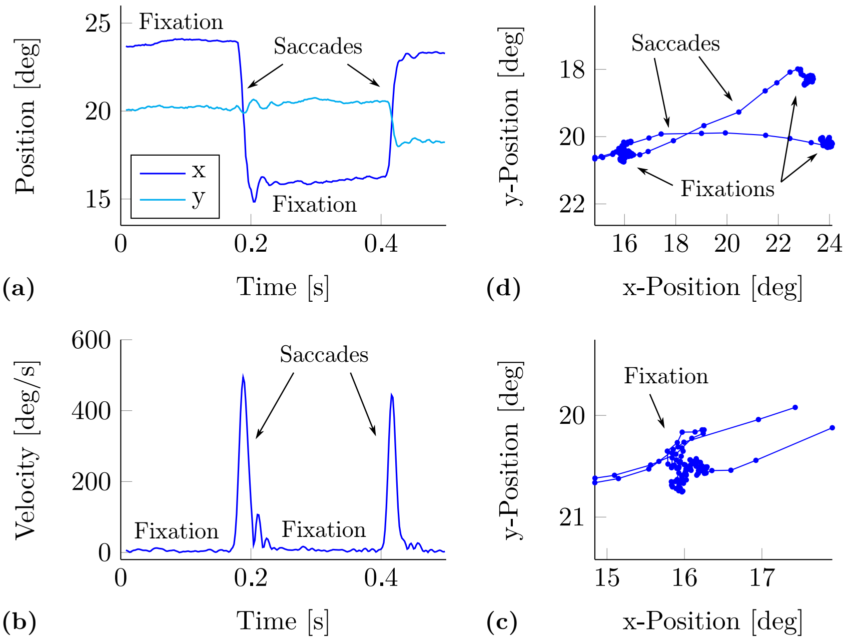


Figure 2.3: Example of the eye movements fixations and saccades. (a) Position over time, (b) velocity over time, (c) position in the spatial domain.

temporarily stops at successive locations across the page. The object of interest is then kept relatively stable upon the fovea while visual information is gathered. In order to have a clear vision of higher spatial frequencies, the image should move less than about 5 °/s and should lie within 0.5 ° of the centre of the fovea [6]. In addition, fixations are characterised by the occurrence of the three involuntary micro-movements: tremor, drift and microsaccades [9]. A tremor is a small and fast wave-like movement of the eyes with a frequency of around 90 Hz and an amplitude of less than 0.01 °. Its exact function is unclear. Drift happens when the image of interest slips on the retina, in that it moves away from the centre of the fovea. Microsaccades are small and fast, jerky eye movements with velocities of around 15 - 50 °/s and amplitudes typically less than one third of a degree. Their mean duration is about 25 ms. The purpose of microsaccades is to return the eyes to the object of interest and thus compensate for the displacements in eye position produced by drifts. The mean duration of an entire fixation is about 200 - 300 ms, depending on the nature of the task [10].

2.2.2 Saccades

Saccades are rapid eye movements which shift the eye from one fixation point to another and thus allow the fovea to fixate different objects of interest within the visual field (cf. Figure 2.3). This type of eye movement can be observed while reading, for instance, when the eyes quickly move to the beginning of the next line once the end of a line is reached [5]. The main characteristics of saccades are: velocity and duration, shape and trajectory, and latency [6]. The relationship between the size, speed and duration of a saccade, often referred to as the main-sequence relationship, is relatively consistent; the larger the saccade is, the higher its peak velocity and the longer its duration. Typically, the velocity lies between 30 and 500 °/s and the duration between 30 and 80 ms [11]. Due to these high velocities, the viewer is blind during most of the saccade, which means that almost no visual information is gathered. Saccades have the shape of a temporal waveform and exhibit a slightly curved trajectory in space, which implies that the eye generally does not take the shortest way (straight line) between the starting and end point [12]. The latency or reaction time of a saccadic eye movement is about 100 - 300 ms [5]. This is the time interval between when a stimulus is first present and when the eye effectively starts to move, which is determined by the time it takes for the brain to program and initiate the saccade.

2.2.3 Smooth Pursuits

Smooth pursuits occur whenever the eye follows a moving object in the visual environment in order to keep it on the fovea (cf. Figure 2.4). This type of eye movement can typically be observed while watching a flying bird or a moving car. The main difference between smooth pursuits and saccades is that smooth-pursuit movements cannot be controlled voluntarily [13]. While saccades can be initiated across a stationary environment, smooth pursuits require an object to follow. Moreover, in an environment consisting only of moving objects, pursuits cannot be completely suppressed. Smooth-pursuit movements can be divided into two phases: an initial acceleration phase and a subsequent correction phase [6]. During the initial acceleration phase, the eye needs to compensate for the latency and catch up with the moving target. The latency is around 100 - 200 ms, which corresponds to the time interval between when an object first moves and when the pursuit eye movement is subsequently initiated[11]. In the subsequent correction phase, the smooth-pursuit system acts as a negative-feedback control system [6]. This means that it tries to match the velocity of the smooth-pursuit movement with the

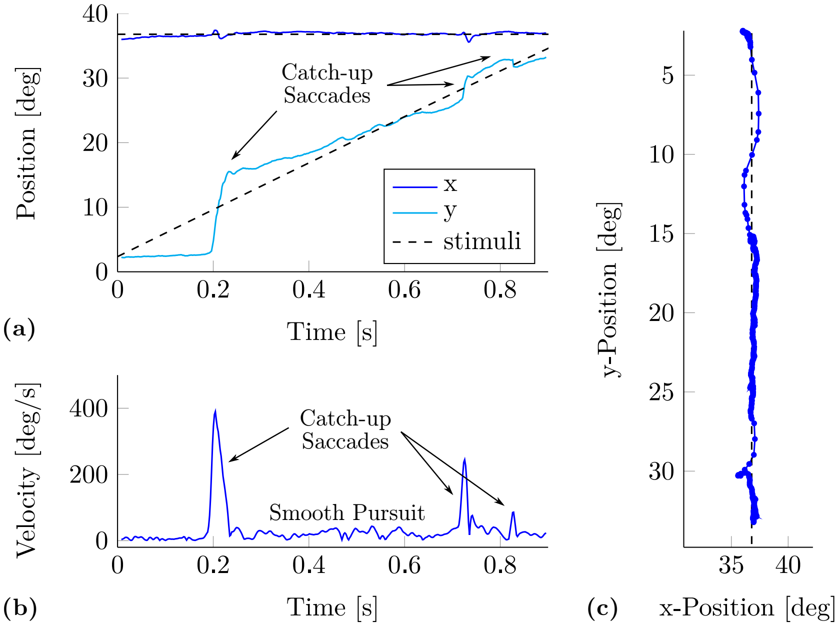


Figure 2.4: Example of smooth-pursuit movements. (a) Position over time, (b) velocity over time, (c) position in the spatial domain.

velocity of the moving target by calculating the retinal error velocity, which is the difference between the velocity of the eye and that of the target. If this error becomes too large, an eye movement that will catch up with the target, known as a catch-up saccade, is generated (cf. Figure 2.4). The amount of catch-up saccades needed depends on the speed and predictability of the moving target. The velocity of smooth pursuits is typically between 10 and 30 °/s [11] but can reach values up to 100 °/s [14].

2.2.4 Vestibulo-ocular Reflex (VOR)

The function of the vestibulo-ocular reflex (VOR) is to stabilise the gaze during head movements such that the viewed object remains on the fovea of the retina. In response, compensatory eye movements are generated in the opposite direction to head movements [6]. An example of this type of eye movement would be when you fixate a point in front of you and start turning your head to the left while fixated on the point. In order to be able to continue to fixate the point, the eyes need to rotate to the right. Head motion consists of three rotational and three translational components. The rotational components are: horizontal, vertical and torsional. Thus, the

VOR also responds with horizontal, vertical and torsional eye movements. The three translational head movement components are lateral (side-to-side), vertical (up-down), and longitudinal (front-back). The VOR responds to these movements by producing horizontal, vertical and vergence eye movements. Head velocities are generally below 100 °/s but the VOR is able to stabilise head velocities of up to 350 °/s [15]. Depending on the viewing distance of the object of interest, the gain of VOR must be adjusted (cf. Section 3.3). The latency of the VOR is in the range of 7 to 15 ms [16], which is extremely short. This corresponds to the time interval between when the head begins to rotate or translate, to when the compensatory eye movement is initiated. No other sensory mechanism generates compensatory eye movements so quickly. Visually-mediated eye movements, for instance, have latencies of at least 70 ms [17].

2.3 Eye- and Head-Tracking Systems

2.3.1 Eye-Tracking Techniques

Eye trackers are measurement devices used to record eye movements. They allow estimation of where a person is looking, providing an insight into where visual attention is localised. There are a number of different eye-tracking methodologies presented in the literature [18, 19, 7, 11, 20, 21]. The simplest and oldest monitoring technique is the direct observation of a person's eye. Obviously, this technique is very subjective and only allows the identification of large eye movements. Thus, it was logical for researchers to seek more sophisticated and particularly, more objective, eye-movement measurement techniques. Starting from the late 1800s, a range of different devices and techniques were developed, the most common of which are discussed in the following paragraphs.

ElectroOculoGraphy (EOG)

EOG is an eye-tracking technique based on the electric potential differences of the skin around the eye. In order to measure eye movements, electrode pairs are placed around the eye. As the eye rotates, the orientation of the corneoretinal electrostatic dipole changes with it. This change will be visible in the measured EOG signal. The EOG was the most widely applied eye-movement measurement technique during the mid-1970s and is still used. An advantage, besides the low cost, is that eye movements can be recorded even if the eyes are closed, for example, during sleep. Accuracy and precision, however, are rather low [19].

Scleral Contact Lens

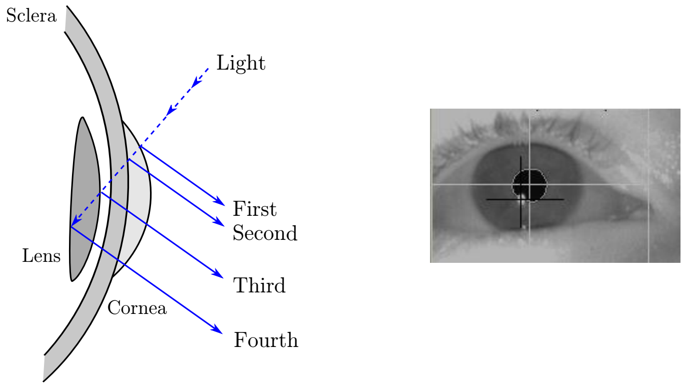
This eye-tracking technique makes use of a large contact lens, which can be worn directly on the eye. On top of the lens, a mechanical or optical device is attached. The principle method, known as the scleral search coil technique, uses small metal wire coils. Rotations of the eyes can then be recorded with the aid of a surrounding electromagnetic field. When the eyes move, the potential difference in the coil varies and can be measured. Although rather uncomfortable, this eye-movement measurement technique is one of the most precise and accurate options [7].

VideoOculoGraphy (VOG)

VOG systems represent a wide variety of video-based eye trackers and are currently the most widely-applied systems for the recording of eye movements. They obtain image data from one or more cameras. Afterwards, the image data is additionally processed in order to estimate where the user of the system is looking. During the first step, the eye is detected and localised within the image. The position of the eye is generally measured using the pupil or iris centre. During the second step, the location of the eye is tracked over subsequent image frames to estimate the path of the gaze. As in [20] and [21], the present study adopts the terms "eye detection" and "gaze tracking" to differentiate between these two steps.

Pupil and Corneal-Reflection VOG

Video-based pupil and corneal-reflection eye tracking has been the dominant method since the early 1990s [11]. The method employs one or more cameras and a single or multiple infrared light sources, which are placed in front of the viewer. Typically, these light sources are close to the stimulus screen and directed towards the eye. The purpose of the infrared light sources is to create reflections on the eye, or more precisely, on the boundary between the lens and the cornea. A total of four reflections, known as Purkinje images, may occur on the external and the internal surface of the cornea as well as on the external and the internal surface of the lens. An illustration is shown in Figure 2.5. The first Purkinje image is called the corneal reflection. This reflection together with the pupil is tracked by the camera system.



(a) **(b)**

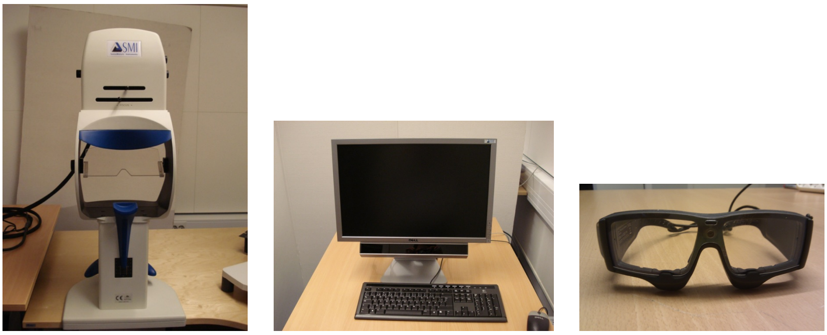
Figure 2.5: (a) The four Purkinje images which occur when light is reflected on the eye. (b) Example of an eye image captured by a video-based eye-tracking system. The detected geometric centres of the pupil and the corneal reflection are marked with a white and a black cross, respectively.

The goal of the eye-detection step is to robustly detect the pupil and the corneal reflection. Eye detection is typically done using either feature-based or model-based approaches or combinations thereof [11, 20, 21]. The feature-based approach involves exploring the characteristics of the eye and extracting distinctive local features. Commonly-used features include the pupil, the limbus, and corneal reflections. Detection criteria often include gradient calculations, in order to find edges or contours, or thresholding, a process which groups pixels according to their intensity distribution. The model-based approach uses a model of the eye, which is matched to the eye image using a similarity measure. Both methods have their advantages and limitations. The feature-based approaches are time efficient and rather robust as long as the image data is of good quality. The major disadvantage of these approaches is that they perform poorly if the images are disturbed, causing parts of the features under investigation to be covered. This can be due to a drooping eyelid or downward-pointing eye lashes. Model-based approaches, by contrast, provide more accurate and robust estimates in such situations. They suffer, however, from a high computational complexity. To overcome the respective shortcomings of both approaches and exploit their benefits, hybrid methods which combine both techniques within a single system can be adopted [20].

After a successful detection phase, the geometric centres of the pupil and corneal reflection are determined (cf. Figure 2.5). They are used in the subsequent gaze-tracking step. During eye movements and small head movements, the relative distance between the pupil and corneal reflection changes systematically. Whereas the pupil rotates together with the eye, the corneal reflection remains relatively stable at its initial position. This means that the corneal reflection can be used as a reference point in the image. Moreover, the vector between the centre of the pupil and the centre of the corneal reflection can be used to determine different gaze positions. In order to establish a relationship between the vector and a position on the stimulus screen, an initial calibration is required. The calibration typically involves presenting between 5 and 13 points on the stimulus space, each of which needs to be fixated by the user.

2.3.2 Types of VOG Systems

Depending on the application, different types of eye tracker are preferable. In [11], VOG eye-tracking systems are classified into three main types: tower-mounted, remote and head-mounted eye trackers. Besides variations in their setup, i.e., the way in which cameras and illuminations are combined, they chiefly differ with regard to the type of data they produce. Representative examples of the different types are shown in Figure 2.6.



(a) **(b)** **(c)**

Figure 2.6: Representative examples of the different types of VOG eye-tracking systems: a tower-mounted system (a), a remote system mounted below a screen (b), and eye-tracking glasses (c).

Tower-Mounted Systems

As shown in Figure 2.6, a tower-mounted system comprises a lower part, where the head can be placed and an upper part, where camera(s) and illumination(s) are located. The lower part typically consists of a forehead and chin rest that are used to restrain head movements. This type of system allows high-quality data recording with typical sampling frequencies of around 1000 - 2000 Hz. The stimuli are generally presented on a monitor.

Remote Systems

In the case of remote eye trackers, the camera(s) and illumination(s) are located in front of the viewer next to the monitor where the stimuli are presented. As a result, this system provides more flexibility, but because small head movements are possible, this comes at the cost of lower quality data. The sampling frequencies of remote systems are usually around 60 - 500 Hz, which is lower than that of tower-mounted systems.

Head-Mounted Systems

While tower-mounted and remote eye trackers are static systems used for experiments inside a laboratory, head-mounted eye trackers provide much more flexibility and mobility, enabling the recording of real-life activities outside the laboratory. Both eye camera(s) and illumination(s) are located on the head of the user and may be mounted on either a helmet, cap or pair of glasses. In addition, a scene camera is attached to the eye tracker. Its purpose is to record the stimuli. Again, the benefits of more mobility comes at the cost of lower sampling frequencies, which in this case are typically around 30 - 60 Hz. Due to the low sampling rate, some types of rapid eye movements such as postsaccadic oscillations will not be visible in the resulting eye-tracking data.

2.3.3 Head-Tracking Techniques

Head trackers or head-tracking systems are used to record head movements in order to calculate the position of the head in space. The applications of head-tracking systems include teleconferencing, virtual reality, and assistive technologies, whereby a wheelchair, keyboard or mouse can be controlled using head movements [21]. However, head trackers are often used in combination with head-mounted eye trackers in order to simplify data analysis. In the literature, several approaches for estimating the pose of the head are proposed which can be used in combination with eye-movement detection. Head trackers can be grouped based on their tracking principle into magnetic, optical, vision-based, acoustic, and inertial head-tracking systems.

Magnetic Head-Tracking Systems

Magnetic tracking systems consist of a source, the transmitter, and a motion-tracking sensor, the receiver. The purpose of the source is to generate a near-field, low-frequency electromagnetic dipole field. The field vectors which are emitted are subsequently detected by the sensor, which is typically embedded or attached to the object which is being tracked. The signals which are sensed enable the receiver's position and orientation relative to the transmitter to be determined. In [22], a magnetic head-tracking system is used in combination with a head-mounted eye tracker. The eye tracker measures the eye-in-head motion, whereas the head tracker, mounted on the eye tracker, measures the head-in-space motion. The setup is used to evaluate the vestibulo-ocular reflex (VOR), for which it is crucial to measure both eye and head movements. The major disadvantages of magnetic tracking systems are that distortions can be caused by metal objects and that the accuracy and resolution rapidly decrease with distance [23].

Optical Head-Tracking Systems

Similarly to magnetic tracking systems, optical tracking systems also consist of a source and a motion-tracking sensor, but apply laser-range measurement techniques instead of using an electromagnetic field. The source continuously scans the work space with laser beams, which are sensed by the sensor that is again attached to the object which is being tracked. The signals which are sensed are further processed into position and orientation data. A system which automatically analysis a driver's visual perception of traffic hazards based on eye-movement detection is presented in [24]. Since the experiment was conducted in a driving simulator in which the base moved, it was important to measure head movements as well as eye movements. The head movements were recorded using an optical LaserBird head tracker. Optical systems are highly accurate but suffer from line of sight problems if the path between the source and the sensor is obstructed [23].

Vision-Based Head-Tracking Systems

Due to increasing computing performance, numerous vision-based tracking approaches have been developed. The main idea behind vision-based approaches is to estimate the orientation and position of the head using images taken by a camera. This can either be done in an "inside-out" style or an "outside-in" style [25]. Inside-out-style approaches use a head-mounted scene camera. The orientation and position of the head are then estimated by processing the images from the resulting video. Outside-in-style approaches, by contrast, use a bird's-eye-view camera, which takes images of the head from a fixed viewpoint. Again, the orientation and position of the head are then estimated by processing the images from the resulting video.

Image processing of the video is often carried out using feature points. The orientation and position of the head are estimated by tracking the movement of a set of points. These points correspond to the key features and are traced across adjacent views in subsequent video frames. This can be achieved, for example, by using two-view geometry to relate image points in two separate views of a scene captured from the same camera from different viewpoints [21]. Feature points can either be markers or natural features. Markers can be placed at known positions around the target scene environment (inside-out-style) or can be attached to the head or the head-mounted eye tracker (outside-in-style). The main disadvantage of using markers is that this method can only be applied within a restricted physical space. Therefore, natural features such as edges, lines, or corners are most often tracked.

Two applications of vision-based systems used in combination with eye-movement detection are presented in [26] and [27]. In [26], an omnidirectional vision sensor mounted on top of a regular eye tracker is used to capture a circular image of the environment. Head rotations (but not translations) are estimated based on analysis of the image sequence by either tracking key-feature points, estimating the optical flow, or by using spherical-harmonic decomposition of the images, the latter of which yielded the best results. In [27], the motion inherent in a scene video of a head-mounted eye tracker is analysed. Using global optical-flow calculations, the relative head motion is estimated and compensated for. Since changes in texture are tracked, the algorithm has problems when the scene video contains textures which are either too similar to each other, or too faint to be detected. In general the disadvantage of vision-based systems is that extremely fast head or body movements can cause motion blur and distortion in the scene video such that the algorithms are no longer able to perform accurately.

Acoustic Head-Tracking Systems

A head-orientation estimation method based on acoustic signals is presented in [28], whereby the direction of the head is estimated by localising the source of the user's voice. There are no reports of this method being used in conjunction with eye-movement detection, however.

Inertial Head-Tracking Systems

Another possible way to obtain information on head movements is to make use of an accelerometer, a gyroscope or a combination of the two, known as an inertial measurement unit (IMU). A huge advantage of inertial systems is that they are not restricted in range and can be used to record in natural environments, in contrast to magnetic, optical, and many vision-based systems. In [27, 29], one or multiple accelerometers were used to measure head movements while simultaneously time performing eye tracking. The recordings, however, suffered from problems related to drift compensation as well as difficulties in synchronising head- and eye-tracking data.

2.4 Existing Event-Detection Algorithms

The goal of an event-detection algorithm is to detect and classify eye movements in eye-tracking data, so that the positional signal is segmented into different types of eye movements. The classification task may vary depending on the type of eye-tracking system, the sampling frequency, the number and types of eye movements which need to be detected, and whether the algorithm is supposed to work in real-time or as an offline procedure.

Most algorithms are designed to distinguish between the two most common types of eye movements, fixations and saccades. Based on the features used for the classification, event-detection algorithms are typically grouped into velocity-based algorithms and dispersion-based algorithms [30, 31]. The first group of algorithms works by analysing the velocity component of the movement signal, taking advantage of the fact that fixation samples have low velocities, whereas saccade samples have high velocities. There are three commonly used velocity-based algorithms reported in the literature: Velocity Threshold Identification (I-VT), Hidden Markov Model Identification (I-HMM), and Kalman Filter Identification (I-KF). The I-VT algorithm is the simplest type of velocity-based algorithm and functions by sorting samples based on their point-to-point velocities. Samples with velocities higher than a given threshold are classified as saccades, whereas samples with velocities lower than the threshold are classified as fixations. The I-HMM algorithm is a more sophisticated version of the I-VT. It uses a two-state HMM (fixation and saccade) whereby the states are characterised by the velocity distributions of saccade and fixation samples, respectively. Although the probabilistic representation inherent in the I-HMM algorithm performs more robustly, this comes at the cost of a more complex parameter space. In the case of the I-KF algorithm, the eye is modelled as a system with two states (position and velocity). In order to classify each eye-positional sample as a part of a fixation or a saccade, a Chi-square test is applied on the difference between the measured and predicted eye velocity.

While velocity-based algorithms typically require data recorded at frequencies higher than 200 Hz, dispersion-based algorithms are typically used for signals with sampling frequencies below 200 Hz [11]. Dispersion-based algorithms analyse the positional properties of the signal, taking advantage of the fact that fixation samples are generally less spread than saccade samples. The two most common algorithms are Dispersion Threshold Identification (I-DT) and Minimum Spanning Tree Identification (I-MST). The I-DT algorithm calculates the spatial dispersion of points within a temporal window and compares it to a threshold. If the dispersion of the window is higher than the threshold, the points within the window are classified as fixations. Otherwise, the first sample of the window is classified as a saccade and the window is moved by one sample. Multiple versions of the I-DT algorithm exist, which differ according to how the dispersion is calculated. An overview of different dispersion measures can be found in [32, 33]. The I-MST algorithm builds an MST whereby a tree connects a set of points such that the total Euclidean distance of the tree's line segments is minimised among all spanning trees. Eye positions are then classified into fixations and saccades based on point-to-point distance thresholds.

The disadvantages of these basic algorithms are that most of them rely on static thresholds and are very sensitive to parameter settings. This may result in poor performance if the data is noisy. As a result, a number of more sophisticated algorithms have been developed, ranging from velocity-based algorithms which take the noise level of the position signal into account [34], to approaches adopting adaptive, data-driven thresholds [35], algorithms which use information about both eyes to classify eye movements [36], and approaches that use acceleration signals instead of, or in addition to, the velocity signals [37].

Among these algorithms, only a few enable ternary classification in order to discriminate between saccades, fixations, and smooth-pursuit movements. The major problem is that the signal characteristics of smooth-pursuit movements overlap with the signal characteristics of saccades and fixations, which makes classification much more complicated [38]. In [39], three basic algorithms to detect fixations, saccades, and smooth-pursuit movements are compared and evaluated: Velocity and Velocity Threshold Identification (I-VVT), Velocity and Movement Pattern Identification (I-VMP), and Velocity and Dispersion Threshold Identification (I-VDT). All three algorithms are modified versions of the I-VT algorithm; Firstly, they identify the saccades by applying a velocity threshold and secondly, they separate fixations from smooth-pursuit movements using either a second velocity threshold, a movement-pattern analysis, or a dispersion threshold. The most successful method combined velocity and dispersion thresholds, whereas the I-VVT algorithm showed the poorest results. In [40], an algorithm which detects saccades, fixations, and smooth pursuits using a velocity threshold in combination with Principal Component Analysis is employed to investigate eye movements in humans and monkeys. The detection performance was not evaluated, however. Another method is proposed in [8, 41]. After the approximate saccadic intervals are detected based on the acceleration signal, the exact onsets and offsets of the saccades are determined by applying three specialised criteria based on directional information in the positional signal. Subsequently, the algorithm detects smooth-pursuit movements by calculating the characteristics of the signal at different stages, which represent the different spatial scales of the data. The algorithm detects movements considerably better than the I-VDT algorithm.

While the algorithms described above [39, 40, 8, 41] are developed for eye-tracking signals with higher sampling frequencies, a few other algorithms exist which are developed for low-speed mobile eye-tracking systems. In [26], the I-HMM algorithm is extended to a four-state HMM in order to also capture smooth-pursuit movements and VORs. In addition to the eye velocities, the head-movement velocities are also integrated as a second observation variable. Preliminary results show that the algorithm is able to classify VORs and saccades. Smooth-pursuit movements, however, were sometimes incorrectly detected. Another approach, which uses a set of shape features that capture the shape characteristics of smooth-pursuit movements, is proposed in [42]. A machine-learning approach is adopted whereby different shape features are combined and used to detect smooth-pursuit movements in the presence of other types of eye movements. Although the machine-learning approach performs well in detecting movements, the disadvantage is that it requires a sufficient amount of training data. Finally, in [24], a driver's visual perception of traffic hazard is analysed using an adaptive online algorithm to detect fixations, saccades and smooth pursuits. The classification is carried out by employing an online Bayesian mixture model [43] in combination with Principal Component Analysis. Although the method shows promise, it is still at a preliminary stage.

Chapter 3

Methods

The objective of the present thesis is to develop a method that can robustly detect the three most common types of eye movements from an eye-tracking signal recorded using eye-tracking glasses, while compensating for head movements which are simultaneously recorded using an IMU. In order to achieve this goal, the project is divided into two main parts, described below.

- *Gaze Estimation.* In the first part, the head-tracking signal is combined with the eye-tracking signal in order to generate a new signal that is as free of head movements as possible. This new signal is used in the subsequent event-detection part. Section 3.1 begins with a description of the apparatus used in the thesis. A classification of different combinations of eye, head and body movements into different complexity levels is presented in Section 3.2. In Section 3.3, a method for combining the head- and eye-tracking signals is derived. Section 3.4 concerns signal analysis, investigating different properties of the two types of signals. The real-world implementation of the method and related problems are discussed in Section 3.5. Finally, a pilot study and a proposed evaluation procedure are described in Section 3.6.

- *Event Detection.* The goal of this second part is to develop and implement a new and enhanced event-detection algorithm to detect saccades, fixations, and smooth-pursuit movements in the setting which was previously described. The proposed algorithm is presented in Section 3.7 and evaluated in Section 3.8. The evaluation is performed by comparing the results to those of two alternative algorithms.

A - Gaze Estimation

In the following section, it is important to distinguish between eye-in-head position and eye-in-space position, also called gaze. The eye-in-head position corresponds to the position of the eye relative to the head. Eye-in-head positions are what the eye-tracking glasses provide, a sequence of coordinates based on a head-centric coordinate system. Eye-in-space position, by contrast, corresponds to the position of the eye relative to the coordinate system of the outside world. In the present case, the aim is to produce signals which only contain eye-in-space coordinates so that a detection algorithm can be successfully applied to them. Eye-in-space motion can basically be derived by combining the head-in-space motion with the eye-in-head motion. This process is called gaze estimation and will be further investigated in Section 3.3. An illustration is shown in Figure 3.1.

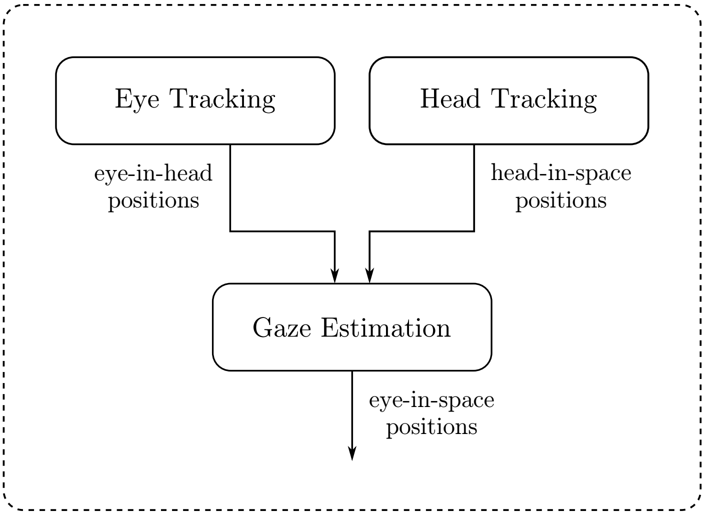


Figure 3.1: Illustration of the gaze-estimation process, which combines eye-in-head and head-in-space motion in order to derive eye-in-space motion.

3.1 Apparatus

The purpose of this section is to shortly introduce the equipment which is used in the thesis to track eye and head movements. Eye movements are measured with mobile eye-tracking glasses, while head motion is recorded using an inertial measurement unit (IMU).

3.1.1 Eye-Tracking Glasses

The eye-tracking signals are recorded using the eye-tracking glasses 2.0 from SensoMotoric Instruments (SMI). It is a non-invasive video-based glasses-type eye tracker with a scene camera and two eye cameras. A picture of the glasses with the corresponding camera locations is shown in Figure 3.2.

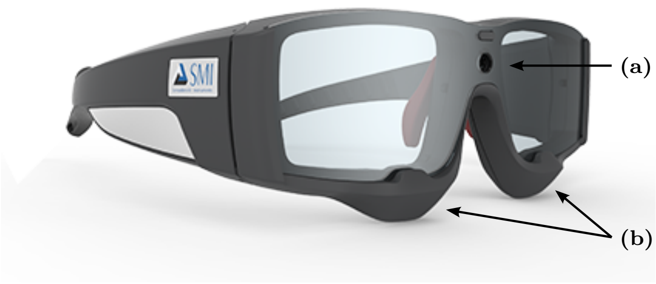


Figure 3.2: SMI eye-tracking glasses. Letters indicate the locations of a) the scene camera and b) the eye cameras, from [44].

The temporal resolution of the eye tracker can be adjusted between 30 Hz and 60 Hz. Few mobile eye trackers are able to record eye movements at such high sampling rates. A sampling frequency of 60 Hz, however, is still very low compared to other types of eye trackers. As a result, the eye-tracking data will not include rapid eye-movement types such as postsaccadic oscillation. Eye positions are determined based on pupil and corneal reflection tracking (cf. Section 2.3) and a total of six corneal reflections are tracked. Furthermore, the eye-tracking glasses record binocularly, which means that data from both eyes is used. The main technical specifications are summarised in Table 3.1.

Eye-Tracking Data

The mobile eye-tracking system outputs a data stream containing x- and y-axis positions in the coordinate system of the scene camera video frame, which has a resolution of 1280×960 pixels. This means that horizontal (x) and vertical (y) eye-in-head positions are mapped onto a 1280×960-pixel plane.

Table 3.1: Main technical specifications of the SMI Eye-Tracking Glasses 2.0, from [44].

Dimensions of glasses	Size: 173 × 58 × 156 mm Weight: 86 g
Calibration	0, 1, and 3 - point calibration modes
Eye-tracking principle	Binocular eye tracking Pupil / CR, dark pupil tracking
Temporal resolution	60 Hz and 30 Hz binocular
Gaze-position accuracy	0.5 ° over all distances, parallax compensation
Tracking distance	40 cm - ∞
Gaze-tracking range	80 ° horizontal, 60 ° vertical
HD scene camera	Resolution: 1280 × 960 px @24 fps Field of view: 60 ° horizontal, 46 ° vertical

3.1.2 Inertial Measurement Unit (IMU)

The head-tracking signals are recorded using the Inertial Measurement Unit (IMU) from x-io Technologies. The IMU consists of a triaxial gyroscope, a triaxial accelerometer, and triaxial magnetometer and has a sampling frequency up to 512 Hz. Moreover, the IMU board includes an AHRS algorithm, which is described in [45]. The AHRS algorithm is a fusion algorithm that combines the signals of all three on-board sensors to compute a measurement of orientation relative to the Earth, which is free from drift. The IMU is very small and lightweight, measuring 57 × 38 × 21 mm and weighing just 49 g including its plastic housing and battery. An illustration of the IMU is shown in Figure 3.3, and the main technical specifications are summarised in Table 3.2.

Head-Tracking Data

The IMU outputs a data stream containing head-in-space orientations stated as ZYX Euler angles. The Euler angles ϕ, θ, and ψ correspond to rotations around the x-, y-, and z-axes of the moving head-centric coordinate system, respectively.

Figure 3.3: Inertial Measurement Unit (IMU) from x-io Technologies, from [46].

Table 3.2: Main technical specifications of the IMU from x-io Technologie, from [46].

Sensor dimensions	Size: $57 \times 38 \times 21\,\mathrm{mm}$ Weight: $49\,\mathrm{g}$
On-board sensors	Triple axis 16-bit gyroscope Triple axis 12-bit accelerometer Triple axis 12-bit magnetometer
On-board algorithms	IMU and AHRS algorithms provide real-time measurement of orientation relative to the Earth
Temporal resolution	Selectable data rates up to 512 Hz

3.2 Complexity Levels

Head-in-space motion consists of three rotational and three translational components, which can result from either moving the head or moving the entire body. Depending on how many of these six components are active, the complexity of the gaze-estimation task varies significantly. In this section, different complexity levels will be introduced. They are summarised in Table 3.3 and explained subsequently.

Table 3.3: Complexity levels of head motion in combination with eye motion.

Level	Eyes	Head (Rotation)			Head
		Horizontal	Vertical	Torsional	(Translation)
1	✓				
2		✓	✓		
3	✓	✓	✓		
4	✓	✓	✓		✓
5	✓	✓	✓	✓	✓

Level 1-3 In the case of complexity level 1, none of the six components of head motion are active, meaning that the head is still and only the eye move. This is similar to cases in which a (head-mounted or remote) static eye tracker is used. In the case of complexity level 2, head rotations are performed but only horizontally and vertically (a nod or shake of the head, for example) but the eyes remain still and try to fixate a single spot. In the case of complexity level 3, horizontal and vertical head rotations are also performed but this time in combination with eye movements.

Complexity levels 2 and 3 can be further divided according to the magnitude of the head rotations which are preformed. Head rotations of less than around $\pm 45\,^\circ$ allow viewing targets to fall comfortably within a two-dimensional plane. Rotations of up to about $\pm 90\,^\circ$ can be performed by moving the head only, whereas rotations of greater than $\pm 90\,^\circ$ require the entire body to be moved.

Level 4-5 In the case of complexity level 4, the translational components of head motion are active in addition to the head and eye movements of level 3. Subjects are now allowed to freely move within the environment. Complexity level 5 allows the maximum degree of mobility, whereby all components of head motion and eye movements can be performed simultaneously.

Starting with complexity level 1, the different levels will be investigated sequentially. The focus of this master's thesis, however, is on complexity levels 1 to 3, which involve horizontal and vertical head rotations in combination with eye movements. Furthermore, it is assumed that subjects are watching distant targets, whereby viewing distances are greater than one meter.

3.3 Model

The goal of this section is to investigate the relationship between eye-in-head motion, head-in-space motion, and eye-in-space motion in order to find a model to combine head- and eye-tracking signals (cf. Figure 3.1). To this end, the vestibulo-ocular reflex (VOR) plays an important role. As described in Section 2.2.4, the function of the VOR is to stabilise the gaze during head movements by generating compensatory eye movements in the opposite direction. Depending on the proximity of the targets being viewed, the gain of these compensatory movements by the eyes varies [6] and thus, so does the relationship between eye position, head orientation, and gaze.

Near-Target Geometry

For a subject viewing near targets, i.e., viewing distances less than one meter [47], the geometric solution has been discussed in [48] and is as follows

$$\begin{aligned} \alpha_{\mathrm{E_R}} &= \tan^{-1}\left(\frac{(\mathrm{D}+\mathrm{R})\,\sin(\gamma-\alpha_H)-\mathrm{I}/2}{(\mathrm{D}+\mathrm{R})\,\cos(\gamma-\alpha_H)-\mathrm{R}}\right), \\ \alpha_{\mathrm{E_L}} &= \tan^{-1}\left(\frac{(\mathrm{D}+\mathrm{R})\,\sin(\gamma-\alpha_H)+\mathrm{I}/2}{(\mathrm{D}+\mathrm{R})\,\cos(\gamma-\alpha_H)-\mathrm{R}}\right), \end{aligned} \tag{3.1}$$

where α_H, $\alpha_{\mathrm{E_R}}$, and $\alpha_{\mathrm{E_L}}$ are the rotation angles of the head, the right eye, and the left eye, respectively, R is the radius of the head, I is the interocular distance, γ is the target eccentricity, and D is the target distance. This signifies that the eyes must compensate for the head rotations with a gain greater than one. Moreover, since the eyes are separated from each other, they must rotate by different amounts. This relationship is summarised in Figure 3.4.

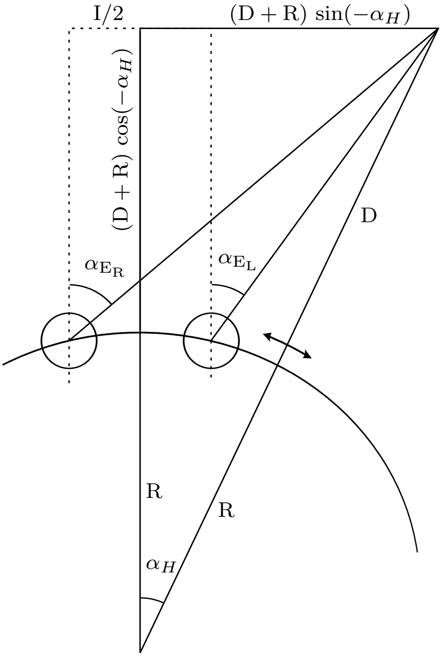


Figure 3.4: The geometric relationship between the rotation angles of the eyes ($\alpha_{\mathrm{E_R}}$, $\alpha_{\mathrm{E_L}}$) and the head-rotation angle (α_H) for a finite radius of rotation (R), target distance (D), and interocular distance (I), from [48].

Distant-Target Geometry

For a subject viewing distant targets, the geometric solution can be simplified. The separation between the eyes (I) as well as the radius of head rotation (R), which was used to account for the eyes not being at the centre of rotation of the head, can be neglected [6]. This results in compensatory VOR eye movements equal and opposite to the head movements. It further implies that the eye-in-space rotation angle α_G can be approximated as the sum of the eye-in-head rotation angle α_E and the head-in-space rotation angle α_H.

$$\alpha_G = \alpha_E + \alpha_H \tag{3.2}$$

3.3.1 Common Coordinate System

In order to be able to apply the relationship of Equation (3.2) to the signals from the eye-tracking glasses and the IMU, the signals first need to be converted to a common coordinate system. For this purpose, it seems advantageous to choose one of the existing systems as the common coordinate system, either the coordinate system of the eye-tracking glasses or that of the IMU. As discussed in Section 3.1, the eye-in-head positions are calculated relative to the scene camera video frame of the eye-tracking glasses so that horizontal and vertical eye positions are mapped to the x- and y-coordinates of a 1280×960-pixel plane. The estimated head-in-space orientation, by contrast, is described in angles in the three-dimensional Euclidean space. The choice of a common coordinate system is affected by multiple factors. On one hand, it is influenced by the differing complexity levels involved (cf. Subsection 3.2). While a two-dimensional reference system would meet all the requirements at the lower complexity levels, the entire three-dimensional space would need to be taken into account at higher complexity levels. On the other hand, the selection of a common coordinate system is also affected by the limitations of our ability to represent stimuli. To calculate the eventual performance measurements, it is much easier to control the experiment by presenting different stimuli in two dimensions rather than in three dimensions. Furthermore, in order to compare the signals which have been adjusted for head movement after gaze estimation with the initial eye-tracking signals, it is advantageous to have both in the same coordinate system. Bearing in mind these considerations, the two-dimensional coordinate system of the eye-tracking glasses has been chosen as the common coordinate system, and is outlined below. This means that all coordinates will be mapped to the x- and y-coordinates of a 1280 x 960-pixel plane and Equation (3.2) can be reformulated as

$$\begin{bmatrix} x_G \\ y_G \end{bmatrix} = \begin{bmatrix} x_E \\ y_E \end{bmatrix} + \begin{bmatrix} x_H \\ y_H \end{bmatrix}, \tag{3.3}$$

where x and y are the x- and y-coordinates in pixels and the subscripts G, E, and H denote eye-in-space (gaze), eye-in-head, and head-in-space positions, respectively.

Mapping to Pixel Plane

As the eye positions are already calculated in pixels relative to the scene camera video frame of the eye-tracking glasses, no further mapping needs to be performed. This means that x_E and y_E of Equation (3.3) correspond to the x- and y-coordinates obtained from the eye-tracking glasses. The head-in-space orientations recorded by the IMU are reported in angles, however, and thus need to be mapped to the common two-dimensional reference system. To do so, firstly, the heading vector of an arbitrary head position $\vec{v}_H$ is calculated. For the sake of convenience, the initial heading vector $\vec{v}_H(0)$ is defined as the vector which points in the direction perpendicular to the centre of the common coordinate system and which coincides with the y-axis of the head-centric coordinate system, i.e., $\vec{v}_H(0) = (0, 1, 0)$. This scenario is depicted in Figure 3.5.

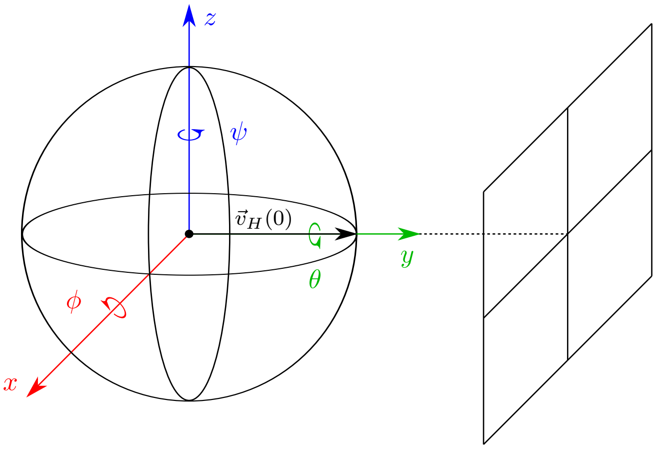


Figure 3.5: The head-centric coordinate system with corresponding head-rotation angles and initial heading vector $\vec{v}_H(0)$ pointing in the direction perpendicular to the centre of the common coordinate system.

Any heading vector $\vec{v}_H$ can then be computed as

$$\vec{v}_H = \mathbf{R}\,\vec{v}_H(0), \tag{3.4}$$

where the matrix $\mathbf{R}$ is a composition of elemental rotations around the principal axes of the head-centric coordinate system

$$\mathbf{R} = \mathbf{R}_z(\psi)\,\mathbf{R}_y(\theta)\,\mathbf{R}_x(\phi), \tag{3.5}$$

with

$$\mathbf{R}_x(\phi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & -\sin\phi \\ 0 & \sin\phi & \cos\phi \end{bmatrix},$$

$$\mathbf{R}_y(\theta) = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix}, \tag{3.6}$$

$$\mathbf{R}_z(\psi) = \begin{bmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The three angles ϕ, θ, and ψ of Equation (3.5) are the ZYX Euler angles reported by the IMU, which correspond to rotations around the x-, y-, and z-axes of the moving head-centric coordinate system, respectively.

In a second step, the heading vector will be mapped to x- and y-pixel-coordinates of the common coordinate system as illustrated in Figure 3.6. To this end, the angles α and β between the initial heading vector and the projections of the heading vector to the initial xy-plane and yz-plane, respectively, are calculated. By considering only a cross section for the x- and y-direction separately, as depicted in Figure 3.7, one can observe that there is a simple relationship between a change in angle (α, β) and a change in pixel (x_H, y_H), as both the resolution as well as the field of view of the scene camera of the eye-tracking glasses are known. For the cross section in the x-direction it is

$$\tan(\alpha) = \frac{x_H}{d}, \tag{3.7}$$

and

$$\tan(\alpha_{max}/2) = \frac{x_{max}/2}{d}, \tag{3.8}$$

which leads to

$$x_H = \frac{x_{max}/2}{\tan(\alpha_{max}/2)} \tan(\alpha), \tag{3.9}$$

where x_{max} denotes the resolution in pixels in the x-direction of the scene video of the eye-tracking glasses and α_{max} denotes the corresponding horizontal angle of view.

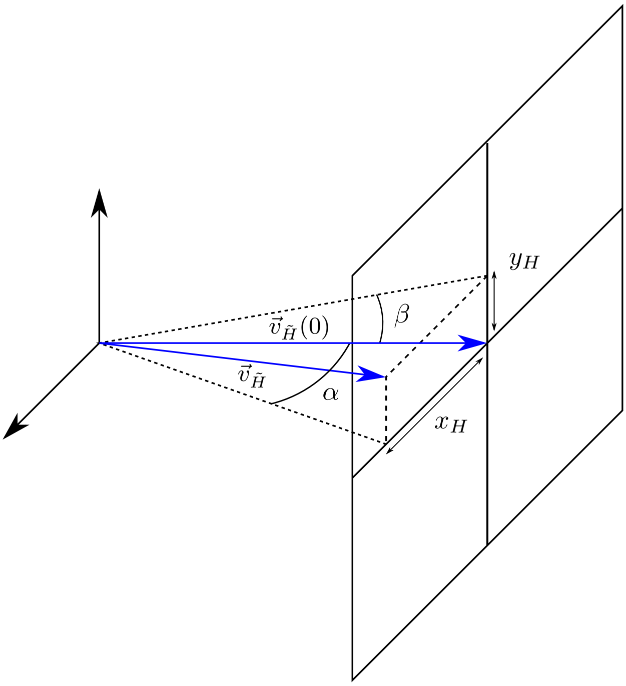


Figure 3.6: The mapping of the heading vector $\vec{v}_H$ to the common coordinate system, whereby α and β denote the angles between the initial heading vector $\vec{v}_H(0)$ and the projections of the heading vector to the initial xy- and yz-planes, respectively. x_H and y_H are the corresponding coordinates in the common pixel coordinate system.

Equations (3.7) and (3.8) are applied to a cross section in the y-direction where y_{max} denotes the resolution in pixels in the y-direction of the scene video and β_{max} denotes the corresponding vertical angle of view. This leads to the final result

$$\begin{bmatrix} x_H \\ y_H \end{bmatrix} = \begin{bmatrix} \frac{x_{max}/2}{\tan(\alpha_{max}/2)} \tan(\alpha) \\ \frac{y_{max}/2}{\tan(\beta_{max}/2)} \tan(\beta) \end{bmatrix}. \tag{3.10}$$

Neglecting torsional head motion and assuming head-rotation angles of around ±30 ° - valid assumptions for complexity levels 1 to 3 - the projection angles α and β can be approximated by the horizontal and vertical head-rotation angles ψ and ϕ, respectively, which are reported by the IMU. Equation (3.10) may then be approximated as

$$\begin{bmatrix} \widetilde{x_H} \\ \widetilde{y_H} \end{bmatrix} = \begin{bmatrix} \frac{x_{max}/2}{\tan(\alpha_{max}/2)} \tan(\psi) \\ \frac{y_{max}/2}{\tan(\beta_{max}/2)} \tan(\phi) \end{bmatrix}. \tag{3.11}$$

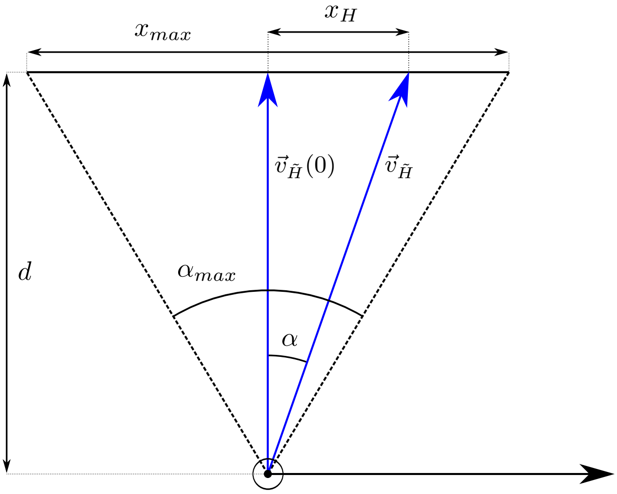


Figure 3.7: The mapping of the heading vector $\vec{v}_H$ to the common coordinate system for a cross section in the x-direction, whereby x_{max} and α_{max} denote the horizontal resolution in pixels and the corresponding angle of view of the scene camera of the eye-tracking glasses, respectively. α is the angle between the initial heading vector $\vec{v}_H(0)$ and the projection of the heading vector to the initial xy-plane, x_H is the corresponding x-axis coordinate in the common pixel coordinate system, and d is the distance to the coordinate system.

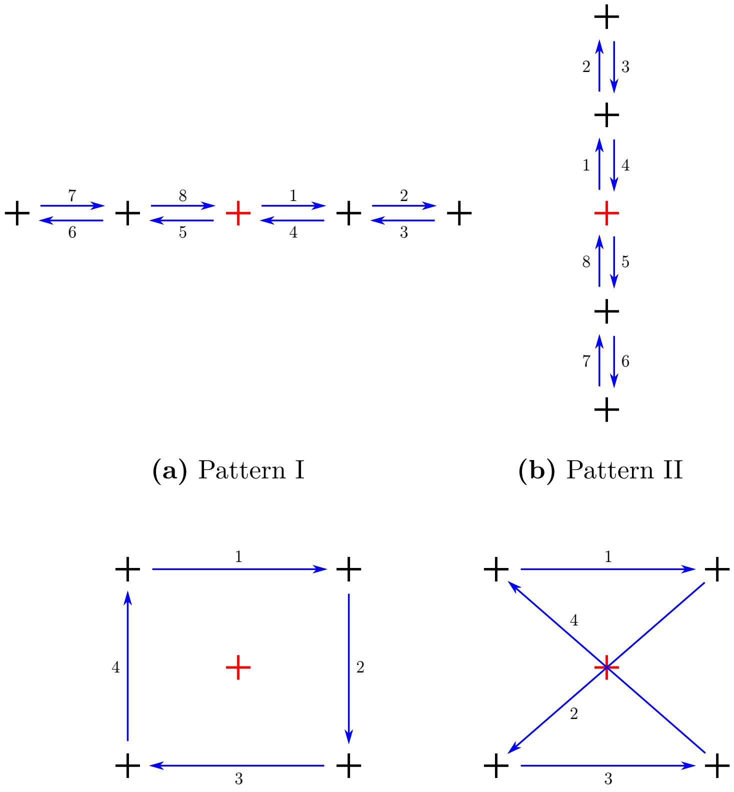
3.4 Signal Analysis

The purpose of the signal analysis section is firstly to illustrate the content of the signals recorded by the different pieces of equipment, i.e., the eye-tracking glasses and the IMU (cf. Section 3.1). Secondly, the section shows how the different types of eye and head movements appear in the eye- and head-tracking signals. This analysis of the eye- and head-tracking signals allows us to gain knowledge about the two types of signals separately. This knowledge will be applied in subsequent sections where the combination of the two signals will be investigated. Therefore, various eye movements, head movements, and combinations thereof are recorded with the eye-tracking glasses and the IMU. The results are presented and discussed in Section 4.1.

3.4.1 Eye-Tracking Glasses

The three most common types of eye movements, saccades, fixations, and smooth pursuits, are described in the background chapter in Section 2.2 and examples are shown in Figures 2.3 and 2.4. These examples are recorded with a static system, which is either tower-mounted or remote. In order to investigate how the three types of eye movements appear in the signals generated by the eye-tracking glasses, different eye movements in combination with head movements are recorded.

Figure 3.8 shows four different movement patterns I - IV. In a first step, the subject performs these four patterns at complexity level 1 (cf. Section 3.2). This means that only eye movements are performed and the head remains still. The subject fixates their gaze at each of the crosses successively, starting with the centre cross, indicated in red, and following the path indicated by the arrows and the corresponding numbers next to them. This should result in an eye-movement sequence composed of fixations and saccades. Thus, when the subject a fixates the crosses, it should result in a fixational eye movement,. When the subject transitions between the crosses, it should result in saccadic eye movements. Pattern I, depicted in Figure 3.8a, and Pattern II, depicted in Figure 3.8b, consist of horizontal and vertical movements only. Pattern III, depicted in Figure 3.8c, alternates between horizontal and vertical movements. Finally, pattern IV, depicted in Figure 3.8d, consists of a combination of horizontal and vertical movements, i.e., diagonal movements.



(a) Pattern I

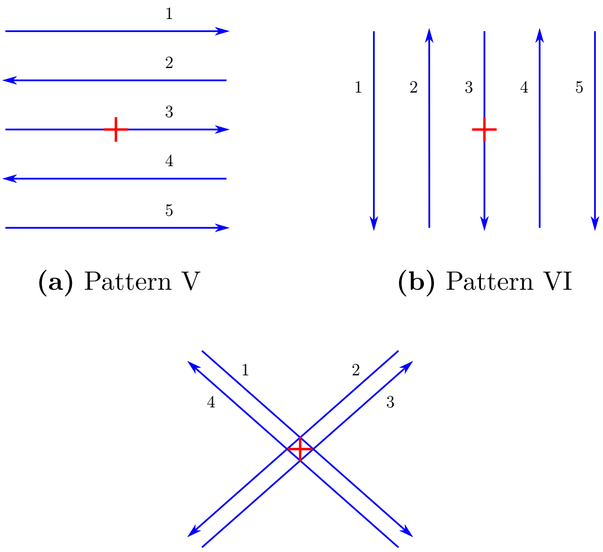
(b) Pattern II

(c) Pattern III

(d) Pattern IV

Figure 3.8: Eye- and head-movement patterns I - IV.

The movement patterns I - IV illustrated in Figure 3.8 aim to stimulate saccadic and fixational eye-movements sequences only. Figure 3.9 shows three additional movement patterns V - VII, which aim to stimulate saccadic, fixational, and smooth-pursuit movements sequences. Again, during the first step, these three patterns are performed by the eyes while the head remains still. The subject's gaze follows a moving target which moves along in the direction of the different arrows. At the beginning and between these movements, the eye fixates the cross in the centre, indicated in red. Patterns I and II, depicted in Figures 3.9a and 3.9b, respectively, consist of only horizontal and vertical smooth-pursuit movements, whereas pattern III, depicted in Figure 3.9c, consists of a combination of horizontal and vertical movements, i.e., diagonal smooth-pursuit movements.



(a) Pattern V

(b) Pattern VI

(c) Pattern VII

Figure 3.9: Eye-movement patterns V - VII.

During the second step, the four patterns I - IV in Figure 3.8 are performed at complexity level 2. In other words, the subject only performs head rotations while fixating the cross in the centre, indicated in red. These head movements are performed for the four patterns in the same way as the eye movements in step one, in that the nose points successively in the direction of each of the crosses, starting with the centre cross, indicated in red, and following the path indicated by the arrows and the corresponding numbers next to them.

Finally, in the third step, the subject performs the four patterns I - IV in Figure 3.8 at complexity level 3, meaning that a combination of eye and head movements is performed. Both the gaze and the head are pointed in the direction of each of the crosses successively. At the beginning, the subject points in the direction of the centre cross, indicated in red. Subsequently, the first transition in performed, whereby firstly the eyes and then the head move in the direction of the arrow and the number one, so that the head movement follows the eye movement. Afterwards, the second transition is performed in the same manner, whereby firstly the eyes and subsequently the head move. This continues sequentially until all the crosses have been fixated.

3.4.2 Inertial Measurement Unit (IMU)

In the previous section, the appearance of the three most common types of eye movements were examined in cases where the eye-tracking signals derive from both eye movements and head movements. In this section, however, the head-tracking signals generated by the IMU will be investigated. Only head movements are recorded, therefore, as eye movements do not influence the head-tracking signals recorded by the IMU. The subject performs the head movements corresponding to the patterns I - IV in Figure 3.8. The nose points successively in the direction of each of the crosses, starting with the centre cross, indicated in red, and following the path indicated by the arrows and the corresponding numbers next to them.

3.5 Model Implementation

This section provides an overview of the different steps which are necessary in order to successfully use the method derived in Section 3.3 to combine the head- and eye-tracking signals. Section 3.5.1 discusses how to calibrate the signals whereas Section 3.5.2 introduces a method to synchronise them. Finally, in Section 3.5.3, an adjustment of the model is presented.

3.5.1 Calibration

While in the previous section, Section 3.4, various signals were recorded with the eye-tracking glasses and the IMU separately, the goal of this section is to perform the recordings simultaneously. To this end, the two recording systems need to be calibrated and synchronised. In the calibration step, the two measurement units are combined. It is advantageous to combine them in such a way that the recording conditions are as similar as possible for the different test persons. After testing different configurations, it was decided to place the IMU at the centre of the test person's forehead, mounted above the glasses as depicted in Figure 3.10.

The calibration itself takes place when the test person stands at an initial position where the recordings are made, e.g., in front of a wall where the stimuli are presented. An initial spot, typically at eye level, is fixated with the

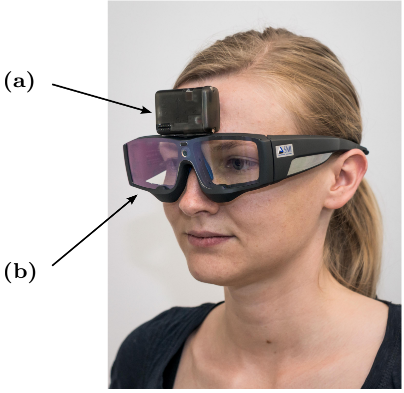


Figure 3.10: The setup of the apparatus, where the IMU (a) is mounted above the eye-tracking glasses (b).

eyes. This spot will be the centre of the common coordinate system described in Section 3.3.1. At the beginning of every recording, the subject fixates the spot for a while without moving the head. Postprocessing of the IMU and the eye-tracking data from this initial sequence enables compensation for possible offsets from the desired initial positions. The desired initial position lies at an x-axis position of 640 px and y-axis position of 480 px, which corresponds to the centre point of the common coordinate system, and the rotation angles ϕ and ψ of 0 °.

3.5.2 Synchronisation

After calibration, the recordings of the eye-tracking glasses and the IMU also need to be synchronised in order to successfully apply the model derived in Section 3.3. The data provided by the eye-tracking glasses and the IMU both contain time stamps for each sample. While the time stamp of the IMU can be set, the time stamp of the eye-tracking glasses is calculated by the recording software provided by the manufacturer and cannot be set. In this case, the time is reported in microseconds from when software is started. This fact complicates the synchronisation task significantly. Synchronisation software and frameworks do exist, such as the ioHub Event Monitoring Framework [49]. Among other things, this framework provides a common time base to automatically synchronise device events from multiple physical and virtual devices, including eye-tracking devices, by using a common eye-tracking interface. This eye-tracking interface, however, currently only supports the static eye-tracking systems of SMI. Problems in synchronising the IMU data with the eye-tracking data were already reported in [27, 29]. In [29], three accelerometers were bumped simultaneously against a fixed object before the recordings started in order to generate extreme acceleration values which could be used for synchronisation. While this method may work if the devices are similar to each other, it is difficult to apply when combining different devices such as the IMU and the eye-tracking glasses. Typical extreme values from eye-tracking data such as blinks do not produce extreme values in the IMU data and vice versa. Investigating the patterns which are present in both signals, however, leads to the idea of applying the VOR for synchronisation. As discussed in Section 2.2.4, the latency of the VOR is extremely short, in the range of 7 and 15 ms. Performing a repetitive head-movement pattern such as nodding, while fixating a single spot with the eyes, generates IMU and eye-tracking data sequences which are very similar in shape. These data sequences can be used for synchronisation by maximising the cross-correlations between them.

Cubic-Spline Data Interpolation

To calculate the cross-correlation and subsequently apply Equation (3.3) on a sample-to-sample basis, i.e., combine the head- and eye-tracking signals, the two signals need to have identical sampling rates. As the eye-tracking signal has a lower sampling rate, therefore, it is resampled. With the aid of these different steps, the two signals are very well synchronised in the beginning but not in the end. A closer inspection of the two signals reveals that the eye-tracking signal is not uniformly sampled, in contrast to the IMU signal, as illustrated in Figure 3.11. Therefore, an additional cubic spline interpolation step is introduced to produce uniformly distributed samples.

The results of combining the head and eye-tracking signals after successful calibration and synchronisation are presented in Section 4.2.1.

3.5.3 Compensatory Factors

During the derivation of Equations (3.3) and (3.11) in Section 3.3, multiple assumptions and approximations were made. In particular, the assumption is made in Equation (3.11) that the field of view of the scene camera of the eye-tracking glasses is 60 ° in horizontal direction and 46 ° in vertical direction. Although these specifications were provided by the manufacturer, they proved to be a poorer than expected estimation of reality. In order to compensate for the different approximations, assumptions, and deviations,

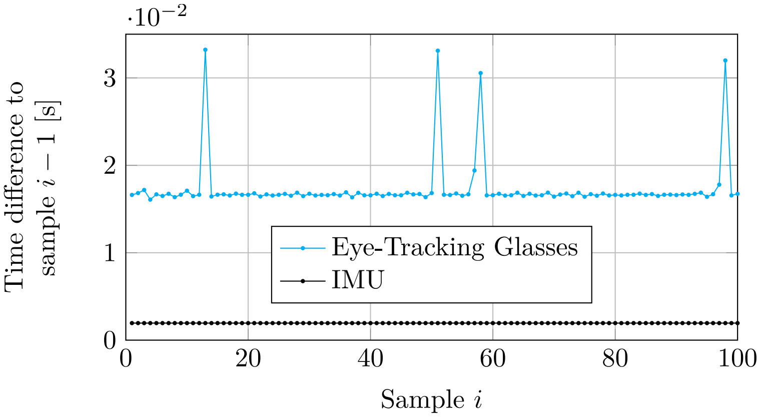


Figure 3.11: The time differences between subsequent samples, showing the uniform and non-uniform sampling rates of the IMU and the eye-tracking data, respectively.

Equation (3.11) is adjusted by introducing two new parameters. One factor compensates for the horizontal direction A, while the other compensates for the vertical direction B.

$$\begin{bmatrix} \widehat{x_H} \\ \widehat{y_H} \end{bmatrix} = \begin{bmatrix} A \, \frac{x_{max}/2}{\tan(\alpha_{max}/2)} \tan(\psi) \\ B \, \frac{y_{max}/2}{\tan(\beta_{max}/2)} \tan(\phi) \end{bmatrix} \tag{3.12}$$

The compensatory factors A and B are designed to optimise the combination of the eye- and head-tracking signals in Equation (3.3). More precisely, they are tuned to minimise the standard deviation of the data from three different recordings where the participant is fixating a stationary target while performing the head-movement patterns I - IV, illustrated in Figure 3.8. The standard deviation is calculated as

$$\sigma_x = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(x_G(n) - \bar{x}_G)^2}, \tag{3.13}$$

$$\sigma_y = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(y_G(n) - \bar{y}_G)^2}, \tag{3.14}$$

for the horizontal and vertical directions separately, where x_G and y_G are the x- and y-coordinates of the resulting eye-in-space data, $\bar{x}_G$ and $\bar{y}_G$ are their respective means, and N is the length of the signal.

The results of tuning the parameters and combining the head- and eye-tracking signals after introducing the new parameters are presented in Section 4.2.2.

3.6 Evaluation

The goal of this section is to describe how the performance of the final model for combining the head- and eye-tracking signals and for determining the eye-in-space motion, derived in the previous sections, is evaluated. For this purpose, a pilot study is performed. The experiment setup and the database is described in Section 3.6.1, and the performance evaluation procedure is described in Section 3.6.2. The corresponding results are presented and discussed in Sections 4.3 and 5.1, respectively.

3.6.1 Experiment Setup and Database

Participant

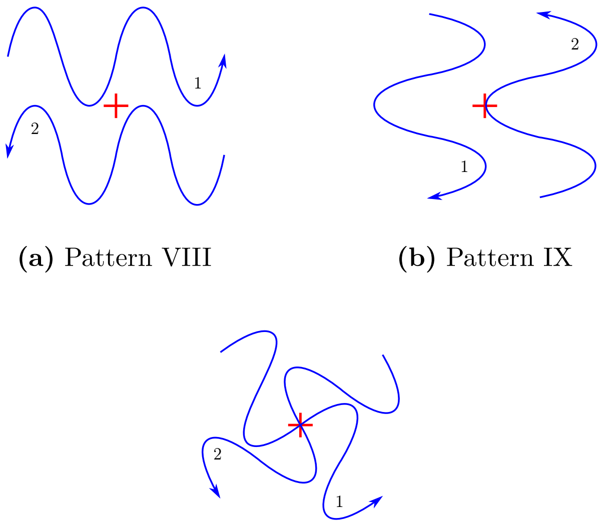
In this pilot study, the eye and head movements of a 25-year-old, female participant with blue eyes are recorded. The participant is wearing neither glasses nor contact lenses.

Apparatus

Eye and head movements are recorded using the tracking equipment described in Section 3.1, which encompasses eye-tracking glasses and an IMU with sampling frequencies of 60 Hz and 512 Hz, respectively. The IMU is mounted above the glasses as described in Section 3.5.1 and depicted in Figure 3.10.

Stimuli

The study is conducted with controlled stimuli as it is important to know which eye movements where actually performed in order to evaluate the performance of the model. Nine different stimuli videos are designed to make the participant perform different eye movements. The stimuli videos 1 - 5 contain the movement patterns I - IV, illustrated in Figure 3.8, whereas stimuli videos 6 and 7 contain the movement patterns V - VII, illustrated in Figure 3.9 (cf. Section 3.4). The stimuli videos 8 and 9 contain three additional movement patterns, VIII - X, shown in Figure 3.12. Similar to the movement patterns V - VII, these additional movement patterns aim to stimulate sequences composed of saccades, fixations, and smooth-pursuit movements. The only difference is that sinusoidally-shaped smooth-pursuit movements are performed instead of straight ones, but still in horizontal, vertical and diagonal directions.



(a) Pattern VIII

(b) Pattern IX

(c) Pattern X

Figure 3.12: Eye-movement patterns VIII - X.

The different stimuli videos represent nine different scenarios, corresponding to different complexity levels. They are summarised in Table 3.4. All of the stimuli videos consist of sequences of stationary, jumping, and moving targets which provoke the fixational, saccadic and smooth-pursuit eye movements, respectively. The targets are presented as coloured dots on a white background and are approximately 1 ° in diameter. The centre is marked with a black cross to facilitate higher targeting accuracy. Depending on the movement pattern and scenario, the dots are coloured differently. Whereas green are followed with the eyes, blue dots are followed with the head.

The saccades are presented with amplitudes ranging from 8 ° to 25 ° of visual angle, whereas the smooth-pursuit movements are presented with amplitudes

Table 3.4: Different scenarios of eye- and head-movement patterns of different complexity levels (CL) presented in the nine stimuli videos (SV).

SV	CL	Scenario
1	1	Movement patterns I - IV are performed with the eyes, while the head remains still.
2	2	Movement patterns I - IV are performed with the head while fixating a single spot with the eyes.
3	3	Movement patterns I - IV are performed with a combination of eye and head movements, whereby each transition is first performed with the eyes and then with the head.
4	3	Movement patterns I - IV are performed with a combination of eye and head movements, whereby each transition is first performed with the head and then with the eyes.
5	3	Movement patterns I - IV are performed with a combination of eye and head movements, whereby each transition is performed with the eyes and the head simultaneously.
6	1	Movement patterns V - VII are performed with the eyes, while the head remains still.
7	3	Movement patterns V - VII are performed with a combination of eye and head movements, whereby each smooth-pursuit movement is performed with the eyes and the head simultaneously.
8	1	Movement patterns VIII - X are performed with the eyes, while the head remains still.
9	3	Movement patterns VIII - X are performed with a combination of eye and head movements, whereby each smooth-pursuit movement is performed with the eyes and the head simultaneously.

ranging from 16 ° to 42 ° of visual angle. The stimuli velocities of the smooth-pursuit movements are between 2.5 °/s and 9 °/s and are constant at each interval. A detailed overview of the stimulus behaviour for the horizontal, vertical, and diagonal directions, separately, can be found in Table 3.5. The total durations of the stimuli videos are between 52 s and 118 s.

Table 3.5: Overview of the individual stimulus behaviour for the horizontal, vertical and diagonal directions.

Stimulus Behaviour	Horizontal	Vertical	Diagonal
Amplitude Range of Saccades	10 - 20 °	8 - 16 °	9 - 25 °
Amplitude Range of SP	20 - 36 °	16 - 30 °	25 - 42 °
Velocity Range of Straight SP	6.5 - 9 °/s	5 - 7.5 °/s	6 - 8.5 °/s
Velocity Range of Sinusoidal SP	3.5 - 5 °/s	2.5 - 4 °/s	3 - 4.5 °/s

In order to calibrate and synchronise the eye-tracking glasses and the IMU, each stimulus video starts with a short sequence, whereby the centre of the screen is indicated, which is also the centre of the common coordinate system. This is followed by a short sequence in which the synchronisation pattern is shown (cf. Sections 3.5.1 and 3.5.2).

Stimuli Presentation

The nine different stimuli videos are presented using a video projector on a large white wall with dimensions of 1.4 m × 1.9 m. The dimensions are chosen as large as possible in order to force the participant not only to move the eyes, but also the head, for some of the stimuli videos. The participant is placed in front of the screen at a distance of 2.5 m and aligns the eyes with the centre of the screen.

Tracking Procedure

The study begins with a 3-point calibration procedure to internally calibrate the eye-tacking glasses. To verify the internal calibration, the participant is asked to fixate some target locations distributed across the screen on a 5 × 5 grid. This calibration procedure is repeated after the presentation of each stimulus video during the experiment. The stimuli videos are explained to the participant prior to the experiment, and some initial rounds of experimentation are conducted to ensure that the participant learns the stimuli and gets used to their speed. Two identical rounds of experimentation are performed and used for data analysis.

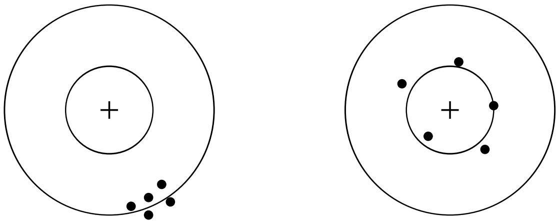
3.6.2 Performance Evaluation Procedure

In order to evaluate the performance of the method, information on the actual eye-in-space movements is required. One way to obtain the true eye movements is to use the stimuli signals as the true eye movements and compare the calculated eye-in-space motion to the presented, known, positions. However, one has to be aware that this method evaluates not only the performance of the method, but also the user's ability to follow the stimulus. As the geometry of the experiment setup is known, i.e., the screen dimensions, the distance to the screen and the stimuli positions, the stimuli coordinates can easily be mapped to the common coordinate system. The synchronisation sequence at the beginning of each stimulus video, which was used to synchronise the eye- and head-tracking signals, can also be used to synchronise the stimulus signal with the gaze-estimation signal. One has to be aware, however, that the synchronisation between stimuli and estimated signals might not be as good as between the head- and eye-tracking signals, as the latencies of visually-mediated eye movements are at least 70 ms, compared to VOR latencies of 7 - 15 ms (cf. Section 2.2.4).

Besides the actual eye-in-space movements, a performance measure is needed. The method is evaluated in terms of precision and accuracy. While precision is the ability of the method to reliably produce an estimation of the eye-in-space motion, accuracy is the average difference between the estimated gaze position and the true gaze position [11]. An illustration of the difference between precision and accuracy is shown in Figure 3.13. High precision is important for calculations of fixation, saccade, and smooth-pursuit measures, whereas high accuracy is crucial in area-of-interest analysis or gaze-contingent studies, where exact gaze positions need to be known.

Precision

Precision (P) is calculated from data samples recorded while the participant is fixating on a stationary target with the eyes. There are two common ways to calculate precision which involve either calculating the standard deviation of the data samples or calculating the root mean square of the inter-sample



(a) High precision, low accuracy

(b) Low precision, high accuracy

Figure 3.13: The difference between precision and accuracy, whereby the cross indicates the true gaze position and the dots indicate the estimated gaze positions, from [11].

distances [11]. In this thesis the standard deviation is used, which can be calculated as

$$P_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_G(n) - \bar{x}_G)^2}, \tag{3.15}$$

$$P_y = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_G(n) - \bar{y}_G)^2}, \tag{3.16}$$

for the horizontal and vertical directions separately, where x_G and y_G correspond to the x- and y-coordinates of the estimated eye-in-space data, $\bar{x}_G$ and $\bar{y}_G$ are their respective means, and N is the length of the signal.

In order to evaluate the effect of compensating for head movements in the eye-tracking data, the precision of four different cases is calculated.

A) Fixating a stationary target while keeping the head still and making no compensation for head movements.

B) Fixating a stationary target while moving the head and making no compensation for head movements.

C) Fixating a stationary target while keeping the head still and compensating for head movements.

D) Fixating a stationary target while moving the head and compensating for head movements.

Accuracy

Accuracy (A) is the average distance between the estimated gaze position and the true gaze position and is calculated as

$$A_x = \frac{1}{N} \sum_{i=1}^{N} (x_T(n) - x_G(n)), \tag{3.17}$$

$$A_y = \frac{1}{N} \sum_{i=1}^{N} (y_T(n) - y_G(n)), \tag{3.18}$$

for the horizontal and vertical directions separately, where x_T and y_T are the x- and y-coordinates of the true eye-in-space data from the stimuli, x_G and y_G are the x- and y-coordinates of the estimated eye-in-space data, and N is the length of the signal.

It is not meaningful to calculate the accuracy for saccades, as they generally show a slightly curved trajectory, whereas the corresponding stimuli consist of a straight line between the starting and end point. Therefore, the accuracy is only calculated for fixation and smooth-pursuit locations and only the intersaccadic intervals are taken into account. A saccadic interval S_j consists of the saccadic latency, which corresponds to the amount of time it takes for the brain to program and initiate the saccade after the stimulus has been presented, and the saccade duration, which is dependent on the amplitude of the saccade. In this thesis, an average duration of 200 ms is used to compensate both for the delay and the saccade duration. Thus, Equations (3.17) and (3.18) can be reformulated as

$$\tilde{A}_x = \frac{1}{N} \sum_{i \in I} (x_T(n) - x_G(n)), \tag{3.19}$$

$$\tilde{A}_y = \frac{1}{N} \sum_{i \in I} (y_T(n) - y_G(n)), \tag{3.20}$$

with

$$I = \{i \in \mathbb{Z} \mid 1 \leq i \leq N, i \notin S\}, \tag{3.21}$$

where x_T and y_T are the x- and y-coordinates of the true eye-in-space data, x_G and y_G are the x- and y-coordinates of the estimated eye-in-space data, N is the length of the signal, and S is the set of all saccadic intervals S_j.

In addition to accounting for intersaccadic intervals when calculating accuracy, a time shift between the stimulus signal and the estimated signal is also introduced to the calculation. This shift is introduced to account for possible delays between the estimated and true signals, which may occur when the stimuli signals are not perfectly synchronised with the estimated signals. Delays may also occur at smooth-pursuit locations, however, because a small difference often exists between the eye and target positions. The time shift is optimised such that the calculation of the accuracy of Equations (3.19) and (3.20) is minimised. This is achieved by choosing a sample shift value in the range of ±50, which corresponds at a frequency of 512 Hz to a time shift of approximately ±100 ms. This approach for calculating the accuracy can be formulated as

$$\hat{A}_x = \min_{m \in [-50,50]} \frac{1}{N} \sum_{i \in I} (x_T(n) - x_G(n-m)), \tag{3.22}$$

$$\hat{A}_y = \min_{m \in [-50,50]} \frac{1}{N} \sum_{i \in I} (y_T(n) - y_G(n-m)), \tag{3.23}$$

where x_T and y_T are the x- and y-coordinates of the true eye-in-space data, x_G and y_G are the x- and y-coordinates of the estimated eye-in-space data, N is the length of the signal, and I is defined as in Equation (3.21).

B - Event Detection

The goal of the event-detection part is to develop and implement a new enhanced event-detection algorithm to detect saccades, fixations, and smooth-pursuit movements from signals containing only eye-in-space motion, i.e., signals which are obtained by combining the eye- and head-tracking signals described in the preceding gaze-estimation part of this chapter. An illustration of this process is shown in Figure 3.14. The performance is then evaluated by comparing the detected events to manual annotations as well as to the detected events of two alternative algorithms.

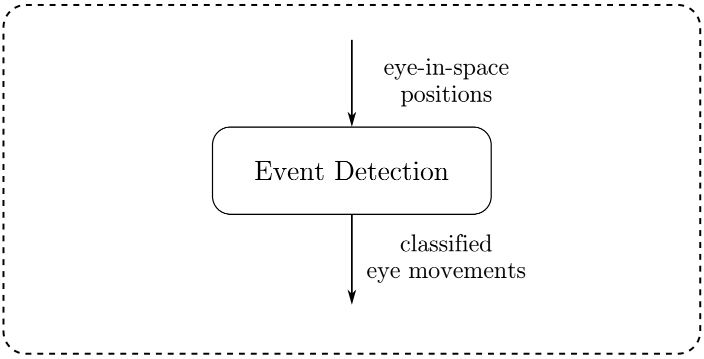


Figure 3.14: An illustration of the event-detection process, whereby the positional eye-in-space signal is segmented into different types of eye movements.

3.7 Proposed Algorithm

Similarly to most event-detection algorithms for ternary classification (cf. Section 2.4), the proposed algorithm comprises three different stages. The first stage is a preprocessing stage, whereby disturbances originating from the recording process are suppressed. In the second stage, the saccades are detected and separated from other types of eye movements. Finally, in the third stage, the remaining samples are classified into fixation and smooth-pursuit movements. A schematic overview of the different stages of the event-detection algorithm is given in Figure 3.15.

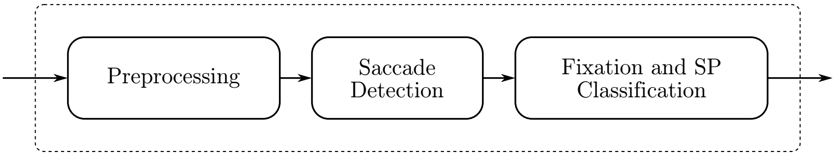


Figure 3.15: The overall structure of the algorithm.

3.7.1 Preprocessing

Before any event detection can be performed, the eye-tracking signals need to be preprocessed in order to remove all parts of the signal that do not correspond to real eye movements such as disturbances originating from the recording process. Disturbances may occur when the pupil and/or the corneal reflection(s) are absent or not correctly detected. Two types of disturbances were observed in the recorded signals: outliers and blinks (cf. Section 4.1). Outliers are samples of the signal with a value outside the gaze-tracking range of 1280 px in horizontal and 960 px in vertical direction. Therefore, all samples corresponding to positions outside a margin of 200 px added to the tracking range are marked as disturbances and excluded from the event detection. During blinks, the x- and y-coordinates of the signal are set to zero by the eye tracker. Thus, all samples with coordinates (0,0) are marked as disturbances and excluded from the event detection. In the beginning and at the end of a blink, the eyelid is not completely closed or open, which may cause saccade-like movements at the start and end of a blink. Hence, additionally to blink-samples, a few samples before and after each blink are marked as disturbances and excluded from the event detection, as the blink rate is not of special interest in this thesis.

3.7.2 Saccade Detection

After the preprocessing stage, the first type of eye movement, the saccades, is detected. The simplest way to detect saccades is to use a single velocity or acceleration threshold [39, 40]. Due to the low sampling frequency of the eye-tracking glasses of 60 Hz, the recorded saccades consist on average of only two to four samples (cf. Section 4.1). Therefore, it is difficult to apply an acceleration threshold to detect the saccades. By calculating the point-to-point velocities of the gaze-estimation signals and applying a velocity threshold instead, the algorithm is able to detect most of the saccades. It has problems identifying the onset and offset of a saccade correctly, however.

In order to solve this problem, the saccade-detection stage is divided into two parts as proposed in [8] (cf. Section 2.4). During the first part, the approximate saccadic intervals are identified, whereas during the second part, the exact onsets and offsets of the saccades are ascertained.

Identification of Approximate Saccadic Intervals

The approximate saccadic intervals are detected using a simple velocity threshold T_V. Assuming a constant sampling frequency, the velocities simply correspond to the distances between the samples. Therefore, the point-to-point velocity for each sample can be ascertained by calculating the distance between the current sample and the previous sample as

$$v(n) = \sqrt{\big(x_G(n) - x_G(n-1)\big)^2 + \big(y_G(n) - y_G(n-1)\big)^2}, \tag{3.24}$$

where x_G and y_G are the x- and y-coordinates of the gaze-estimation signal. Samples with velocities greater than the threshold are classified as saccades, and consecutive saccadic samples are grouped together to gauge the approximate saccadic intervals.

Saccadic Onset and Offset Detection

In [8], the exact onsets and offsets of the saccades are identified for each approximate saccadic interval using three criteria. These criteria are based on directional information in the positional signal, which indicates the deviation from the main direction, inconsistencies between sample-to-sample directions, and the distance between directional changes. Again, it is not possible to apply these criteria to the low-speed signal, as the saccades consist of too few samples.

An inherent physical property of saccades is that their velocity profile shows a triangular shape as depicted in Figure 3.16. The saccade reaches peak velocity sometime in the first half of its total duration, depending on its size. Both before and after this point, the saccade's velocity decreases more or less linearly. In order to take advantage of this fact, two additional velocity thresholds, $T_{V_{ON}}$ and $T_{V_{OFF}}$, are introduced at a lower level, one of which detects the onsets of the saccades and the other of which detects the offsets. This means that for each approximate saccadic interval, adjacent samples are added to the interval as long as their velocities are above the threshold. The search for the exact onsets and offsets is performed in the forward direction

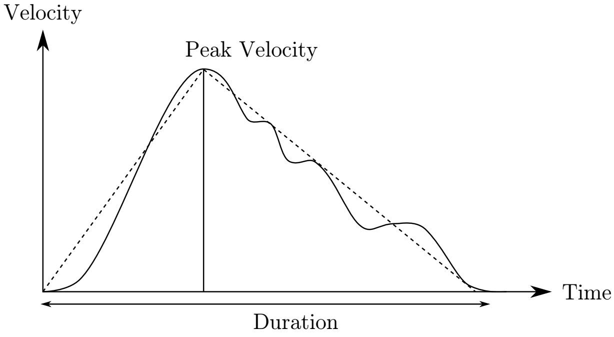


Figure 3.16: Representative velocity profile of a saccade.

to detect offsets using threshold $T_{V_{OFF}}$ and in the backward direction to detect onsets using threshold $T_{V_{ON}}$. To detect onsets, the first sample with a velocity below the threshold is added to the interval, as it also contributes to the high velocity value of the next sample. This is the case since the point-to-point velocities are calculated as the distance between the current sample and the previous one (cf. Equation (3.24)). By introducing the step for saccadic onset and offset detection, the main threshold T_V to identify the approximate saccadic intervals can be made more strict (higher), as only samples around the peak velocities need to be detected (cf. Section 4.4). Even though this method is less sophisticated in nature, it will be demonstrated in Section 4.4 that it performs extremely well.

3.7.3 Fixation and SP Classification

After the saccade detection, the remaining samples of the intersaccadic intervals need to be classified into fixations and smooth-pursuit movements. Therefore, eight different measures are calculated for different sets of consecutive samples of length N. The measures are based on previous work on eye-movement analysis [42, 41, 24, 40, 39] and on the specific characteristics of fixations and smooth-pursuit movements.

Signal Measures

Mean Velocity. The first measure is the mean velocity which is calculated as

$$MV = \frac{1}{N} \sum_{n=1}^{N} v(n), \tag{3.25}$$

where $v(n)$ are the point-to-point velocities of the samples calculated according to Equation (3.24). The mean velocity is generally lower for samples which are fixational eye movements than for samples which are smooth-pursuit movements. The same applies for the next four measures: slope, integral, energy, and dispersion.

Slope. The second measure is the slope of the signal which is calculated as

$$S = S_x + S_y, \tag{3.26}$$

where S_x and S_y are the slopes of first-order polynomials fitted to each of the x- and y-coordinates over time, respectively.

Integral. The third measure is the integral of the signal, i.e., the area under the graph of the signal. This area approximates to a trapezoid and is calculated for the horizontal and vertical directions separately as

$$I_x = N\,\frac{|\,x(N) - x(1)\,|}{2}, \quad \text{and} \quad I_y = N\,\frac{|\,y(N) - y(1)\,|}{2}. \tag{3.27}$$

I_x and I_y are combined to the single measure

$$I = I_x + I_y. \tag{3.28}$$

Energy. The fourth measure is the energy of the signal which is calculated for the horizontal and vertical directions separately as

$$E_x = \sum_{n=1}^{N} x(n)^2, \qquad \text{and} \qquad E_y = \sum_{n=1}^{N} y(n)^2. \tag{3.29}$$

E_x and E_y are combined to the single measure

$$E = E_x + E_y. \tag{3.30}$$

Dispersion. The fifth measure is the dispersion of the signal which is calculated as

$$D = \sqrt{\big(max(x) - min(x)\big)^2 + \big(max(y) - min(y)\big)^2}. \tag{3.31}$$

Directional Variation. The sixth measure describes the directional variation of the signal, which is determined by applying Principle Component Analysis. The first principle component is the direction in which the samples exhibit greatest variation, whereas the second principle component is the direction in which the samples exhibit least variation. As the corresponding eigenvalues quantify the amount of variation observed in the respective directions, they are used to calculate the sixth measure.

$$DV = \frac{\lambda_2}{\lambda_1} \tag{3.32}$$

A value of DV which is close to one means that the samples are equally spread in both directions, such is the case for fixations. Conversely, a lower value implies that the samples are more spatially spread in one direction than the other, which is the case for smooth-pursuit movements.

Consistency in Direction. The seventh measure evaluates whether the samples have a consistent direction or not. Therefore, the Euclidean distance d_{ED} between the positions of the first and last sample of the interval is calculated. This is compared to the length of the projections of the samples onto the direction of the first principle component d_{PC_1}. An illustration of both distances is shown in Figure 3.17.

$$CD = \frac{d_{ED}}{d_{PC_1}} \tag{3.33}$$

A value of CD which is lower than one indicates that the range of the samples in the interval is much larger than the actual distance between the first and last sample, which is typical for fixations.

Positional Displacement. The eight measure describes the positional displacement. Thus, the relationship between the Euclidean distance d_{ED} and the trajectory length d_{TL} between the positions of the first and last sample of the interval is evaluated as

$$PD = \frac{d_{ED}}{d_{TL}}. \tag{3.34}$$

A value of PD equal to one corresponds to a straight line. Thus, higher values of PD are characteristic of smooth-pursuit movements.

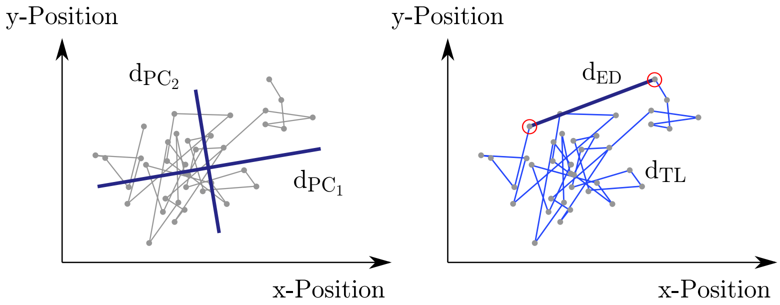


Figure 3.17: An illustration of the distances used to calculate the seventh and eight measures. d_{PC_1} and d_{PC_2} are the lengths of the projections of the samples onto the directions of the principle components. d_{ED} and d_{TL} are the Euclidean distance and the trajectory length between the positions of the first and last sample of the interval, respectively.

Classification Based on Sliding Windows

In order to extract each signal measure and subsequently classify the samples of the intersaccadic intervals into fixations and smooth-pursuit movements, two different sliding-window approaches are applied. They are explained below:

Basic Sliding Window (BSW). This approach is similar to the sliding-window approach used in the basic algorithms I-DT and I-VDT (cf. Section 2.4). It can be performed in the forward and in the backward direction. For the forward direction (BSW_F), a temporal window w of length l_w is defined, which initially spans the first l_w consecutive samples of the intersaccadic interval. The signal measure is calculated for the samples within the window and is compared to a threshold. If the value of the signal measure is below (or above) the threshold, depending on the type of measure, the window is expanded by one sample (to the right) and the value of the signal measure is recalculated for the new set of samples. This step is repeated until the value of the signal measure is above (or below) the threshold, respectively. All samples within the window are then classified as fixations and a new window is initialised with the first l_w consecutive samples which remain. If the value of the signal measure is above (or below) the threshold instead, the window is moved one sample (to the right) and the first sample of the previous window is marked as a smooth-pursuit movement.

For the backward direction (BSW_B), the temporal window w of length l_w initially spans the last (instead of the first) l_w consecutive samples of the intersaccadic interval and the window is expanded and moved to the left (instead of to the right). Otherwise, the procedure is the same.

Weighted Average of Sliding Windows (WASW). In this approach, a temporal window w of length l_w is defined, which, beginning at the start of the interval, moves over the intersaccadic interval one sample at a time until the end of the interval is reached. This results in a total of $N_w = N - l_w + 1$ windows w_i, where N is the number of samples in the intersaccadic interval, and l_w is the length of the temporal window. Each sample may, thus, belong to more than one window, and at most to $N_s = l_w$ different windows, which corresponds to the size of the window. For each set of samples $S_i = \{s_1, s_2, \ldots, s_{l_w}\}$ within a window w_i, the signal measure M_i is calculated as described in the previous subsection. The value of the signal measure for a specific sample n is then calculated as the weighted average value of the measures M_i of all windows to which the sample belongs to, as

$$\overline{M}(n) = \frac{\sum_{i \in I} a_i \, M_i}{\sum_{i \in I} a_i}, \qquad \text{for } 1 \leq n \leq N \tag{3.35}$$

where I is the set of windows to which sample n belongs to, a_i are the corresponding weighting factors, and N is the number of samples in the intersaccadic interval. An illustration is shown in Figure 3.18.

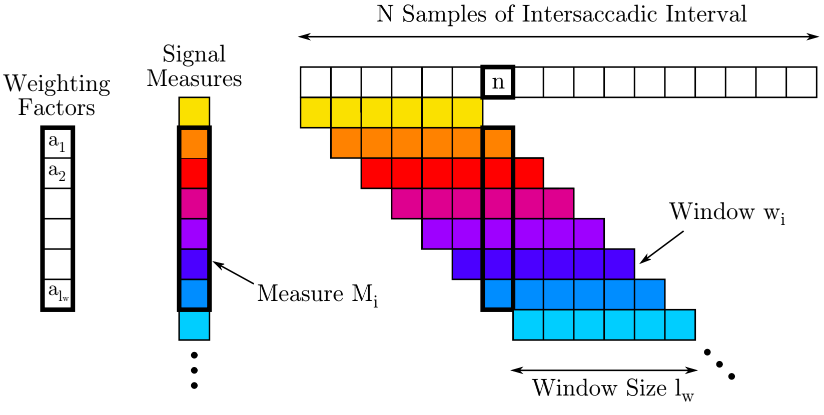


Figure 3.18: Illustration of the WASW approach.

Equation (3.35) can also be expressed in the form of a discrete convolution of two vectors as

$$\overline{M}(n) = \frac{(\mathbf{M} * \mathbf{a})(n)}{(\mathbf{1} * \mathbf{a})(n)}, \qquad \text{for } 1 \leq n \leq N \tag{3.36}$$

where $\mathbf{M} = [M_1, M_2, \ldots, M_{N_w}]$ is a vector containing the signal measures calculated for the different windows w_i, $\mathbf{1}$ is an all-ones vector of size N_w, and $\mathbf{a} = [a_1, a_2, \ldots, a_{l_w}]$ is a vector which contains the weighting factors.

Four different versions of weighting vectors are applied in the $WASW$ approach: a constant weighting vector $\mathbf{a}$ ($WASW_1$), a triangular weighting vector $\mathbf{b}$ ($WASW_2$), an exponential (Hann-Poisson) vector $\mathbf{c}$ ($WASW_3$), and an inverse-exponential vector $\mathbf{d}$ ($WASW_4$). These are

$$a_n = 1, \tag{3.37}$$

$$b_n = \frac{2}{l_w}\left(\frac{l_w}{2} - \left|(n-1) - \frac{l_w - 1}{2}\right|\right), \tag{3.38}$$

$$c_n = \frac{1}{2}\left(1 - \cos\left(\frac{2\pi n}{N+1}\right)\right)\exp\left(\frac{-5\,|N+1-2n|}{N-1}\right), \tag{3.39}$$

$$d_n = [c_{n/2+1}, c_{n/2+2}, \ldots, c_{l_w}, c_1, c_2, \ldots, c_{n/2}], \tag{3.40}$$

for $1 \leq n \leq l_w$, where l_w is the size of the window, and N is the number of samples in the intersaccadic interval. An illustration of the different weighting vectors for a window size $l_w = 15$ samples is shown in Figure 3.19.

The use of a constant weighting vector $\mathbf{a}$ in the calculation of Equation (3.36) means that all the measures of the different windows to which sample n belongs to have the same influence on the average $\overline{M}(n)$. The use of a triangular $\mathbf{b}$, exponential $\mathbf{c}$ or inverse-exponential weighting vector $\mathbf{d}$ instead, means that the measures of the different windows to which sample n belongs to have a differing influence on the average $\overline{M}(n)$, depending on whether the sample n is located in the centre or on the margin of the sliding window.

After successfully calculating the weighted-average value of the signal measure for each sample $\overline{M}(n)$, the samples are classified into fixations and smooth-pursuit movements by comparing the average value $\overline{M}(n)$ to a threshold T_M,

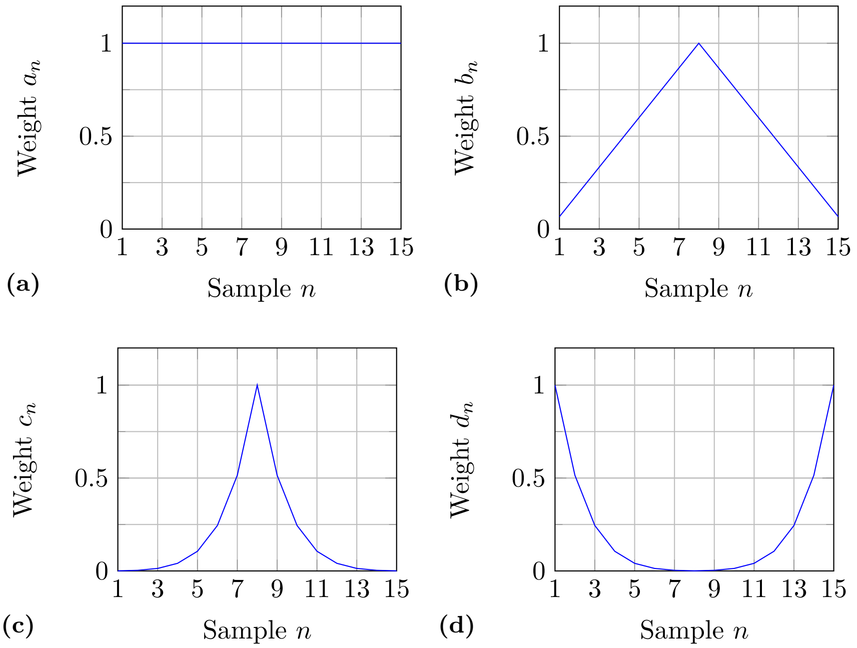


Figure 3.19: Illustration of the four different weighting vectors of Equations (3.37) - (3.40) for a window size $l_w = 15$ samples: (a) constant, (b) triangular, (c), exponential, and (d) inverse-exponential weighting vector.

which is set for each signal measure. The samples are classified as fixations if the value is below (or above) the threshold, depending on the type of measure. Otherwise, they are classified as smooth-pursuit movements.

Combination of Signal Measures

The performance evaluation of the classification into fixations and smooth-pursuit movements for the eight different measures and six different sliding-window approaches is presented in Sections 4.5.1 and 4.5.2. The goal of this section is to investigate whether the classification performance can be further improved by classifying the samples based on a combination of different signal measures and thresholds. Therefore, the WASW approach is slightly adapted. Instead of only calculating an average value $\overline{M}(n)$ for every sample for one signal measure, an average value for each signal measure is calculated, i.e., $\overline{M}_{MV}(n)$, $\overline{M}_{I}(n)$, …, $\overline{M}_{PD}(n)$. The results of combining two signal measures to carry out the classification are presented in Section 4.5.3.

3.8 Evaluation

The goal of this section is to evaluate the performance of the proposed event-detection algorithm. Therefore, the data of the pilot study after gaze estimation is used, namely the signals containing the estimated eye-in-space motion. The database is briefly summarised in Section 3.8.1. An evaluation procedure and corresponding measures are introduced in Section 3.8.2 and a description of two algorithms used for comparison is given in Section 3.8.3. The results are presented and discussed in Section 4.6.

3.8.1 Database

A detailed description of the experiment and the database can be found in Section 3.6.1. Nine different stimuli videos are presented, and two identical rounds of experimentation are performed and used for data analysis. The recorded data from one round is used to develop the algorithm, so that the different thresholds can be tuned. The data from the second round is used to make the final evaluation of the algorithm.

3.8.2 Evaluation Procedure

In order to evaluate the performance of the event-detection algorithm, information about the eye movements which are actually performed is required. Three common ways are reported in the literature to obtain this information. The first method is to use the stimuli signals as the true eye movements [31, 39, 50, 42]. This method, however, evaluates not only the detection performance of the algorithm, but also the user's ability to reliably follow the stimuli signals. Another method is to simulate the eye movements by generating position signals [51, 27]. The drawback of simulations is that it is difficult to generate eye-movement signals that are comparable to real signals. The last method, which is applied in the present thesis, is to manually annotate the eye-tracking data [11, 15, 8, 24]. For this purpose a Matlab GUI is used, which shows the horizontal and vertical eye positions over time and in the spatial domain, and the corresponding velocity over time. Although it is very time consuming, manual annotation enables the classification of every sample, which is important in order to facilitate comparison between different algorithms. One has to be aware, however, that this method might suffer from human subjectivity and inconsistency.

In addition to the true eye movements, a performance measure is also needed. In [31, 39], an evaluation procedure based on scores is proposed for saccades, fixations, and smooth-pursuit movements. Again, this procedure is highly dependent on the user's ability to reliably follow the stimulus. Therefore, the performance of the algorithm is evaluated in terms of sensitivity and specificity, which are the performance measures most commonly used when discriminating between two or more groups [52].

Sensitivity and Specificity

Sensitivity and specificity are calculated for each type of eye movement i separately, with $i \in \{\mathrm{S} = \mathrm{Saccade}, \mathrm{F} = \mathrm{Fixation}, \mathrm{P} = \mathrm{Smooth\ Pursuit}\}$.

They are defined as

$$SENS_i = \frac{TP_i}{TP_i + FN_i}, \tag{3.41}$$

$$SPEC_i = \frac{TN_i}{TN_i + FP_i}, \tag{3.42}$$

where TP_i are the *true positives*, i.e., the number of correctly classified samples for eye-movement type i, FN_i are the *false negatives*, i.e., the number of samples that should have been classified as eye-movement type i, but have incorrectly been classified as another type of eye movement, TN_i are the *true negatives*, i.e., the number of samples that the algorithm correctly classified as another type of eye movement than i, and FP_i are the *false positives*, i.e., the number of samples that the algorithm falsely classified as eye-movement type i.

The sensitivity, also referred to as the *true-positive rate*, describes the algorithm's ability to correctly classify a certain type of eye movement. The specificity, also referred to as *true-negative rate*, describes the algorithm's ability to find only the samples of eye-movement type i, i.e., to correctly exclude the other types of eye movements. For both measures, a value close to one is desired.

In order to determine TP_i, FN_i, TN_i, and FP_i for each type of eye movement i, the confusion matrix is calculated. This is a matrix which visualises the performance of the event-detection algorithm. The rows represent the true (manually annotated) classes, whereas the columns represent the estimated classes. This means that each element N_{ij} of row i and column j represents a number of samples which belong to the (true) class i and are detected by the algorithm as class j, with i and $j \in \{$S = Saccade, F = Fixation, P = Smooth Pursuit$\}$. Thus, the diagonal elements of the confusion matrix show the number of correctly classified samples for each class, and the off-diagonal elements show the errors. An example of a confusion matrix for the tree classes of eye movements is shown in Figure 3.20.

		Estimated		
		S	**F**	**P**
True	**S**	N_{SS}	N_{SF}	N_{SP}
	F	N_{FS}	N_{FF}	N_{FP}
	P	N_{PS}	N_{PF}	N_{PP}

Figure 3.20: Confusion Matrix for the three types of eye movements: saccades (S), fixations (F), and smooth pursuits (P).

Using the confusion matrix, the sensitivity and specificity of saccades, for instance, are then explicitly calculated as

$$SENS_S = \frac{N_{SS}}{N_{SS} + (N_{SF} + N_{SP})}, \tag{3.43}$$

$$SPEC_S = \frac{N_{FF} + N_{FP} + N_{PF} + N_{PP}}{(N_{FF} + N_{FP} + N_{PF} + N_{PP}) + (N_{FS} + N_{PS})}. \tag{3.44}$$

ROC Curve

In order to study the behaviour of the event-detection algorithm for different parameter settings, the sensitivity and specificity can be combined into a receiver-operating characteristic (ROC). This can be achieved by plotting the sensitivity against the *false-positive rate*, which is the complementary of the specificity (1 - specificity). The ROC curve helps in choosing optimal parameter values, such that an acceptable trade-off between the two counterbalancing measures is achieved. This can be done by maximising the balanced accuracy which is defined as

$$BACC_i = \frac{SPEC_i + SENS_i}{2}. \tag{3.45}$$

3.8.3 Algorithm Comparison

In order to further evaluate the detection performance of the proposed algorithm, the detected events are compared to those detected by two alternative algorithms, the I-VDT algorithm described in [39] and the event detector built-in to the eye-tracking glasses.

I-VDT Algorithm

The I-VDT algorithm is chosen because it is one of the few algorithms which is able to perform ternary classification, meaning that it can discriminate between saccades, fixations, and smooth-pursuit movements (cf. Section 2.4). In addition, it was the algorithm which performed the best in a previous study, where three basic ternary classification algorithms were compared and evaluated [39].

Algorithm 1: I-VDT

```
Data: array of eye-position points, velocity threshold $T_V$,
      dispersion threshold $T_D$, temporal window size $T_W$
Result: array of fixations, saccades, and smooth pursuits
Calculate point-to-point velocities for each point;
Mark all points above $T_V$ as saccades;
Initialise temporal window over first points in the remaining
eye-movement trace;
while temporal window does not reach the end of array do
    Calculate dispersion of points in window;
    if dispersion $< T_D$ then
        while dispersion $< T_D$ do
            Add one more unclassified point to window;
            Calculate dispersion of points in window;
        end
        Mark the points inside the window as fixations;
        Clear window;
    else
        Remove first point from window;
        Mark first point as a smooth pursuit;
    end
end
Return saccades, fixations, and smooth pursuits;
```

Built-in Event Detector of Eye-Tracking Glasses

The gaze-analysis software distributed with the eye-tracking glasses has a built-in event detector, which uses a velocity-based algorithm to detect saccades and fixations. Saccades are explicitly detected, whereas all other eye movements are collectively labelled as fixations. Therefore, the algorithm is not able to discriminate between fixations and smooth-pursuit movements. There are no user-adjustable parameters for the algorithm since it relies on a combination of fixed parameters. These parameters are set according to the sampling rate and the physiological limits of eye movements, and the adaptive thresholds used to discriminate between saccades and other types of fast eye movements. The algorithm is applied directly to the recorded data from the eye-tracking glasses, without preliminary head-movement compensation.

Chapter 4

Results

In this chapter, the results of the two main parts, *Gaze Estimation* and *Event Detection*, are presented. The chapter is outlined in the same way as Chapter 3. A discussion of the results can be found in Chapter 5.

A - Gaze Estimation

The *Gaze Estimation* part begins with an investigation of the properties of the eye- and head-tracking signals in Section 4.1. The combination of the two signals, using both the initial and the adjusted model derived in Sections 3.3 and 3.5.3, is presented and discussed in Section 4.2. Finally, in Section 4.3, the combination model is evaluated in terms of precision and accuracy.

4.1 Signal Analysis

4.1.1 Eye-Tracking Glasses

Figures 4.1 - 4.3 show the eye-tracking signals which were produced when the eye-movement patterns I - VII, illustrated in Figures 3.8 and 3.9, were performed without head movement. Figures 4.1 and 4.2 show the horizontal and vertical eye positions over time reported by the eye-tracking glasses as well as the spatial velocity over time, whereas Figure 4.3 shows the x- and y-positions in the spatial domain for the different movement patterns I - VII, respectively.

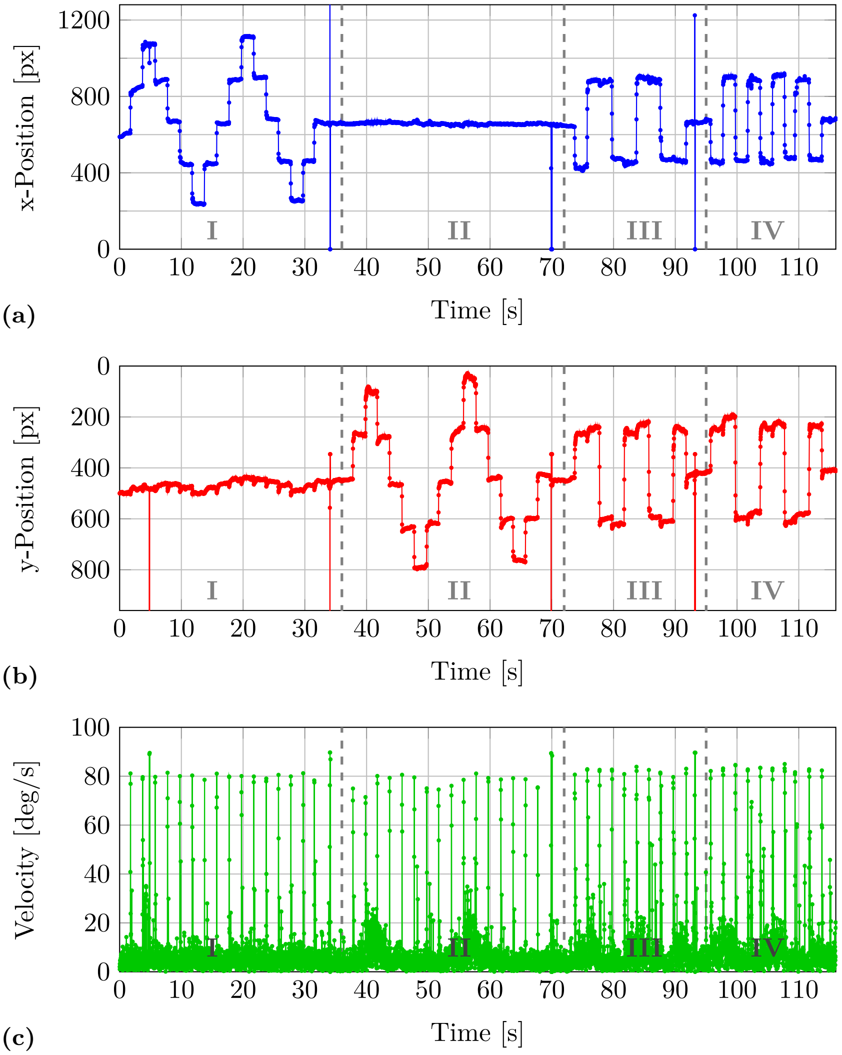


Figure 4.1: Horizontal (a) and vertical (b) eye positions over time reported by the eye-tracking glasses while performing the eye-movement patterns I - IV, illustrated in Figure 3.8, head movement. Corresponding spatial velocity over time (c).

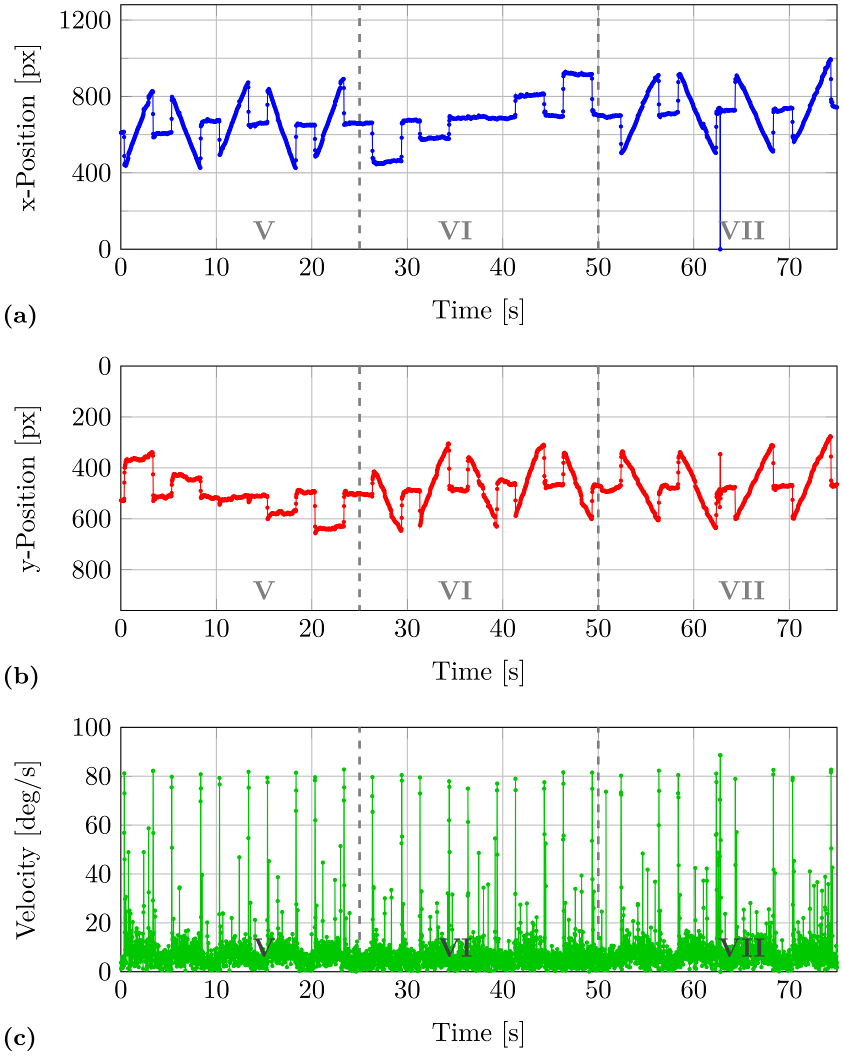
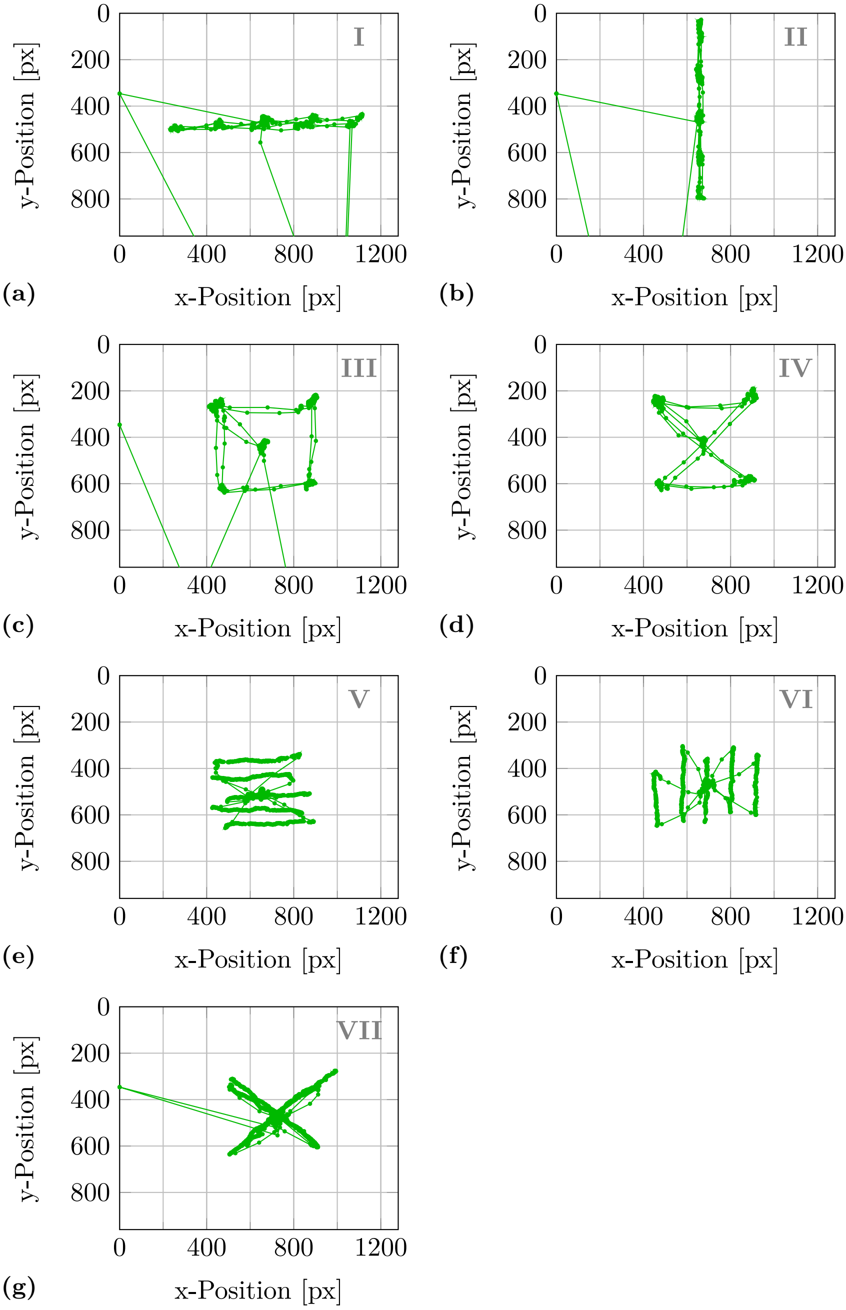


Figure 4.2: Horizontal (a) and vertical (b) eye positions over time reported by the eye-tracking glasses while performing the eye-movement patterns V - VII, illustrated in Figure 3.9, without head movement. Corresponding spatial velocity over time (c).



(a) (b) (c) (d) (e) (f) (g)

Figure 4.3: Horizontal and vertical eye positions in the spatial domain reported by the eye-tracking glasses while performing the eye-movement patterns I - VII, illustrated in Figures 3.8 and 3.9, without head movement.

As expected, when the eye-movement patterns I - IV are performed, it results in an eye-movement sequence comprised of fixations and saccades. This is evidenced by the clustering of samples at positions where crosses are located in the movement patterns and by the scarcity of samples at transitions between the crosses (arrows) in the movement patterns (cf. Figures 4.1 and 4.3a-d). By contrast, when the eye-movement patterns V - VII are performed, it results in a eye-movement sequence comprised of fixations, saccades, and smooth pursuits. This is evidenced by the clustering of samples at the position where the centre cross is located in the movement patterns and the scarcity of samples at transitions between the cross and the starting and end points of the moving target. Furthermore, a kind of elongated clustering of samples is in evidence at locations of the moving targets in the movement patterns (cf. Figures 4.2 and 4.3e-g).

Due to the low sampling frequency of 60 Hz, saccades consist on average of only two to four samples, which makes it difficult to apply sophisticated detection criteria in an eventual event-detection algorithm. The fact that there are few samples, however, also means that the movements were fast, which is indicative of saccades. This can also be observed in the velocity profile of the signals in Figures 4.1 and 4.2. Saccades reach velocities of around 80 °/s and can easily be distinguished from the other two types of movements. The velocity ranges of fixations and smooth pursuits, by contrast, overlap in the recordings (cf. Figures 4.1 and 4.2), which makes it more difficult for an eye-movement detection algorithm to distinguish between these movements.

In addition to the three types of eye movements, it can be observed that parts of the signals do not correspond to real eye movements, e.g., at times 5 s, 34 s, 70 s, and 93 s in Figure 4.1. These disturbances may occur if the pupil and/or the corneal reflection(s) are absent or cannot correctly be detected. The signal is either zero, which is the case during blinks, or reaches a value outside the gaze-tracking range of 1280 px in the horizontal and 960 px in the vertical direction. Another type of disturbance referred to as one- or two-sample spikes may occur if the estimated position of the corneal reflection alternates between two possible locations. This type of disturbance, however, is not present in the signals recorded by the eye-tracking glasses. It is important to be aware of possible disturbances and to remove them before performing any event detection.

Figures 4.4 and 4.5 depict the eye-tracking signals at complexity levels 2 and 3, whereby the movement patterns I - IV, illustrated in Figure 3.8, were performed. Figure 4.4 illustrates complexity level 2, which involves head movement while fixating a single spot with the eyes, whereas Figure 4.5 illustrates complexity level 3, which involves a combination of eye and head movements. The figures show the horizontal and vertical eye positions reported by the eye-tracking glasses over time as well as in the spatial domain for the different movement patterns I - IV, respectively. Although, the eyes were fixating a single spot in the case of complexity level 2, eye movements were recorded by the eye-tracking glasses (cf. Figure 4.4). These movements are the eye-in-head movements which were generated to compensate for the head-in-space movements in order to stabilise the gaze. It is easy to recognise that they look very similar to smooth-pursuit eye movements (cf. Figures 4.2 and 4.3e-g). This means that in contrast to complexity level 1, it is not possible to draw any conclusion on the eye-in-space movement. In the case of complexity level 3, whereby the eyes as well as the head are moving, it is even harder to extract information about the eye-in-space motion from the recorded eye-tracking data. As a result, it is barely possible to perform any event detection without conducting a preliminary head-movement compensation step.

4.1.2 Inertial Measurement Unit (IMU)

Figure 4.6 depicts the head-tracking signals generated by the IMU when the head-movement patterns I - IV, illustrated in Figure 3.8, were performed. The figure shows three plots corresponding to the three head rotation angles ϕ, θ, and ψ over time, respectively. There is almost no visible variation in the rotation angle θ, meaning that no torsional head rotation is present. Furthermore, by comparing the signals of the rotation angles ψ and ϕ to the x- and y-positions of the eye-tracking signals, illustrated in Figure 4.4, whereby the same head-movement patterns were performed while fixating a single spot with the eyes, it can be observed that the shapes of the signals are extremely similar. In other words, it stands to reason that the horizontal eye movements resulting from the VOR can be compensated using the rotation angle ψ, whereas the vertical eye movements can be compensated using the rotation angle ϕ. These two observations, thus, support the assumptions made at the end of Section 3.3 as well as the subsequent approximation in Equation (3.11).

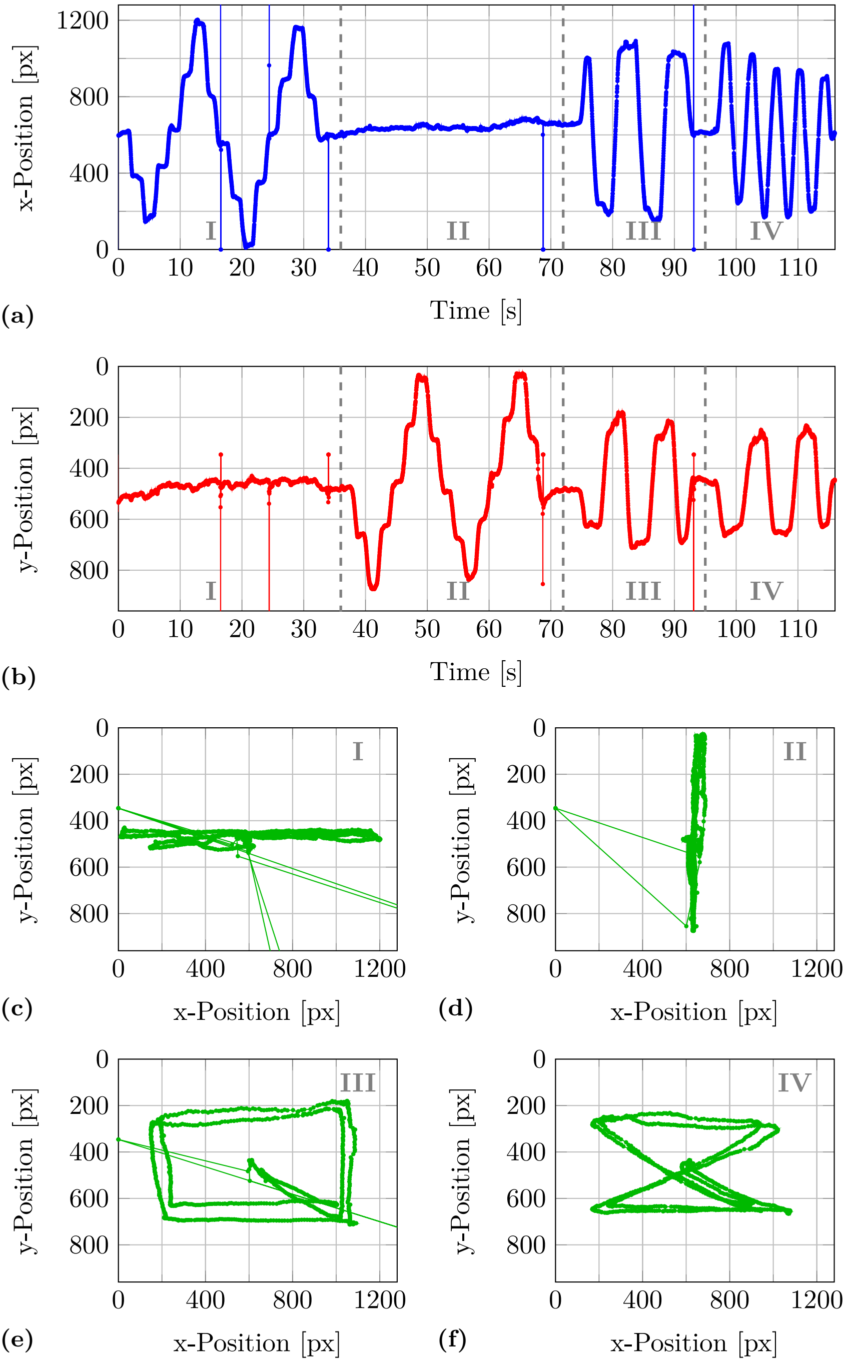


Figure 4.4: Eye positions over time (a)-(b) and in the spatial domain (c)-(f) reported by the eye-tracking glasses while performing the movement patterns I - IV with the head, illustrated in Figure 3.8, and fixating a single spot with the eyes.

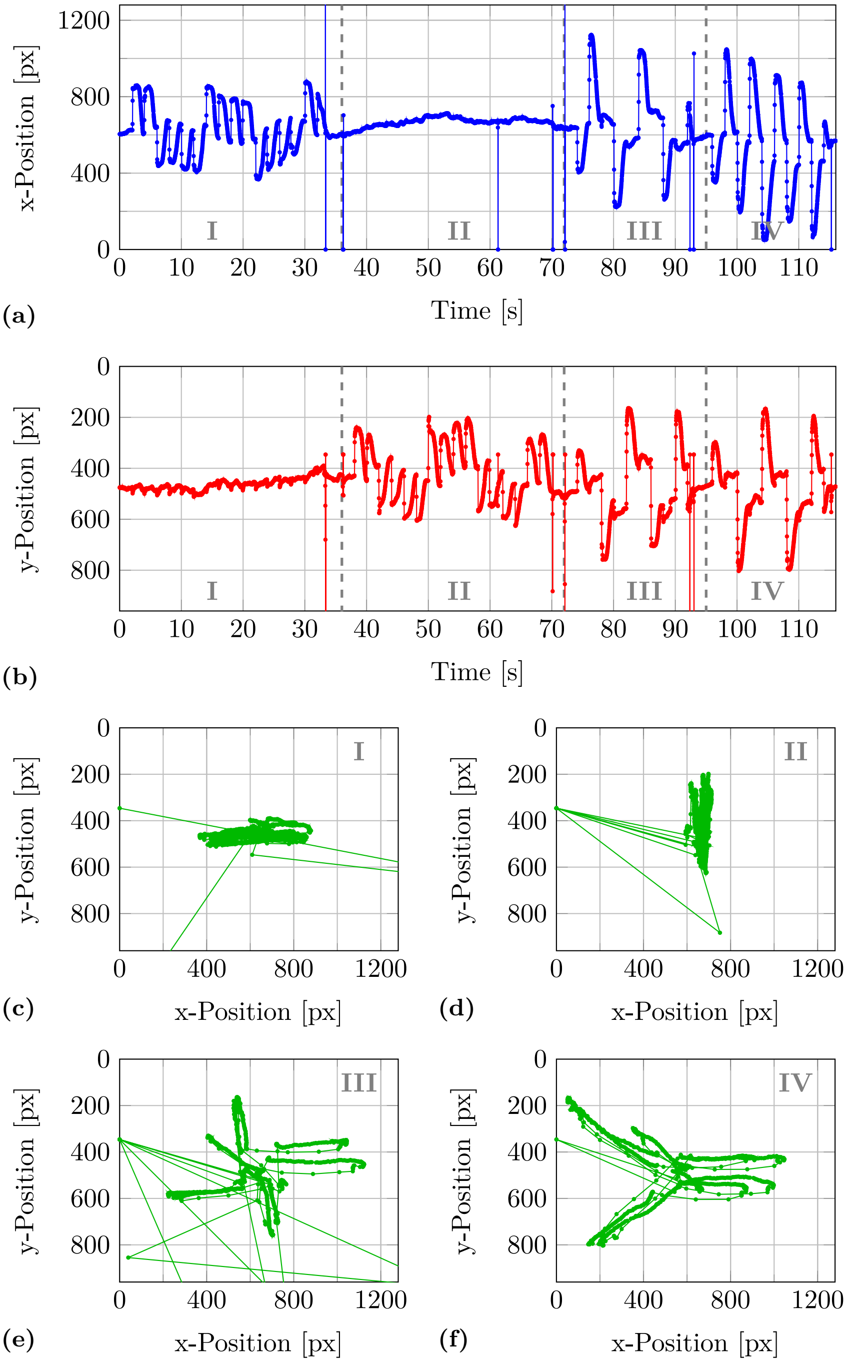


Figure 4.5: Eye positions over time (a)-(b) and in the spatial domain (c)-(f) reported by the eye-tracking glasses while performing the movement patterns I - IV, illustrated in Figure 3.8, involving a combination of eye and head movements.

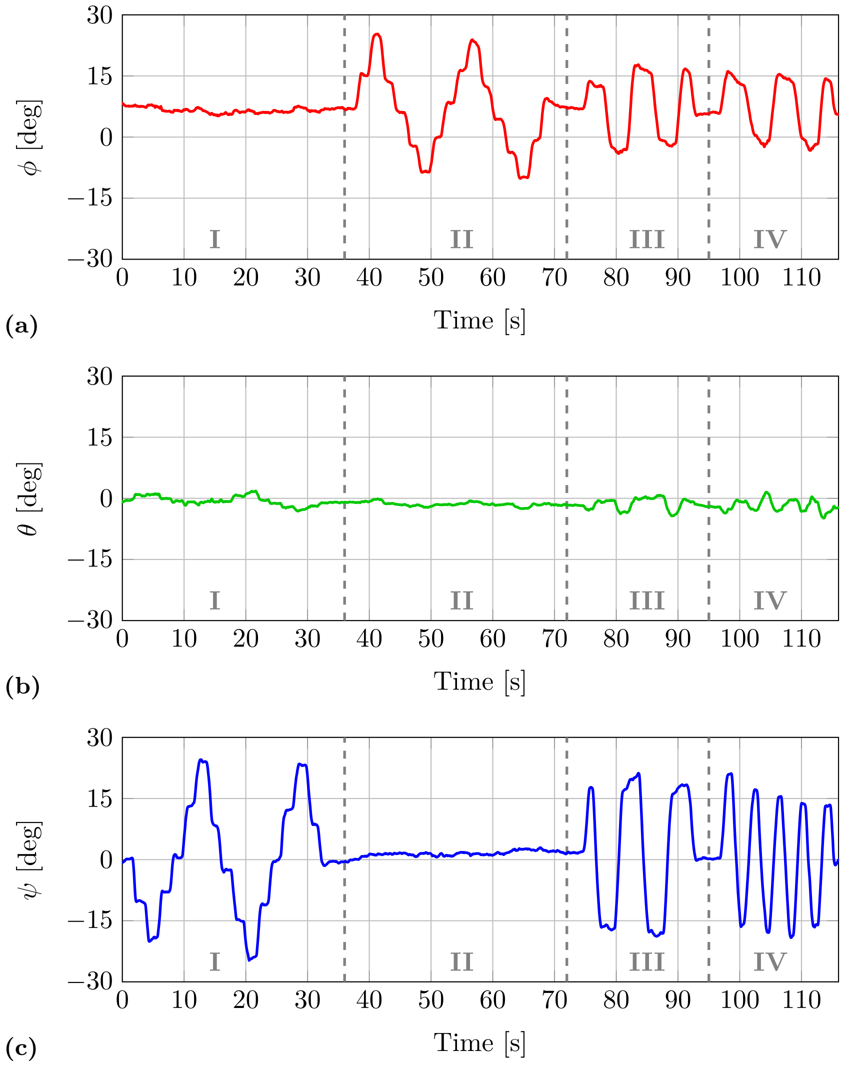


Figure 4.6: Head rotation angles ϕ (a), θ (b) and ψ (c) over time reported by the IMU the the head-movement patterns I - IV, illustrated in Figure 3.8, are preformed.

4.2 Model Implementation

4.2.1 Combination of Eye- and Head-Tracking Signals

Various synchronisation patterns, mainly sinusoidal patterns of different frequencies, were tested in both the horizontal and vertical directions. The best results were achieved using a sinusoidal pattern in the horizontal direction with a frequency of approximately 2 Hz. Figure 4.7 shows the results when such a pattern is performed by the head, while fixating at a single spot with the eyes. The pattern is visible both in the data of the eye-tracking glasses and the IMU. It can be clearly observed that the two signals are almost identical in shape and that they can be synchronised almost perfectly by maximising the cross-correlation between them.

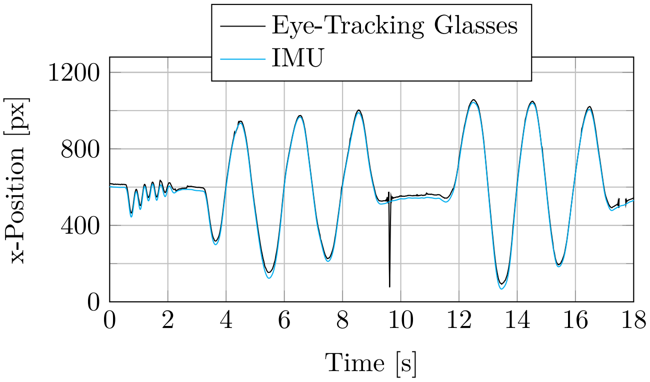


Figure 4.7: Synchronisation of the eye-tracking data and the IMU data using a sinusoidal VOR synchronisation pattern.

After calibration and synchronisation have successfully taken place, the eye- and head-tracking signals are combined by applying Equations (3.3) and (3.11) in order to calculate the eye-in-space motion. Figure 4.8 depicts the results of combining the eye-tracking signals, illustrated in Figure 4.4, with the head-tracking signals, illustrated in Figure 4.6. These eye- and head-tracking signals were recorded while the movement patterns I - IV, illustrated in Figure 3.8, were performed with the head while fixating a single spot with the eyes. Figure 4.8 shows the horizontal and vertical positions of the resulting eye-in-space motion over time as well as in the spatial domain for the movement patterns I - IV, respectively. Through observation of the different plots, is can be assumed that the test person was fixating a spot

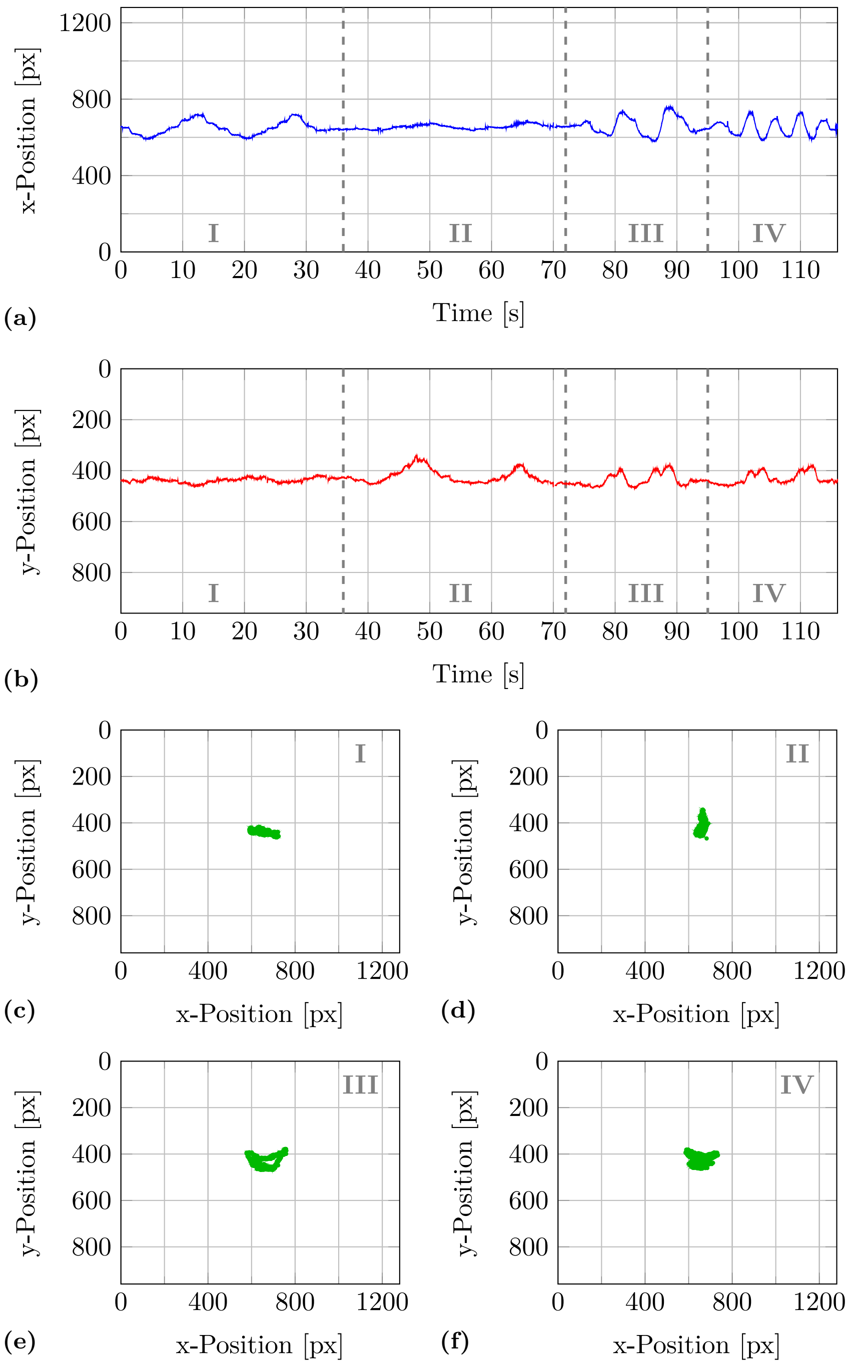


Figure 4.8: Eye-in-space positions over time (a)-(b) and in the spatial domain (c)-(f) derived with the initial model while performing the movement patterns I - IV with the head, illustrated in Figure 3.8, and fixating a single spot with the eyes.

with the eyes, approximately at the centre of the common coordinate system, i.e., at position $(x_G, y_G) = (640, 480)$. However, there is still quite a high degree of dispersion of roughly ±100 px in the horizontal and ±50 px in the vertical direction. The origin of this dispersion is not obvious. It is certainly influenced by the accuracy and precision of the two tracking systems and it is also highly dependent on the ability of individual participants to fixate a stationary target reliably. Primarily, however, it reflects the quality of the applied model for combining the head- and eye-tracking signals.

4.2.2 Compensatory Factors

Section 3.5.3 discussed the compensatory factors A and B, which are formulated in Equation (3.12). These factors are tuned to minimise the standard deviation of the different signals recorded while performing the movement patterns I - IV with the head, illustrated in Figure 3.8, and fixating a single spot with the eyes. The result of one recording, whereby different values are applied for A and B, are depicted in Figure 4.9. The figure shows the horizontal and vertical positions of the resulting eye-in-space motion over time for the movement patterns I - IV. If values are chosen for the compensatory factors A and B so that they are both equal to one, this is the same as applying the initial model, without compensatory factors. Choosing values for A and B which are not equal to one, results in a clearly visible improvement in performance. It can be observed that the optimum values for the compensatory factors A and B must lie somewhere between 1.1 and 1.2. This is evident because when values for A and B between 1.1. and 1.2 are chosen, the degree of dispersion of the eye-in-space motion is much lower than when larger or smaller values are selected. Compensatory factors of $A = 1.15$ and $B = 1.19$ were found to produce the best results.

Figure 4.10 illustrates the results of applying the adjusted model, i.e., Equations (3.3) and (3.12) with $A = 1.15$ and $B = 1.19$, to the combination of the eye- and head-tracking signals while performing the movement patterns I - IV with the head, illustrated in Figure 3.8, and fixating a single spot with the eyes. The figure shows the horizontal and vertical positions of the resulting eye-in-space motion over time as well as in the spatial domain for the movement patterns I - IV, respectively. By comparing the results of the initial model, depicted in Figure 4.8, with the results of the adjusted model, illustrated in Figure 4.10, a clearly improved performance can be observed. The dispersion resulting form the initial model of roughly ±100 px in the

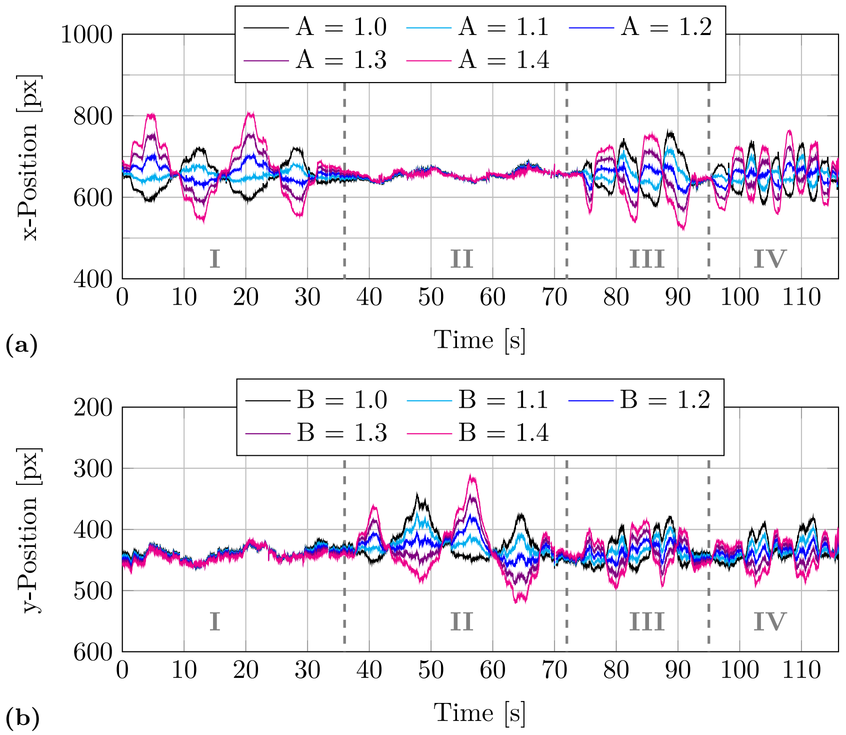


Figure 4.9: Horizontal (a) and vertical (b) eye-in-space positions determined with the adjusted model for different values of the compensatory factors A and B while performing the movement patterns I - IV with the head, illustrated in Figure 3.8, and fixating a single spot with the eyes.

horizontal and ±50 px in the vertical direction is reduced to approximately ±30 px in both directions. The presumption that the test person was fixating a spot with the eyes at approximately the centre of the common coordinate system is now even more valid. It would not have been possible to make such a presumption, however, by only investigating the signals of the eye-tracking glasses (cf. Figure 4.4).

Figure 4.11 shows the result of applying the adjusted model to the combination of the eye- and head-tracking signals, when the movement patterns I - IV, illustrated in Figure 3.8, are performed, which involve both eye and head movements. The figure shows the horizontal and vertical positions of the resulting eye-in-space motion over time as well as in the spatial domain for the different movement patterns I - IV, respectively. It is easy to observe

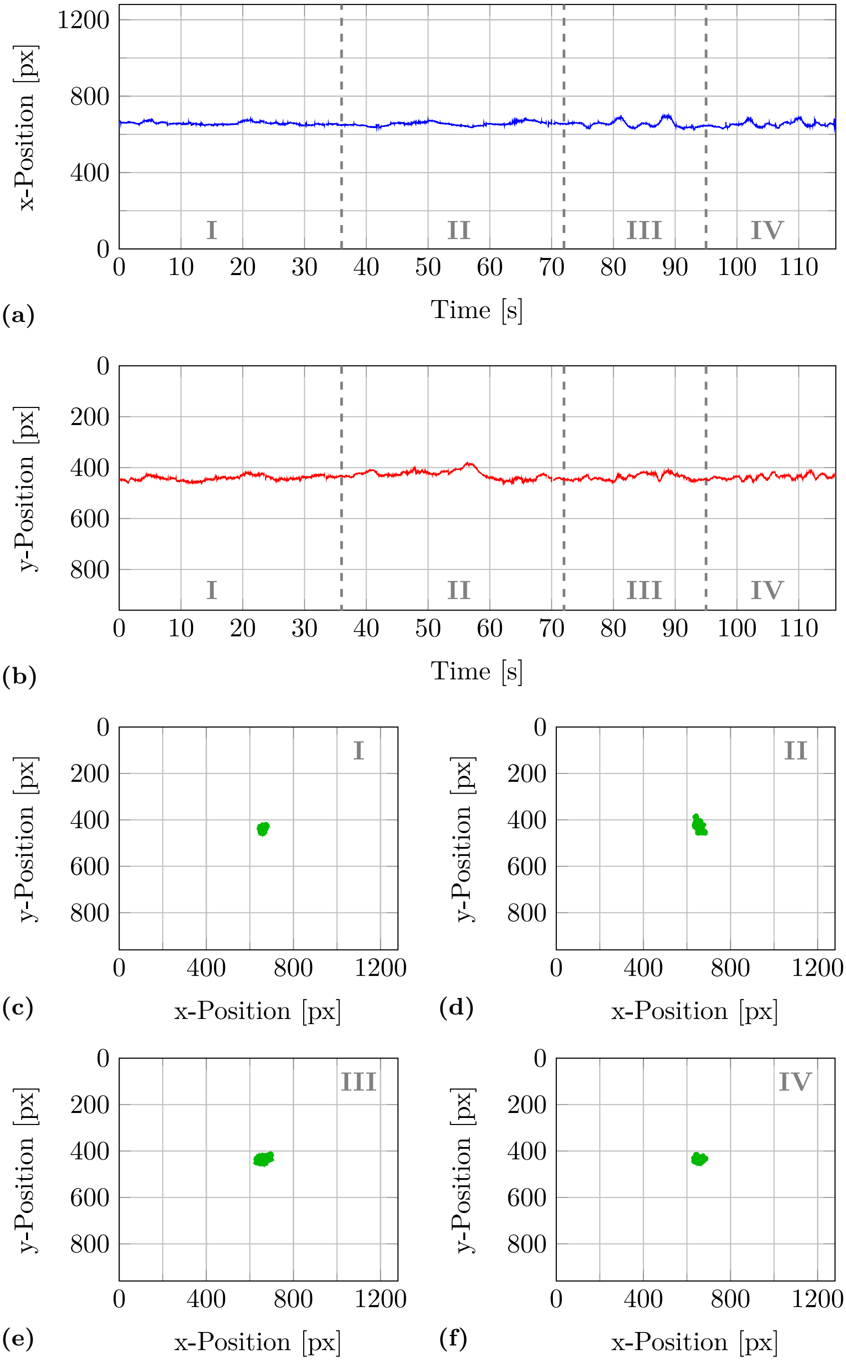
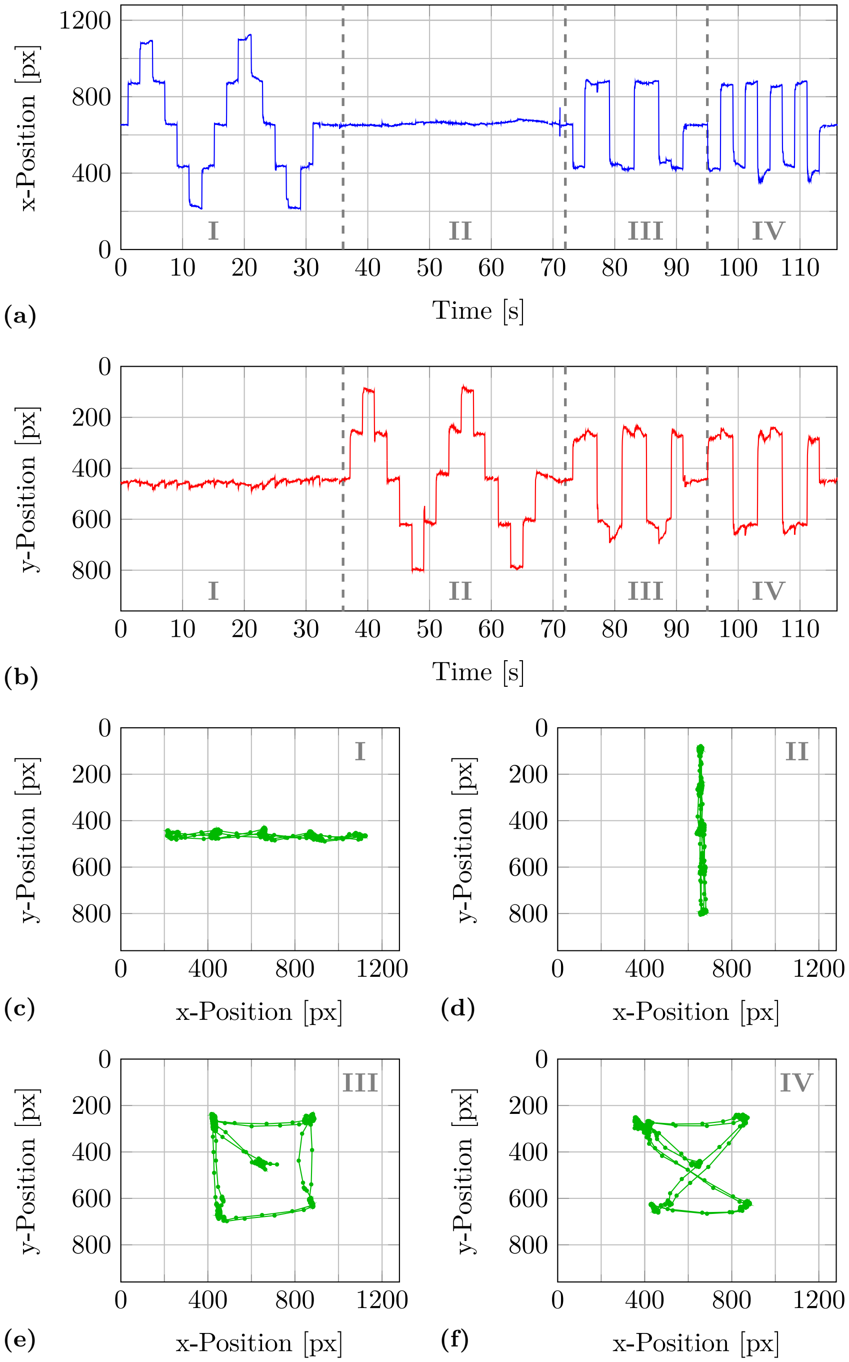


Figure 4.10: Eye-in-space positions over time (a)-(b) and in the spatial domain (c)-(f) derived with the adjusted model, while performing the movement patterns I - IV with the head, illustrated in Figure 3.8, and fixating a single spot with the eyes.



(a)

(b)

(c)

(d)

(e)

(f)

Figure 4.11: Eye-in-space positions over time (a)-(b) and in the spatial domain (c)-(f) derived with the adjusted model, while performing the movement patterns I - IV, illustrated in Figure 3.8, involving a combination of eye and head movements.

that the plots showing the position in the spatial domain (Figure 4.11c-f), closely match the performed eye-movement patterns I - IV, illustrated in Figure 3.8. This means that while it was not possible to draw any conclusion on the eye-in-space motion by only investigating the recorded eye-tracking data for the same type of eye and head movements (cf. Figure 4.4), it is possible if the head movement is also taken into account. In particular, it is possible to distinguish between the different types of eye movements.

4.3 Evaluation

4.3.1 Precision

The calculated precision of the four different cases described in Section 3.6.2 is presented in Tables 4.1 and 4.2. Table 4.1 shows the result without compensation for head movements, whereas in Table 4.2, the result incorporating head-movement compensation is shown. To calculate the precision for cases A) and C), i.e., fixating a stationary target while keeping the head still, 3 different recorded sequences, each 116 s in length, were evaluated. To calculate the precision for cases B) and D), i.e., fixating a stationary target while moving the head, 18 different recording sequences, each 4 s in length, were evaluated.

Table 4.1: Horizontal and vertical precision calculated without compensating for head movements for recordings where a stationary target is fixated with the eyes, with and without head movement, respectively, i.e., cases A) and B) described in Section 3.6.2.

Precision	Horizontal (x)	Vertical (y)
Head still	0.45 °	0.44 °
Head moving	12.89 °	7.99 °

When the participant is asked to not move the head, the precision is reduced from around 0.5 ° without head-movement compensation to around 0.2 ° with head-movement compensation. When the participant is asked to intentionally move the head, the precision is reduced from 12.89 ° to 0.61 ° in the horizontal direction and from 7.99 ° to 0.85 ° in the vertical direction. These results clearly show that by compensating for head movements the precision can

Table 4.2: Horizontal and vertical precision calculated with head-movement compensation for recordings where a stationary target is fixated with the eyes, with and without head movement, respectively, i.e., cases C) and D) described in Section 3.6.2.

Precision	Horizontal (x)	Vertical (y)
Head still	0.16 °	0.20 °
Head moving	0.61 °	0.85 °

be significantly improved, even for data sequences where the test person is asked to keep the head as still as possible.

4.3.2 Accuracy

The accuracy of the gaze estimation calculated in two different ways as described in Section 3.6.2 is presented in Tables 4.3 - 4.5. Table 4.3 shows the accuracy of the data recorded when all nine different stimuli videos are presented. Table 4.4 shows the accuracy when only the fixational movements (stimuli videos 1 - 5) are shown. Finally, Table 4.5 shows the result when mostly smooth-pursuit movements (stimuli videos 6 - 9) are shown. Two different recording sequences were evaluated for the presentation of each stimulus video.

Table 4.3: Horizontal and vertical accuracy calculated in two different ways as described in Section 3.6.2 from data recorded while presenting stimuli videos 1 - 9.

Accuracy	Horizontal (x)	Vertical (y)
$\tilde{A}$	1.49 °	1.64 °
$\hat{A}$	0.99 °	1.24 °

Over all recordings, accuracy values of 1.49 ° in the horizontal and 1.64 ° in the vertical direction are achieved if only the intersaccadic intervals are taken into account. The corresponding accuracy values are 0.99 ° and 1.24 ° for the horizontal and vertical directions, respectively, when an additional time shift between the stimulus signal and the estimated signal is introduced

Table 4.4: Horizontal and vertical accuracy calculated in two different ways as described in Section 3.6.2 from data recorded while presenting stimuli videos 1 - 5.

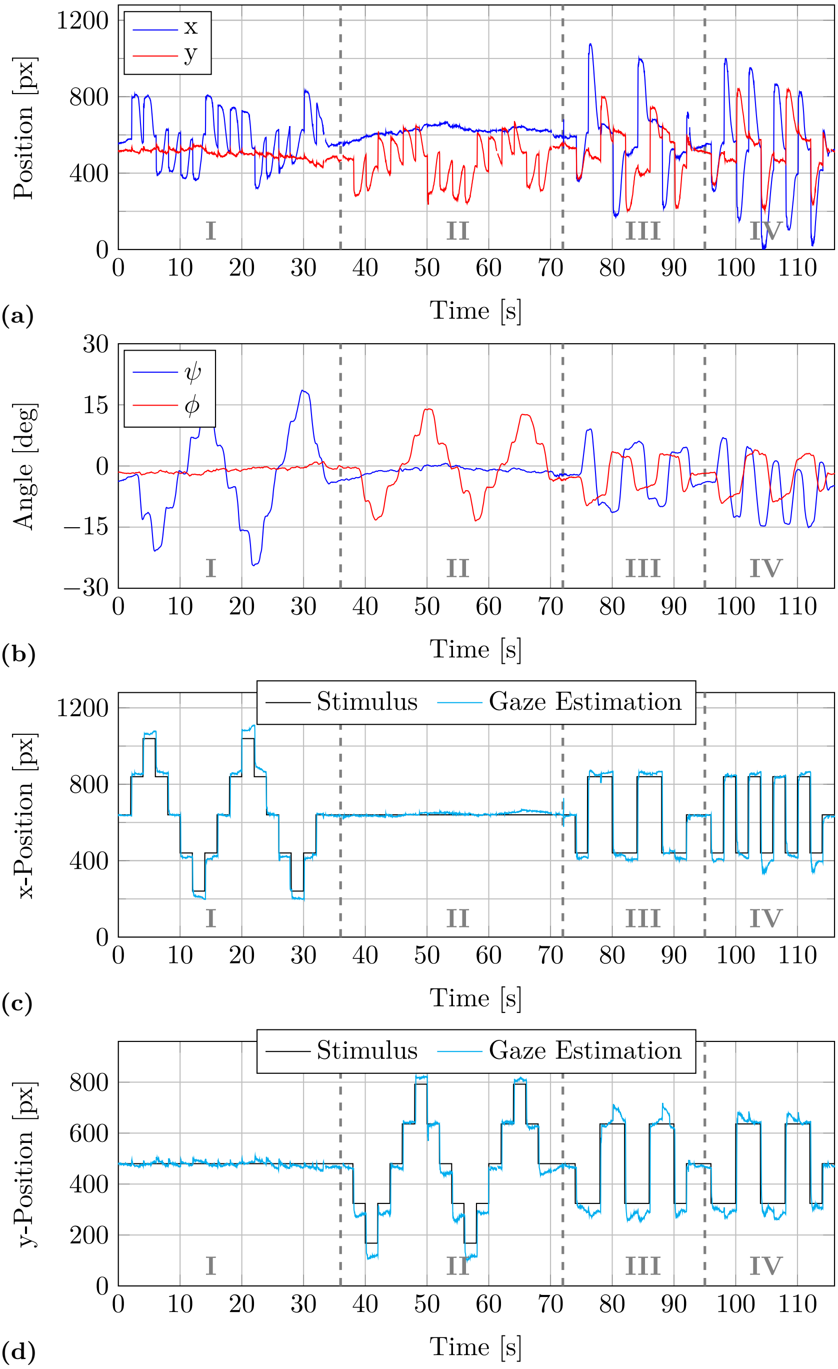
Accuracy	Horizontal (x)	Vertical (y)
$\tilde{A}$	1.25 °	1.66 °
$\hat{A}$	1.19 °	1.64 °

Table 4.5: Horizontal and vertical accuracy calculated in two different ways as described in Section 3.6.2 from data recorded while presenting stimuli videos 6 - 9.

Accuracy	Horizontal (x)	Vertical (y)
$\tilde{A}$	1.79 °	1.61 °
$\hat{A}$	0.74 °	0.75 °

into the calculation (cf. Section 3.6.2). By comparing Tables 4.4 and 4.5, it can be observed that the introduction of the additional time shift into the accuracy calculation barely influences the results of the fixational movements (stimuli videos 1 - 5). For the smooth-pursuit movements (stimuli videos 6 - 9), however, the introduction of the additional time shift improves the accuracy values by almost 1 °.

Figures 4.12 - 4.14 depict representative results of the stimuli videos 3, 7, and 9, whereby the movement patterns I - X were performed with a combination of eye and head movements. The figures show the horizontal and vertical eye positions and head-rotation angles over time reported by the eye-tracking glasses and the IMU, respectively. Moreover, to facilitate comparison with the estimated gaze positions, the horizontal and vertical positions of the stimuli are also shown. They were used to calculate the accuracy of the gaze estimation. The quality of the gaze estimation was similar for all recordings, independent of the type of stimulus video. Horizontal gaze estimation, however, was generally slightly better than vertical gaze estimation.



(a)

(b)

(c)

(d)

Figure 4.12: Results for stimulus video 3. (a)-(b) horizontal and vertical eye-in-head positions and head-in-space orientations, respectively. (c)-(d) stimuli and gaze positions in horizontal and vertical directions, respectively.

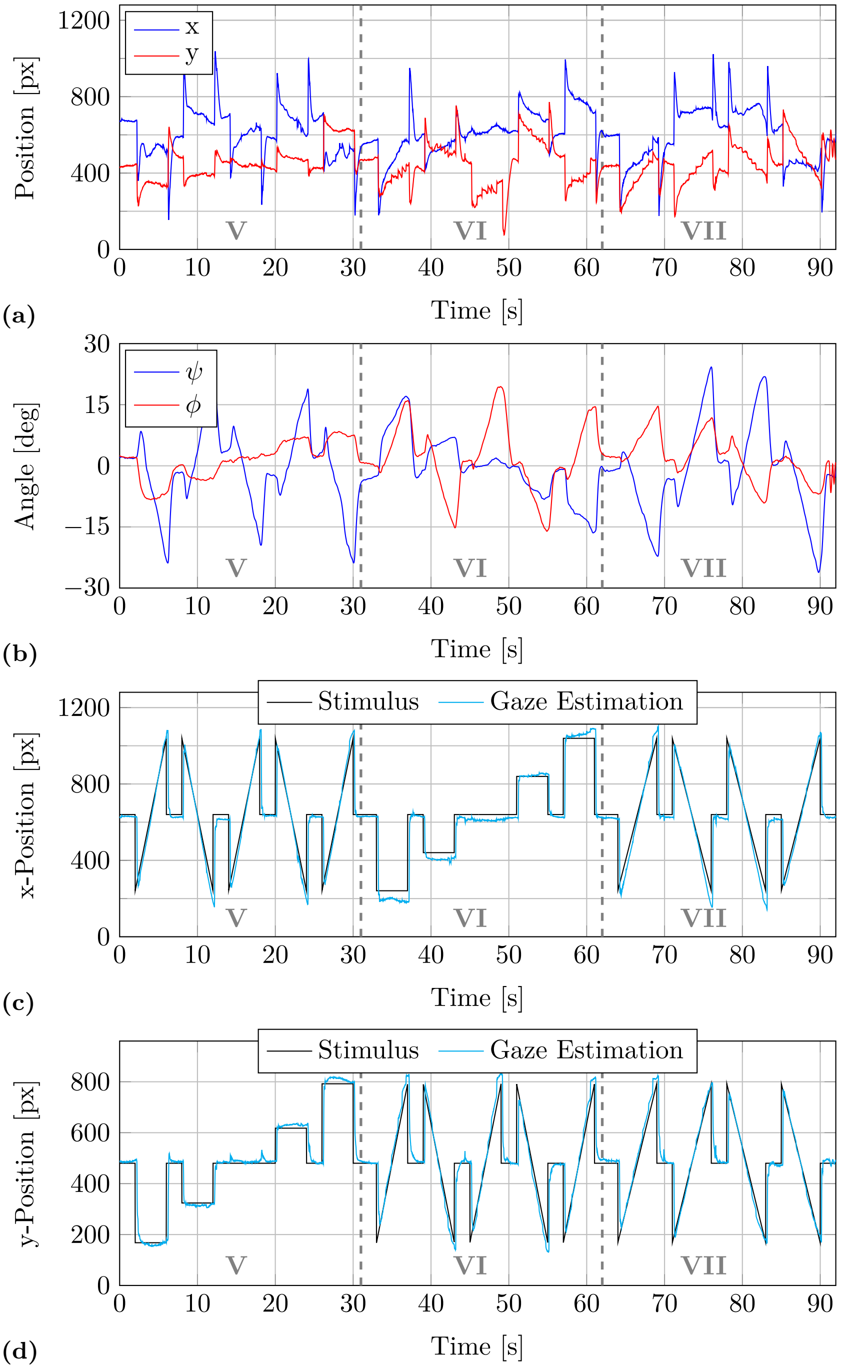


Figure 4.13: Results for stimulus video 7. (a)-(b) horizontal and vertical eye-in-head positions and head-in-space orientations, respectively. (c)-(d) stimuli and gaze positions in horizontal and vertical directions, respectively.

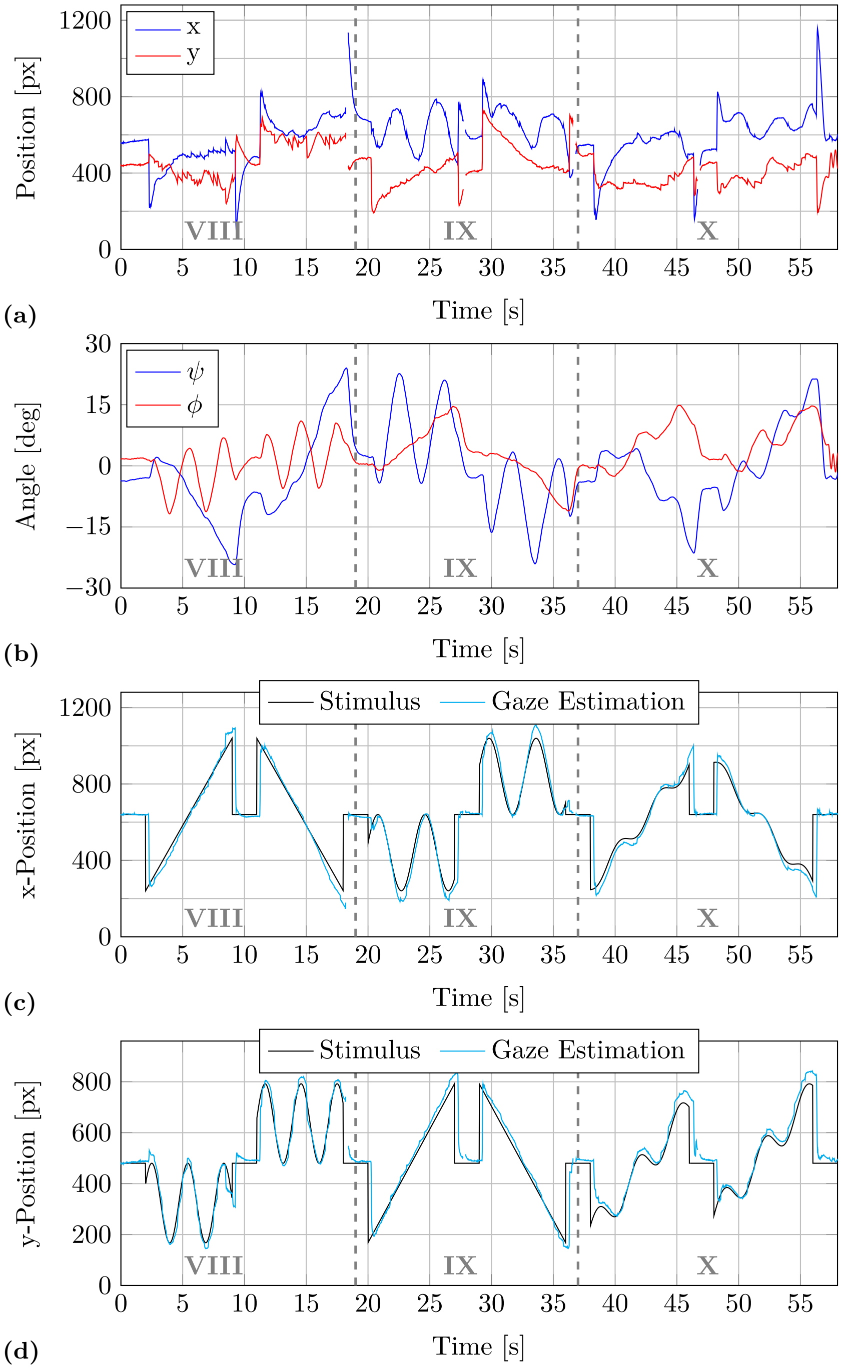


Figure 4.14: Results for stimulus video 9. (a)-(b) horizontal and vertical eye-in-head positions and head-in-space orientations, respectively. (c)-(d) stimuli and gaze positions in horizontal and vertical directions, respectively.

B - Event Detection

This second part of the results chapter presents the outcomes of the *Event Detection* part of Chapter 3. The evaluation of the different stages of the event-detection algorithm, i.e., the saccade detection and the classification of fixations and smooth pursuits, is presented in Sections 4.4 and 4.5, respectively, and in Section 4.6, the detection performance of the algorithm is evaluated and compared to the two alternative algorithms described in Section 3.8.3.

4.4 Saccade Detection

The parameter settings for the saccade-detection stage for both the proposed algorithm and the I-VDT algorithm are presented in Table 4.6. The proposed algorithm uses a two-step saccade-detection procedure with three different velocity thresholds, one for the detection of the approximate saccadic intervals and two for the saccadic onset and offset detection. In contrast, the I-VDT algorithm only uses one threshold. The thresholds were found by maximising the balanced accuracy of the saccade detection $BACC_S$. They were adjusted using the developmental part of the database only.

Table 4.6: The parameter settings for the saccade-detection stage of the proposed algorithm and the I-VDT algorithm.

Algorithm	Parameter	Value
Proposed Algorithm	T_V	45 °
	$T_{V_{ON}}$	25 °
	$T_{V_{OFF}}$	30 °
I-VDT Algorithm	T_V	15 °

The resulting sensitivities and specificities for the saccade-detection stage of both algorithms are presented in Table 4.7. The proposed algorithm clearly outperforms the I-VDT algorithm in terms of both sensitivity and specificity with values of 99.41 % and 99.33 % compared to 86.92 % and 94.55 %, respectively. The lower level of sensitivity for the I-VDT algorithm indicates that too few samples are detected as saccades in comparison with the manual annotations. In contrast to the proposed algorithm, which adopts a two-step saccade-detection process, the algorithm was not primarily able to detect the onsets and offsets of the saccade correctly.

Table 4.7: The sensitivity, specificity and balanced accuracy of the saccade-detection stage for the proposed algorithm and the I-VDT algorithm.

Algorithm	$SENS_S$	$SPEC_S$	$BACC_S$
Proposed Algorithm	99.41	99.30	99.35
I-VDT Algorithm	86.92	94.55	90.74

4.5 Fixation and SP Classification

4.5.1 Signal Measures

In order to classify the remaining samples into fixations and smooth-pursuit movements, the eight different signal measures described in Section 3.7.3, i.e., mean velocity, slope, integral, energy, dispersion, directional variation, consistency in direction, and positional displacement, were applied individually and their relative classification performance was compared. For the comparison of the classification performance of the different measures, the balanced accuracies for fixations and smooth-pursuit movements were calculated. The performances of the different signal measures were compared for the different sliding-windows approaches and corresponding versions, i.e., BSW_F, BSW_B, $WASW_1$, $WASW_2$, $WASW_3$, and $WASW_4$, using different window sizes. The different thresholds T_M were again optimised by maximising the balanced accuracies for fixations $BACC_F$ and smooth pursuits $BACC_P$, using the developmental part of the database only. Representative results of the comparison using the $WASW$ approach with a constant weighting vector and with a window size l_w of 200 ms in length, which corresponds to 12 samples, are listed in Table 4.8 and illustrated in Figure 4.15.

Fixation and smooth pursuit are classified very well in the case of seven out of the eight signal measures, with balanced accuracies between 90 % and 94 %. When the mean velocity was used as signal measure, however, it resulted in lower balanced accuracies of 85.06 % and 84.13 % for fixations and smooth pursuits, respectively. This means that it is very difficult to

Table 4.8: Comparison of the balanced accuracies for fixation and smooth-pursuit classification for the eight different signal measures described in Section 3.7.3 applied in a $WASW$ approach with a constant weighting vector and a window size l_w of 200 ms in length.

Signal Measures		$BACC_F$ [%]	$BACC_P$ [%]
PD	Positional Displacement	93.94	93.21
I	Integral	93.15	92.41
CD	Consistency in Direction	92.91	92.16
S	Slope	92.73	92.11
D	Dispersion	92.18	91.38
E	Energy	91.97	91.20
DV	Directional Variation	91.77	90.82
MV	Mean Velocity	85.06	84.13

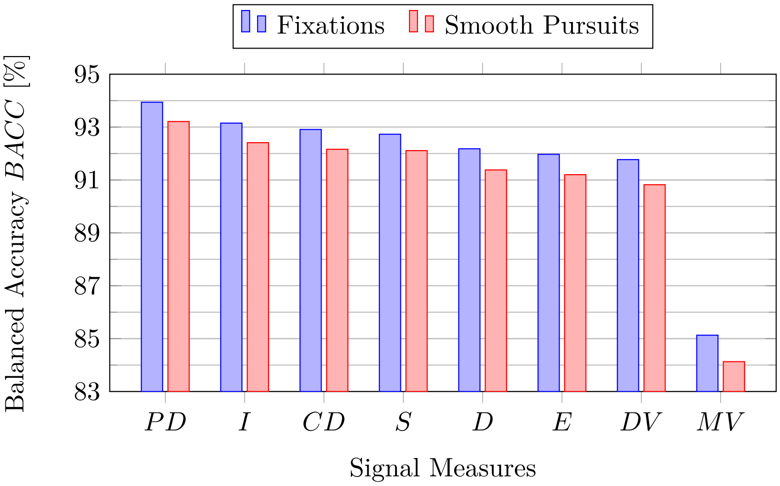


Figure 4.15: Visualisation comparing the balanced accuracies for fixation and smooth-pursuit classification for the eight different signal measures described in Section 3.7.3 applied in a $WASW$ approach with a constant weighting vector and a window size l_w of 200 ms in lenght.

separate fixations and smooth-pursuit movements using a velocity threshold, as was already mentioned in [39]. Among the other seven signal measures, the positional displacement generally showed the best results when different sliding-windows approaches and different window sizes were applied, followed by the integral measure. The performance using one of the other measures was quite similar.

4.5.2 Sliding-Window Approaches

In the previous section, the classification performances when applying the eight different signal measures were compared. The goal of this section is to compare the performances when applying the two different sliding-window approaches and corresponding versions, i.e., BSW_F, BSW_B, $WASW_1$, $WASW_2$, $WASW_3$, and $WASW_4$, to extract the signal measures and subsequently classify fixations and smooth-pursuit movements. To compare the classification performance of the different approaches and versions, the balanced accuracies for fixations and smooth-pursuit movements were calculated for different window sizes, different signal measures and different threshold values T_M. Representative results of the comparison using the positional displacement as the signal measure are depicted in Figures 4.16 and 4.17. Figure 4.16 shows the balanced accuracies for the fixation detection, whereas Figure 4.17 illustrates the balanced accuracies for the smooth-pursuit classification. The positional displacement was chosen as example because it is the measure which showed the best performance among all the signal measures (cf. Section 4.5.1).

The performance of the two different sliding-window approaches, BSW and $WASW$, in classifying fixations and smooth pursuits was similar. Performances with balanced accuracy values between 90 % and 98 % were achieved depending on which approach and signal measure were applied. For six out of the eight measures, i.e., mean velocity, integral, energy, dispersion, and positional displacement, the BSW approach showed slightly better results, whereas for the remaining two measures, i.e., directional variation and consistency in direction, the $WASW$ approach performed a little better. The former approach showed equally good results, regardless of whether it was performed in the forward (BSW_F) or backward (BSW_B) direction. The four versions of the latter approach, which involved using either a constant ($WASW_1$), triangular ($WASW_2$), exponential ($WASW_3$), or inverse-exponential ($WASW_4$) weighting vector, differed slightly in their

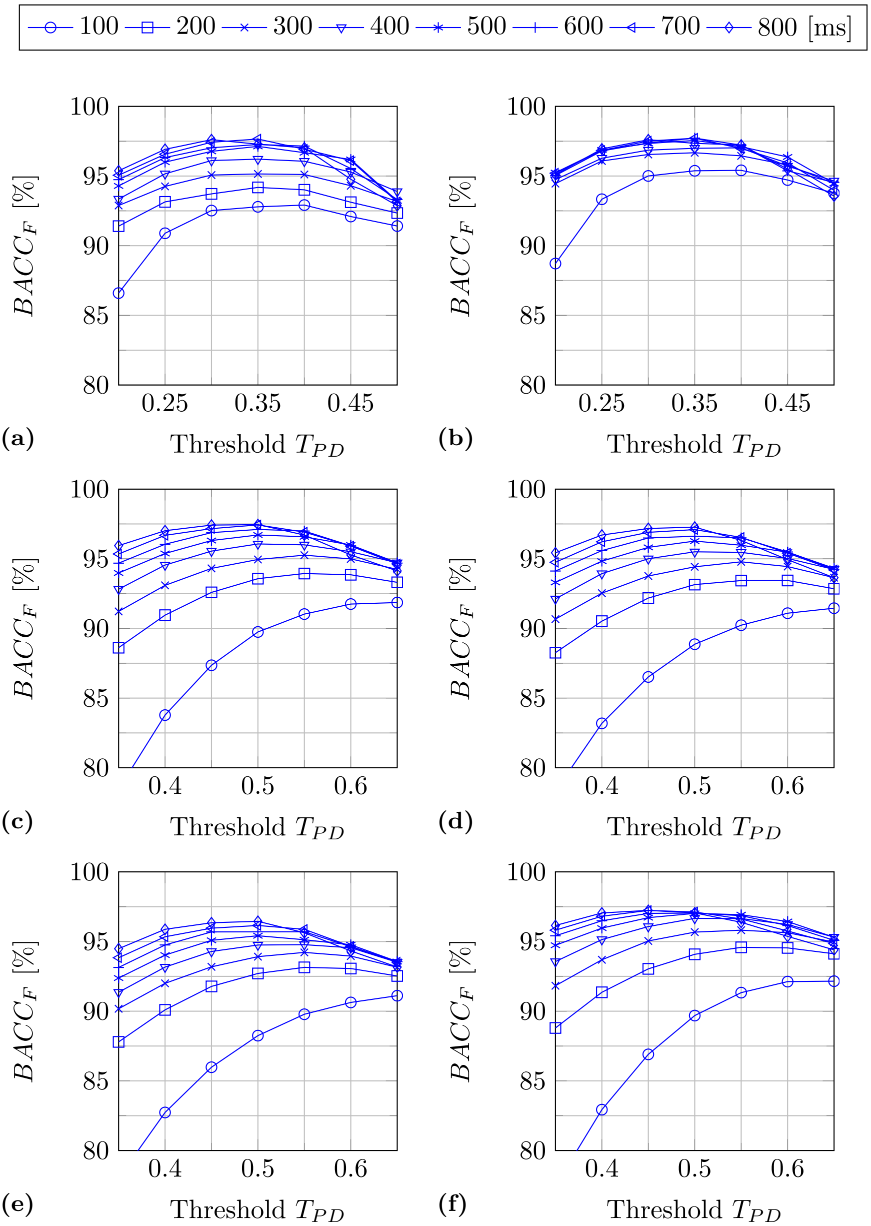


Figure 4.16: A comparison of the fixation classification performance of the different sliding-window approaches and versions described in Section 3.7.3, for different window sizes and threshold values using the positional displacement as the signal measure. (a) BSW_F, (b) BSW_B, (c) $WASW_1$, (d) $WASW_2$, (e) $WASW_3$, (f) $WASW_4$.

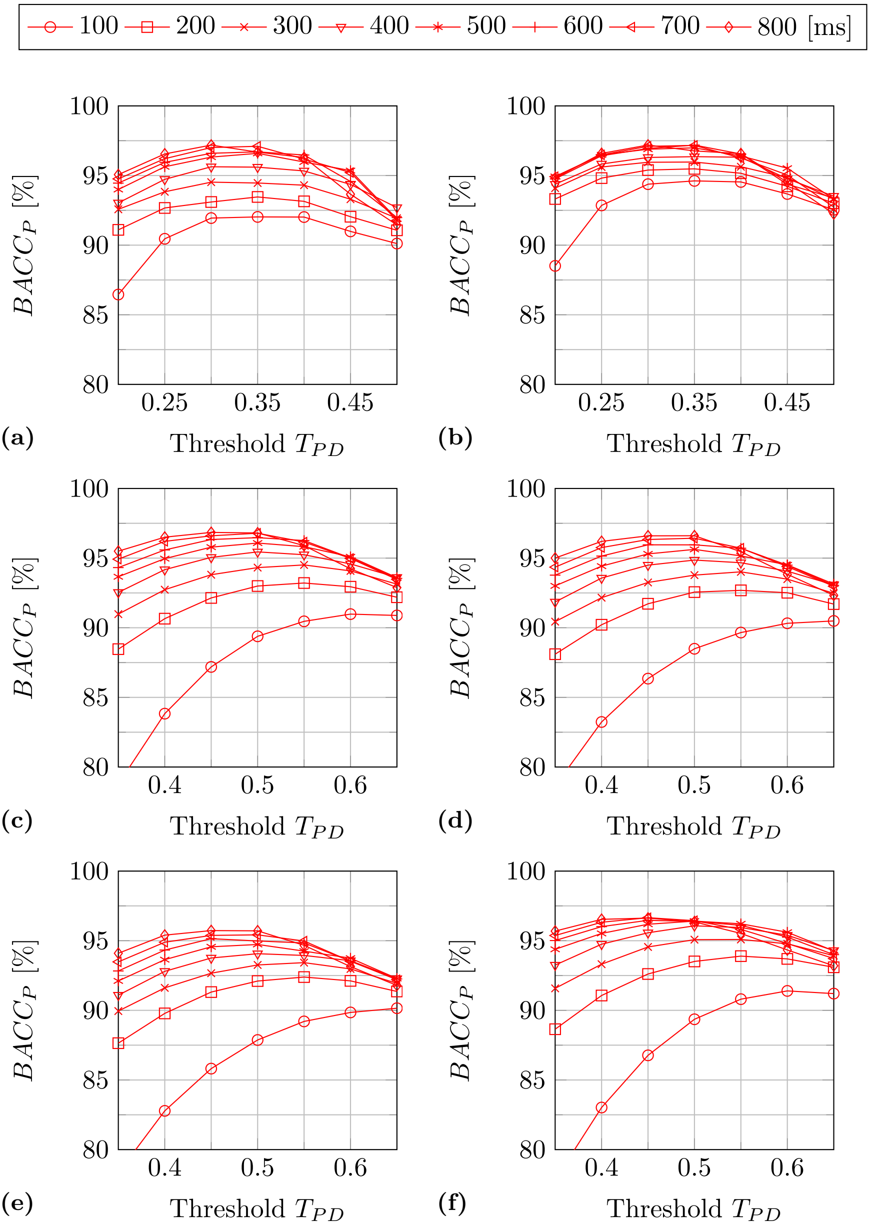


Figure 4.17: A comparison of the smooth-pursuit classification performance of the different sliding-window approaches and versions described in Section 3.7.3, for different window sizes and threshold values using the positional displacement as the signal measure. (a) BSW_F, (b) BSW_B, (c) $WASW_1$, (d) $WASW_2$, (e) $WASW_3$, (f) $WASW_4$.

performance. Applying an inverse-exponential or constant weighting vector showed slightly better results compared to using a triangular or exponential weighting vector. This means that it is advantageous to take a lot of information from distant neighbouring samples into account.

Besides the different sliding-window approaches and the respective versions thereof, the window size also has a big influence on the classification performance. Generally, it can be observed that the larger the window, the better the performance. This phenomenon was already reported in [42]. In some cases, however, a saturation effect can be observed, whereby the balanced accuracy cannot be improved any further by increasing the window size. One reason for the tendency of larger windows to result in better performances might be that the database only contains recordings where the intersaccadic intervals consist of one single type of eye movement, either fixations or smooth pursuits. In order to find the optimal window size, the algorithm is additionally applied to downsampled high-speed data. The data was recorded by a tower-mounted eye-tracking system and contains more than one type of eye movement within the intersaccadic intervals. Although it should be borne in mind that the signal properties of downsampled high-speed data are different from the characteristics of data recorded with a low-speed tracking system, this test might help to find a reasonable window size. The classification performance of the algorithm when it is applied to this alternative database also improves when the window size is increased. A saturation point is already reached when the window size rises to about 250 ms, however, depending on which signal measure is applied. In the example shown in Figures 4.16 and 4.17, in which the positional displacement is used as the signal measure, the BSW approach performed in the backward direction, depicted in subplot (b), exhibits the best performance for such a small window size of 250 ms.

4.5.3 Combination of Signal Measures

The goal of this section is to investigate whether the classification performance can be further improved by classifying the samples based on a combination of different signal measures and thresholds. For this purpose, the two signal measures which showed the best results for different sliding-window approaches and different window sizes are combined, namely the positional displacement and the integral. In order to combine these signal measures, the $WASW_1$ approach is applied, and a window size of 200 ms is used. For both signal measures, the average values $\overline{M}_{PD}(n)$ and $\overline{M}_{I}(n)$ are calculated for each sample as described in Section 3.7.3 and plotted against each other.

The results for the samples of the recordings using stimuli videos 1 - 9 are given in Figures 4.18 and 4.19. The samples are coloured blue if they were manually annotated as part of a fixation, or red if they were manually annotated as part of a smooth-pursuit movement. The respective thresholds T_{PD} and T_I are indicated as black lines when each signal measure is used separately for the classification. The two figures show that it is difficult to improve the classification by introducing new thresholds or other criteria to distinguish between fixational and smooth-pursuit samples. This is in evidence in the sample clusters, which are overlapping. In addition, other combinations of signal measures were tried for the classification, but the results were similar.

4.6 Evaluation

In this section, the proposed algorithm is evaluated in terms of sensitivity and specificity and compared to both the I-VDT algorithm and the event detector which is built-in to the eye-tracking glasses. The proposed algorithm applies the BSW_B approach and uses the positional displacement as the signal measure (cf. Section 3.7.3) to classify fixations and smooth pursuits. The approach and the signal measure are chosen because they showed the best results in the previous comparisons (cf. Section 4.5). The window size is set to 250 ms as determined in Section 4.5.2. The parameter settings for the different detection stages for both the proposed algorithm and the I-VDT algorithm are shown in Table 4.9. The parameters were optimised using the developmental part of the database. To facilitate a fair comparison, the parameters of the I-VDT algorithm were optimised in the same way. As previously mentioned in Section 3.8.3, there are no user-adjustable parameters for the built-in event-detection algorithm of the eye-tracking glasses.

The sensitivities, specificities and balanced accuracies of the detections of saccades, fixations, and smooth-pursuit movements for the three different algorithms are given in Tables 4.10 and 4.11, and they are illustrated in Figures 4.20 and 4.21. Table 4.10 and Figure 4.20 contain the results pertaining to the developmental part of the database, whereas Table 4.11 and Figure 4.21 show the results pertaining to the

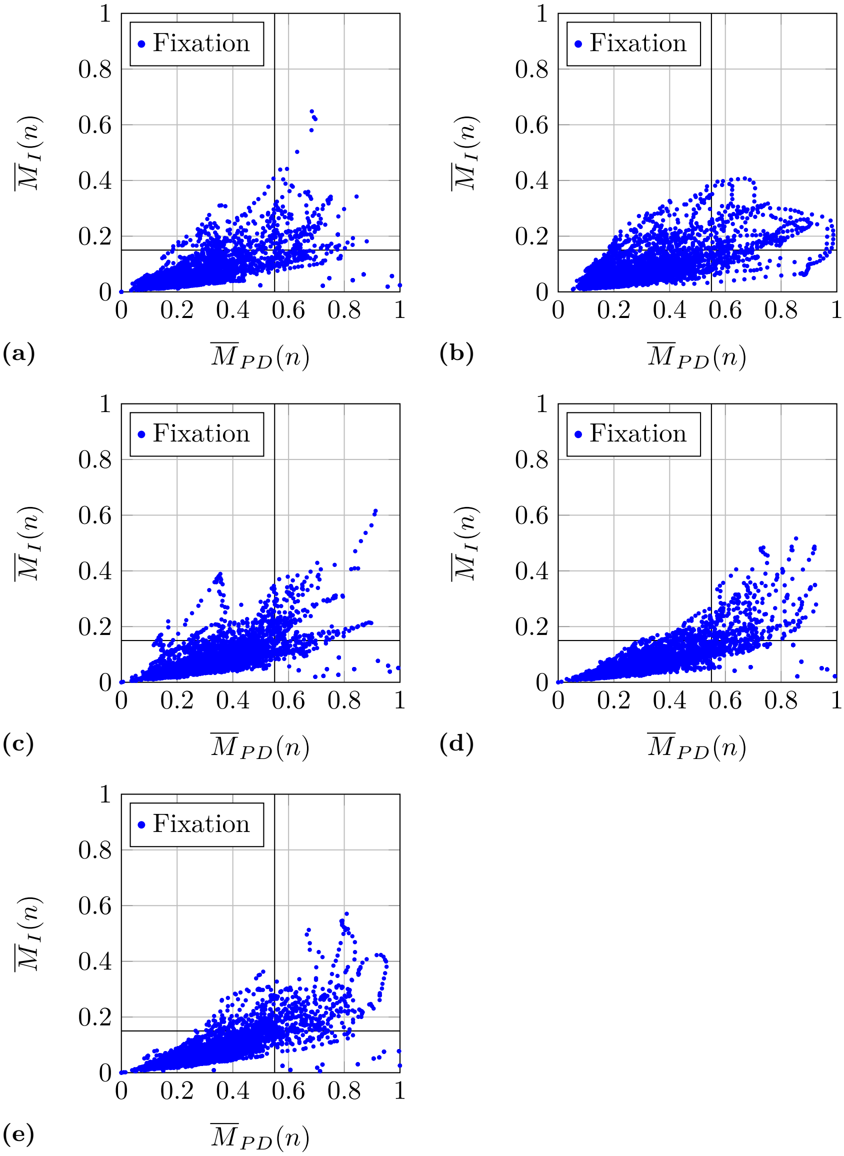
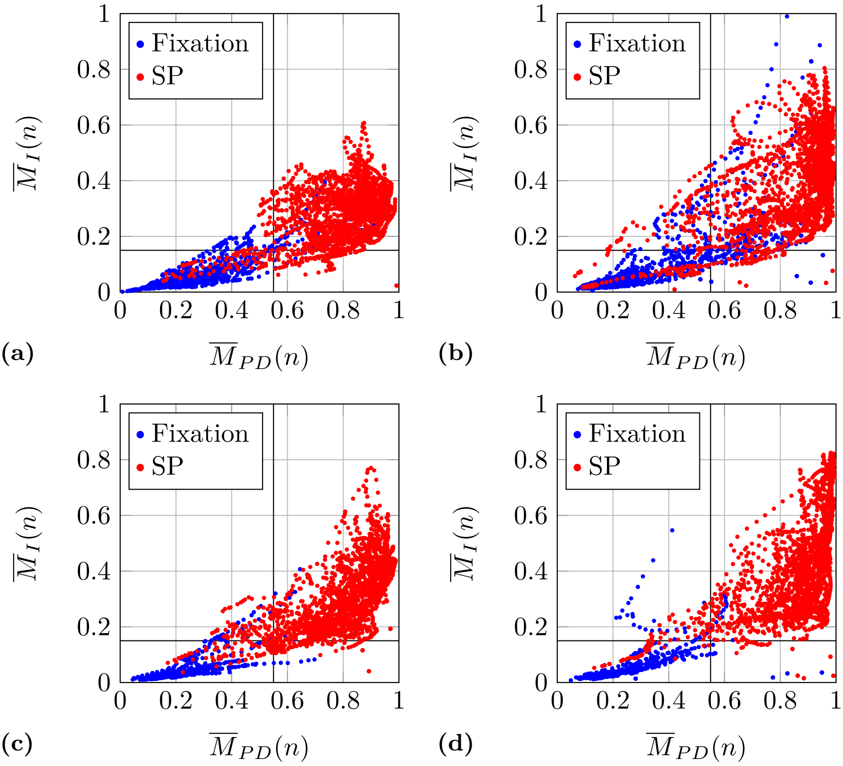


Figure 4.18: The sample average values of two signal measures (positional displacement $\overline{M}_{PD}(n)$ and integral $\overline{M}_I(n)$) plotted against each other for stimuli videos 1 - 5 (a)-(e). The samples are coloured according to the manual annotations as fixations (blue) or smooth-pursuit movements (red), and the respective thresholds T_{PD} and T_I are indicated as black lines when each signal measure is used separately for the classification.



(a)

(b)

(c)

(d)

Figure 4.19: The sample average values of two signal measures (positional displacement $\overline{M}_{PD}(n)$ and integral $\overline{M}_I(n)$) plotted against each other for stimuli videos 6 - 9 (a)-(d). The samples are coloured according to the manual annotations as fixations (blue) or smooth-pursuit movements (red), and the respective thresholds T_{PD} and T_I are indicated as black lines when each signal measure is used separately for the classification.

Table 4.9: The parameter settings for proposed algorithm and the I-VDT algorithm.

Algorithm	Parameter	Value
Proposed Algorithm	T_V	45 °
	$T_{V_{ON}}$	25 °
	$T_{V_{OFF}}$	30 °
	l_w	250 ms
	T_{PD}	0.305
I-VDT Algorithm	T_V	15 °
	l_w	250 ms
	T_D	0.675°

validation part of the database. The results of the developmental and validation part of the database are very similar, which implies that the different parameters were not overfitted to the data of the developmental part.

The performance of the proposed algorithm and the I-VDT algorithm in detecting saccades for was already discussed in Section 4.4. It was established that the proposed algorithm clearly outperforms the I-VDT algorithm. The algorithm which is built-in to the eye-tracking glasses performs slightly better than the I-VDT algorithm in terms of balanced accuracy, even though it has a lower specificity compared to the I-VDT algorithm. The performance is still poorer than that of the proposed algorithm, however, which exhibits values of over 99 % for both the developmental and validation part.

The sensitivities and specificities of the proposed algorithm are a bit lower when classifying fixations and smooth pursuits than when detecting saccades, but the values are very high nevertheless, ranging between 95 % and 97 %. The performance of the I-VDT algorithm is again poorer with values between 77 % and 93 %. The built-in algorithm clearly performs poorest, as it is not able to discriminate between fixations and smooth-pursuit movements (cf. Section 3.8.3). All events which are not detected as saccades are classified as fixations. Therefore, the sensitivity of the fixation detection is quite high, and the specificity is extremely low, meaning that far too

Table 4.10: Sensitivities, specificities and balanced accuracies for the proposed algorithm, the I-VDT algorithm, and the built-in event detector of the eye-tracking glasses, calculated for the developmental part of the database.

Measure	Proposed [%]	I-VDT [%]	Built-in [%]
$SENS_S$	99.41	86.92	92.23
$SPEC_S$	99.30	94.55	91.40
$BACC_S$	99.35	90.74	91.82
$SENS_F$	96.30	89.21	93.13
$SPEC_F$	97.13	90.52	25.08
$BACC_F$	96.72	89.87	59.11
$SENS_P$	95.31	78.92	0
$SPEC_P$	96.92	93.06	100
$BACC_P$	96.12	85.99	50

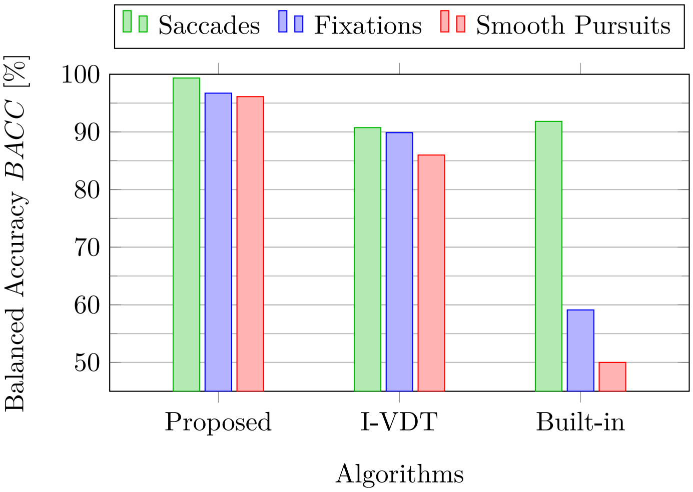


Figure 4.20: A visualisation of the balanced accuracy values for the proposed algorithm, the I-VDT algorithm, and the built-in event detector of the eye-tracking glasses, calculated for the developmental part of the database.

Table 4.11: Sensitivities, specificities and balanced accuracies for the proposed algorithm, the I-VDT algorithm, and the built-in event detector of the eye-tracking glasses, calculated for the validation part of the database.

Measure	Proposed [%]	I-VDT [%]	Built-in [%]
$SENS_S$	99.41	85.07	91.81
$SPEC_S$	99.52	94.54	90.83
$BACC_S$	99.46	89.81	91.32
$SENS_F$	96.49	89.85	92.76
$SPEC_F$	96.40	88.98	26.32
$BACC_F$	96.44	89.42	59.54
$SENS_P$	94.97	76.72	0
$SPEC_P$	96.77	93.51	100
$BACC_P$	95.97	85.11	50

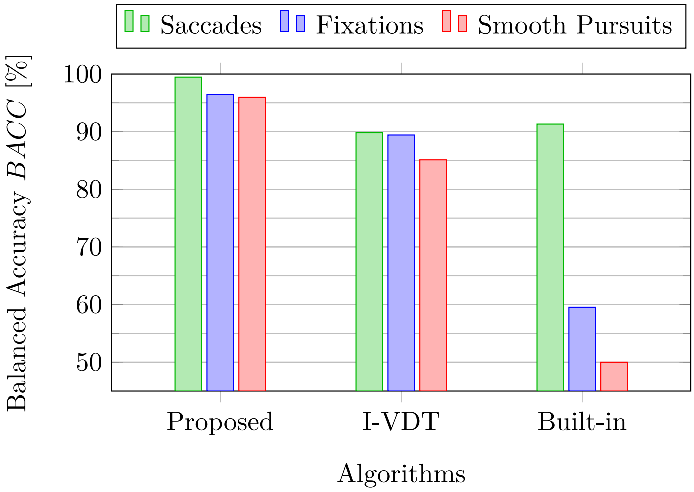


Figure 4.21: A visualisation of the balanced accuracy values for the proposed algorithm, the I-VDT algorithm, and the built-in event detector of the eye-tracking glasses, calculated for the validation part of the database.

many samples are detected as fixations compared to the manual annotations. The sensitivity and specificity of the smooth-pursuit detection are 0 % and 100 %, respectively, as no samples are labelled as smooth pursuits by the built-in detector.

Three examples illustrating the detection performance of the proposed algorithm for saccades, fixations, and smooth-pursuit movements based on the recordings using the stimuli videos 3, 7, and 9 are depicted in Figures 4.22 - 4.24, respectively. The figures show the labelled eye-in-space position for the horizontal and vertical directions over time as well as in the spatial domain.

In the case of stimulus video 3, the movement patterns I - IV are performed with a combination of eye and head movements. As described in Section 3.4, these movement patterns aim to stimulate sequences of saccadic and fixational eye movements only. Accordingly, saccades and fixations were almost exclusively detected by the proposed algorithm in Figure 4.22. However, a few samples were still classified as smooth-pursuit movements. These samples are mainly located at the beginning or the end of a saccade. One reason for this classification could be that the eye was possibly still drifting a bit and not entirely still. Some samples were also marked as smooth pursuits at parts of the signal where the accuracy of the head-movement compensation was low.

For the stimuli videos 7 and 9, the movement patterns V - VII and VIII - X are performed, respectively, with a combination of eye and head movements. These movement patterns aim to stimulate sequences of saccadic, fixational, and smooth-pursuit eye movements. In Figures 4.23 - 4.24, all three eye movements were detected by the proposed algorithm, even the catch-up saccades within the smooth-pursuit movements (cf. Section 2.2.3). Each intersaccadic interval should contain only one type of eye movement, either fixations or smooth pursuits. In this case again, it can be observed that at the beginning and the end of a saccade it is hard to discriminate between fixations and smooth pursuits.

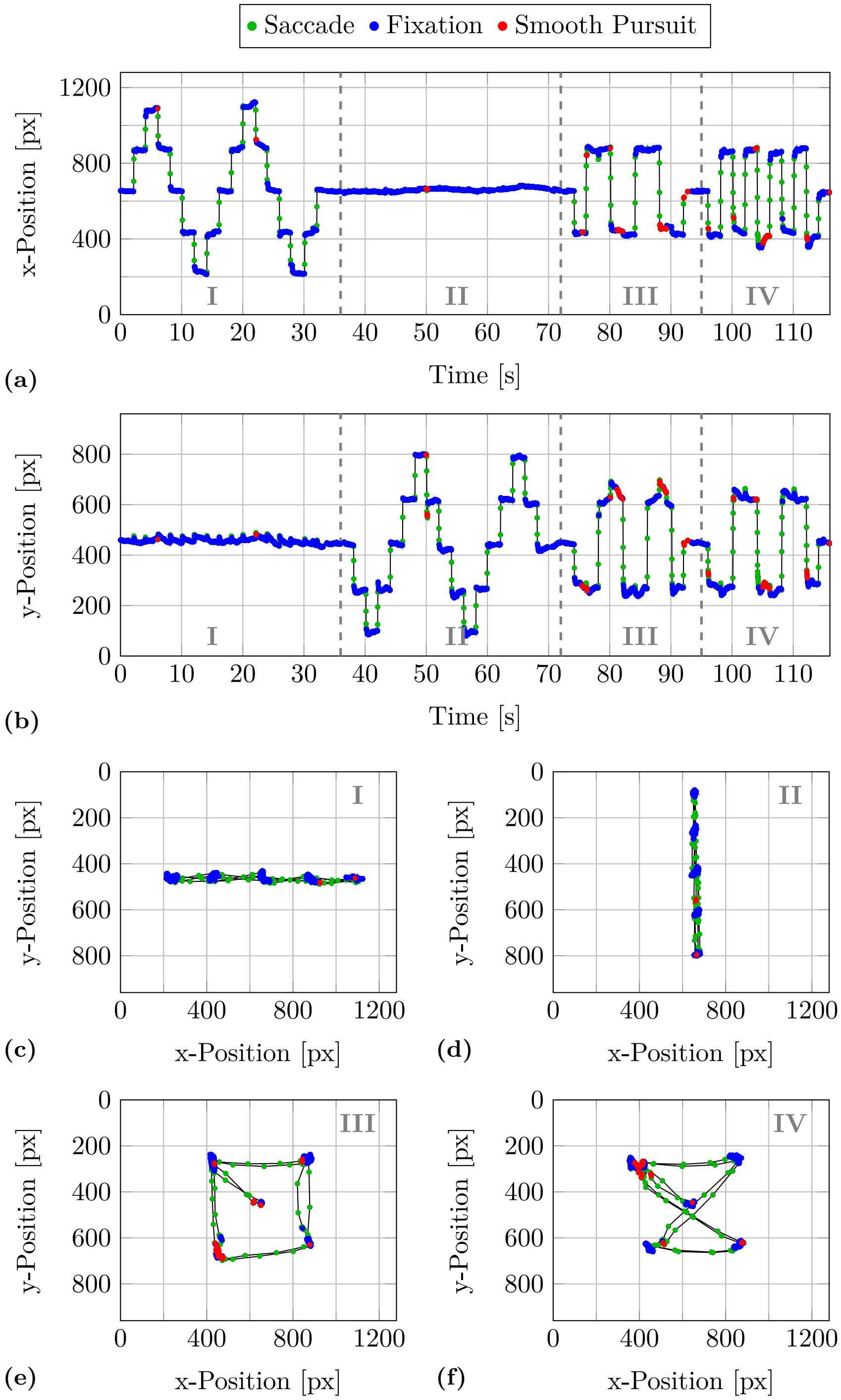


Figure 4.22: Event detection of proposed algorithm for stimulus video 3. Labelled eye-in-space positions over time (a)-(b) and in the spatial domain (c)-(f).

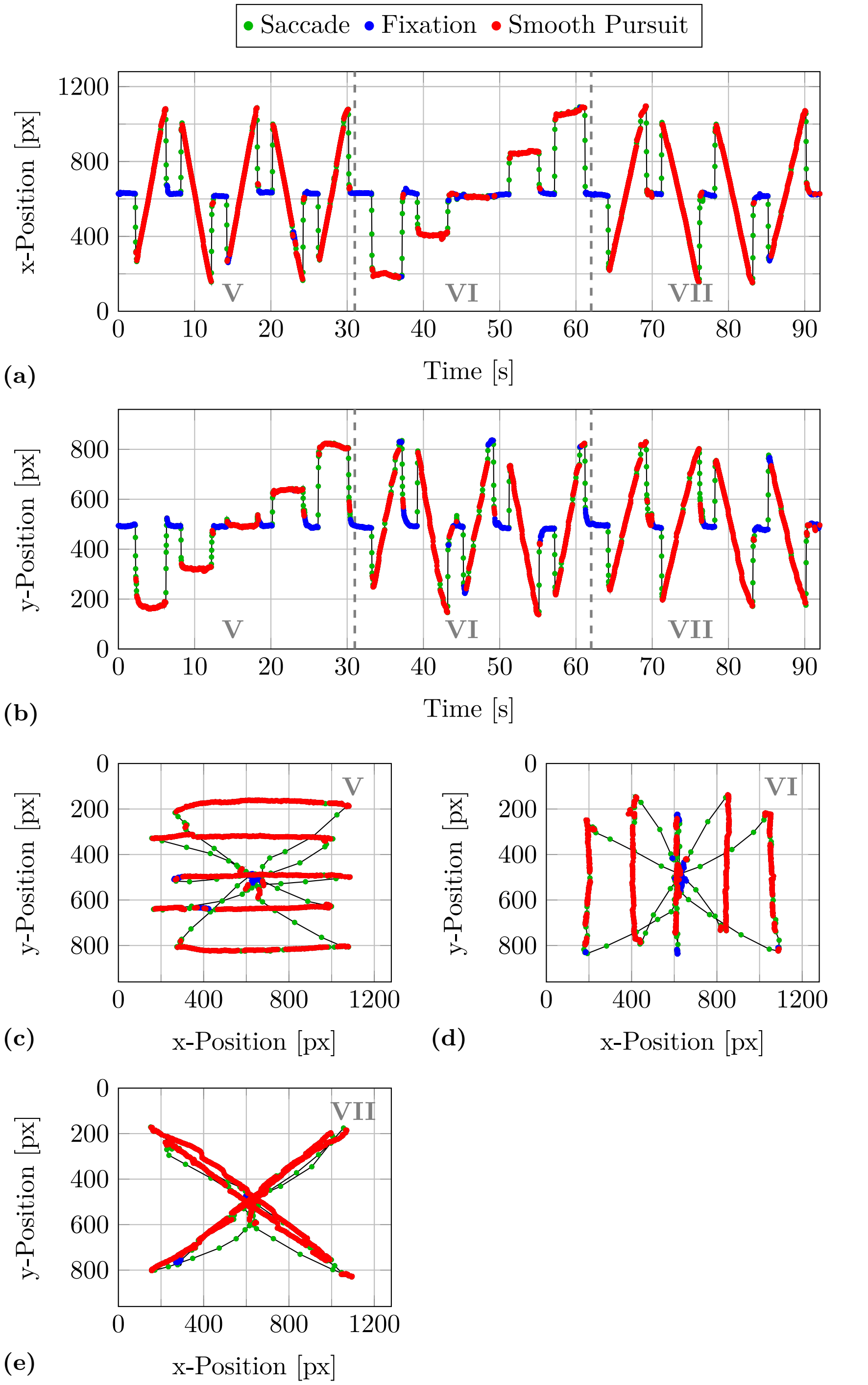


Figure 4.23: Event detection of proposed algorithm for stimulus video 7. Labelled eye-in-space positions over time (a)-(b) and in the spatial domain (c)-(e).

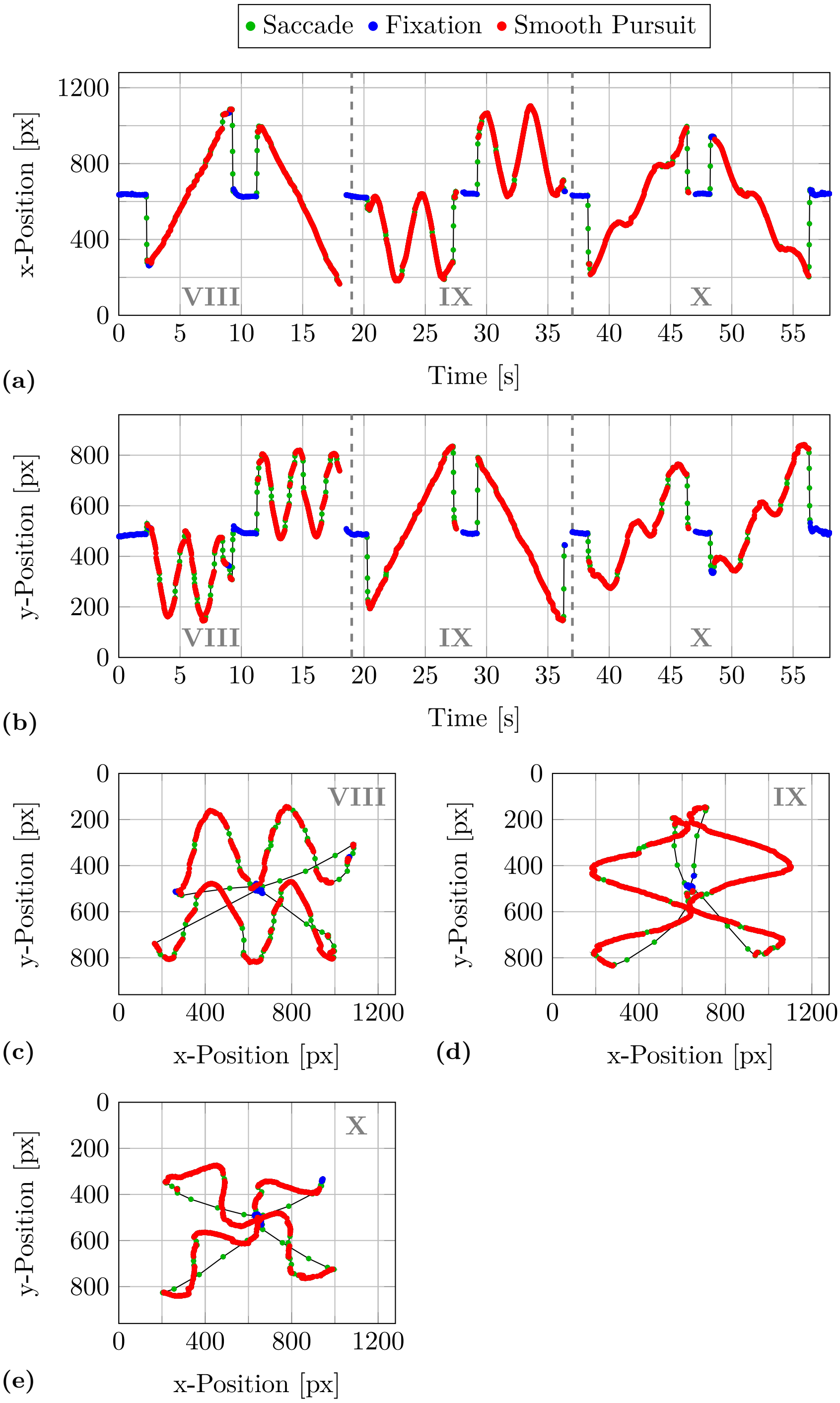


Figure 4.24: Event detection of proposed algorithm for stimulus video 9. Labelled eye-in-space positions over time (a)-(b) and in the spatial domain (c)-(e).

Chapter 5

Discussion

In this chapter, the results presented in Chapter 4 are discussed. An overview of the challenges and limitations of the gaze-estimation model and the proposed event-detection algorithm is given, and some suggestions for further improvements and extensions are presented and discussed.

5.1 Gaze Estimation

The method presented in this thesis estimates the eye-in-space motion by combining eye-tracking signals, recorded with eye-tracking glasses, and head-tracking signals, recorded using an IMU. The results of the performance evaluation of the model presented in Section 4.3 show that by compensating for head movements, the precision can be significantly improved and a considerably high degree of overall accuracy is achieved.

In [11], different remote and tower-mounted eye trackers are investigated and precision values from $0.01\,^{\circ}$ up to $1\,^{\circ}$ are reported. The signals obtained by combining the eye- and head-tracking signals, which result in precision values of around $0.2\,^{\circ}$ to $0.8\,^{\circ}$, are comparable to signals recorded using a remote or tower-mounted eye-tracking system, therefore. This in turn implies that the application of head-movement compensation makes it easier to develop an algorithm capable of distinguishing between different types of eye movements.

Overall, the proposed gaze-estimation method achieved accuracy values between $1\,^{\circ}$ and $1.2\,^{\circ}$, which is relatively high for signals recorded using eye-tacking glasses. In the literature, measured accuracy values between

0.3 ° and 5 ° are reported, depending on the type of eye-tracking systems and the recording environment employed [11, 20, 15]. While tower-mounted eye trackers provide greater accuracy compared to remote and head-mounted systems, experiments in laboratory settings glean more accurate results than when recordings are made in naturalistic, real-life situations. However, it has to be borne in mind that accuracy is influenced by a number of different factors:

- *Participant Factors.* As previously mentioned in Section 3.6.2, comparing the estimated eye-in-space motion to the stimuli signals evaluates not only the performance of the method, but also the user's ability to follow the stimulus, both temporally and spatially. The temporal factor means that eye movement might be performed too fast, or too slow, too early, or too late. The spatial factor involves the ability of the participant to fixate exact positions on the stimulus screen. As the presented dots have a diameter of approximately 1 °, fixating the fringe of the dots instead of the centre might decrease the accuracy by 0.5 °, for instance.

- *Equipment Factors.* Besides the factors which are unique to each participant, the properties of the equipment also influence the accuracy of results. The more precise and accurate the measurements of the two tracking systems are, the more accurate the gaze estimation will be.

- *Calibration and Synchronisation.* In addition, accuracy highly dependent on the quality of the calibration and synchronisation procedure. During the calibration of the eye-tracking glasses with the IMU, a stationary target has to be fixated with the eyes while the head is kept as still as possible. During the synchronisation procedure, a stationary target has to be fixed with the eyes while a synchronisation pattern is performed with the head. The better the participant can follow the instructions, the better the two systems are calibrated and synchronised, and the better accuracy values are achieved.

- *Stimuli Conversion.* Finally, the quality of the mapping of the stimuli coordinates to the common coordinate systems also has an influence on the accuracy of the gaze estimation. During mapping, the geometry of the experiment setup is used, i.e., the screen dimensions, distance to the screen, and field of view of the camera. The following three factors may each impair the accuracy by 0.2 - 0.3 °: inexact measurements of the screen dimension of 1 cm, inexact specifications of the field of

> view of the camera angles specifications of 1 °, and changes of the distance to the screen by 10 cm, which may for instance happen when the participant is standing slightly askew.

A simple but effective way to improve the calibration and synchronisation of the eye-tracking glasses and the IMU would be to combine the two systems within a single system by integrating the IMU into the mobile eye-tracking device. In comparison with other head-movement estimation techniques, an IMU is small and light which would make integration with the glasses relatively easy.

Currently, the model for gaze estimation is able to accurately compensate for head movements in the eye-tracking data in a controlled environment, meaning that the data is gleaned from recordings of sequences of controlled saccadic, fixational and smooth-pursuit movements. Further analysis of the gaze estimation is needed in order to investigate how the method can be extended to more complex situations where the subject is allowed to freely move within the environment. This means that in addition to the horizontal and vertical head rotations, transitional and torsional head movements need to be compensated for, which implies that the IMU needs to track the head position as well as head orientation.

5.2 Event Detection

The proposed algorithm is able to robustly detect saccades, fixations and smooth-pursuit movements from signals containing eye-in-space motion only, which are obtained by combining the eye- and head-tracking signals. The results of the comparison of the algorithm presented in Section 4.6 show that the detection performance of the proposed algorithm is considerably better than either of of the two alternative algorithms, namely the I-VDT algorithm and the event detector built-in to the eye-tracking glasses.

The lower level of sensitivity of the saccade-detection stage exhibited by the I-VDT algorithm indicates that there are too few samples detected as saccades compared to the manual annotations. The I-VDT algorithm was not able to detect the onsets and offsets of the saccades correctly, in contrast to the proposed algorithm, which uses a two-step saccade-detection process which explicitly detects saccadic onsets and offsets. On the basis of testing a variety of different signal measures and sliding-window approaches, the proposed algorithm uses the positional displacement as the signal measure and applies the $WASW_B$ approach to classify the samples into fixations and smooth-pursuit movements. This combination discriminates between fixations and smooth pursuits better than to the I-VDT algorithm which uses the dispersion of the signal applied in a basic sliding-window approach.

In contrast to the proposed algorithm and the I-VDT algorithm, the event detector built-in to the eye-tracking glasses is applied directly to the data recorded by the eye-tracking glasses without preliminary head-movement compensation. The saccade-detection performance of the built-in event detector is slightly better than that of the I-VDT algorithm, but is poorer than that of the proposed algorithm nevertheless. Thus, the velocity-based built-in algorithm adopts adaptive thresholds and is able to reliably detect saccades from raw eye-tracking data. It is not able to discriminate between fixations and smooth-pursuit movement, however, in contrast to the proposed algorithm, which operates on signals containing eye-in-space motion only. This emphasises the need for and importance of head-movement compensation when analysing eye-tracking data recorded with a mobile eye-tracking device.

Further investigation of the detection performance of the proposed algorithm showed that a few samples were detected by the algorithm as smooth pursuits in intervals where only fixational eye movements are assumed to have been performed. A possible reason for the misclassification could be that in the case of moving-dot stimuli, the eye is still at the beginning of the movement, although the dot is moving on the screen. Some samples were also marked as smooth pursuits at parts of the signal where the accuracy of the head-movement compensation was low. Conversely, detections of smooth-pursuit movements in intervals where only fixational movements are assumed to have been performed may have been caused by drifts during fixations and remainders from saccadic movements in the case of the jumping-dot stimuli.

In order to assess the limits of the algorithm, further testing is needed. For this purpose, however, a larger database is required. Future work would involve a larger test study with a greater number of participants and a larger variation of stimuli characteristics. The stimuli signals should provide data to verify properties such as the slowest saccade, the fastest smooth pursuit, or smallest eye movement detectable by the algorithm. Moreover, the stimuli signals should be extended such that the intersaccadic intervals consist of different types of eye movements.

The performance of the proposed algorithm was quantitatively evaluated on a sample-to-sample basis in terms of sensitivity and specificity by comparing the detected events to manual annotations. Manual annotations may suffer from human subjectivity and inconsistency, however, because different experts may have different annotation behaviour. Therefore, with a larger database, a combination of different evaluation strategies would be beneficial. This may include the calculation of event properties such as the mean duration, the peak velocity or the total number of different types of eye movements. Other possible strategies include Cohen's kappa analysis, which evaluates the overall agreement between manual annotations and the detections of an algorithm, or calculating scores as proposed in [31, 39].

The results of Section 4.5.3 show that it is barely possible to improve the classification performance of fixations and smooth pursuits by combining different signal measures and thresholds. This is evidenced by the overlapping clusters of fixational and smooth-pursuit movement samples. With a larger database, however, all signal measures may be combined as proposed in [50, 42] without risking problems such as overfitting. To classify events, different classification and clustering algorithms may be tested in combination with prior feature-selection algorithms in order to select the most relevant signal measures.

Chapter 6

Conclusion

In this master's thesis, a two-step procedure to automatically classify the three most common types of eye movements from mobile eye-tracking data is proposed. The eye-in-space motion is estimated by combining the eye-tracking signals with the head-tracking signals recorded using an IMU, and a new enhanced event-detection algorithm is applied to detect saccades, fixations, and smooth-pursuit movements.

The results of a pilot study show that by compensating for head movements, the precision of the mobile eye-tracking data can be significantly improved, and a relatively high overall degree of accuracy can be achieved. Furthermore, the proposed event-detection algorithm is able to accurately perform ternary classification of eye movements based on mobile eye-tracking data. With sensitivities and specificities over 95 %, it clearly outperforms two alternative algorithms used for comparison.

The present work demonstrates that, in a controlled environment, head movements can accurately be compensated for in eye-tracking data by using an IMU, and that robust event detection can be achieved. Future work involves further analysis of the gaze-estimation model in order to extend it to handle more complex situations. Moreover, a larger test study with more participants and a larger database of stimuli will help to assess the limits of the presented method.

Appendix A

Paper to PETMEI 2014

Conference paper contribution to 4th International Workshop on Prevasive Eye Tracking and Mobile Eye-Based Interaction (PETMEI) at September 13th, 2014, in Seattle, USA [53].

Compensation of Head Movements in Mobile Eye-Tracking Data Using an Inertial Measurement Unit

Linnéa Larsson
Department of Biomedical Engineering
Lund Univeristy, Sweden
linnea.larsson@bme.lth.se

Marcus Nyström
Lund University Humanities Laboratory
Lund Univeristy, Sweden
marcus.nystrom@humlab.lu.se

Andrea Schwaller
Department of Biomedical Engineering
Lund Univeristy, Sweden
schwaand@ee.ethz.ch

Martin Stridh
Department of Biomedical Engineering
Lund Univeristy, Sweden
martin.stridh@bme.lth.se

Kenneth Holmqvist
Lund University Humanities Laboratory
Lund Univeristy, Sweden
kenneth.holmqvist@humlab.lu.se

UbiComp '14, September 13 - 17 2014, Seattle, WA, USA
Copyright 2014 ACM 978-1-4503-3047-3/14/09?$15.00.
http://dx.doi.org/10.1145/2638728.2641693

Abstract

Analysis of eye movements recorded with a mobile eye-tracker is difficult since the eye-tracking data are severely affected by simultaneous head and body movements. Automatic analysis methods developed for remote-, and tower-mounted eye-trackers do not take this into account and are therefore not suitable to use for data where also head- and body movements are present. As a result, data recorded with a mobile eye-tracker are often analyzed manually. In this work, we investigate how simultaneous recordings of eye- and head movements can be employed to isolate the motion of the eye in the eye-tracking data. We recorded eye-in-head movements with a mobile eye-tracker and head movements with an Inertial Measurement Unit (IMU). Preliminary results show that by compensating the eye-tracking data with the estimated head orientation, the standard deviation of the data during vestibular-ocular reflex (VOR) eye movements, was reduced from 8.0° to 0.9° in the vertical direction and from 12.9° to 0.6° in the horizontal direction. This suggests that a head compensation algorithm based on IMU data can be used to isolate the movements of the eye and therefore simplify the analysis of data recorded using a mobile eye-tracker.

Author Keywords

Signal Processing, Eye-tracking, Head Movement Measurement, Inertial Measurement Unit

ACM Classification Keywords

G.4 [Mathematical software]: Algorithm design and analysis.

General Terms

Measurement

Introduction

The interest and popularity of mobile eye-tracking have increased significantly in recent years. Mobile eye-trackers make it possible to record eye movements outside the laboratory in a natural environment. When the degrees of freedom to move the head and the body are increasing, the complexity of the data analysis increases dramatically [3]. In many mobile eye-trackers, the recorded data are mapped into the coordinate system of the simultaneously recorded scene video. Since the coordinate system of the scene video is not fixed when the head and body move, analysis of the recorded data is difficult to perform. Algorithms developed for the analysis of data recorded from remote-, and tower-mounted eye-trackers do not take head movements into account and are therefore not suitable to use when head-, and body movements are present [4]. In this paper, we present preliminary work on the feasibility to measure head movements with an IMU, and subtract it from the eye-tracking signals to isolate eye movements.

Related work

In the literature, there are several methods presented for recording head movements together with eye movements, e.g., with a magnetic field [2], optically with a laserBird system [8], with an omnidirectional vision sensor [7], image processing of the scene video [5], and an accelerometer [1]. Both of the methods in [2] and [8] are only applicable when the recording is performed in the laboratory or in a limited area. The other three methods can be used in recordings with natural environments. Previously, recordings using an accelerometer have suffered from problems with drift and difficulties to synchronize the accelerometer data with the eye-tracking data [5]. The main goal with compensation of the head movements is to be able to perform automatic analysis of the signals recorded with a mobile eye-tracker.

Method

When an eye-tracker is used in a natural situation where the head can be moved, the signal that the eye-tracker is recording, $s(n)$, is a combination of eye movements and head movements and can be expressed as the sum of these movements. The expression for the movement in the horizontal direction, x, is

$$s_x(n) = e_x(n) + h_x(n) + \eta_x(n) \qquad (1)$$

where $e_x(n)$ is the movements of the eye, $h_x(n)$ the movements of the head, and $\eta_x(n)$ is a noise component. An analogue expression for the recorded signal can be written for the vertical direction, $s_y(n)$. In order to compensate for the effect of $h_x(n)$ and $h_y(n)$, head movements are measured with the sensors in the IMU. The IMU board includes an AHRS-algorithm described in [6]. The AHRS-algorithm is a fusion algorithm that combines the three signals from each of the gyro, the accelerometer, and the magnetometer into a three dimensional signal containing the orientations of each of the Euler angles. The Euler angles describe the pitch, ϕ, roll, θ, and yaw, ψ, which correspond to rotations around the x', y', and z'-axes, respectively [6].

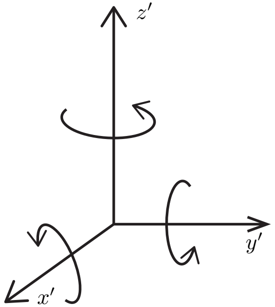


Figure 1: A description of the coordinate system for the IMU, where the y'-direction points in the direction of the nose.

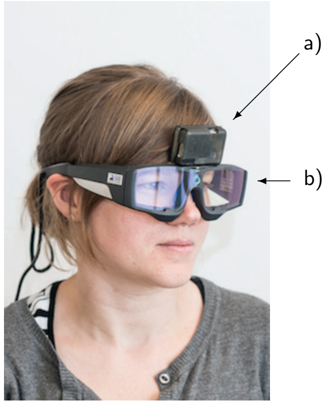


Figure 2: The apparatus setup used in this work. a) The IMU, and b) the eye-tracking glasses.

The description of the coordinate system for the IMU is shown in Figure 1, where the y'-direction points in the direction of the nose. In this study we are only using the estimated orientations of the head movements around the x' and z'-axes, e.g., nodding corresponds to a rotational movements around the x'-axis and shaking the head corresponds to a rotational movement around the z'-axis.

The estimated orientation is described in angles, while the eye-tracking data extracted from the eye-tracker are expressed in pixels of the scene video. In order to compensate for the head movements in the eye-tracking data, the signals need to be converted to the same coordinate system. We decided to map the angles of the head movements to the coordinate system of the eye-tracking data. The projections of the angles, ϕ, and ψ, are described by

$$\hat{h}_x = A\frac{x_{max}}{2\tan(\frac{\alpha_{max}}{2})}\tan(\phi) \tag{2}$$

$$\hat{h}_y = B\frac{y_{max}}{2\tan(\frac{\beta_{max}}{2})}\tan(\psi) \tag{3}$$

where x_{max} and y_{max} are the resolution of the scene video in pixels in the x and y directions, respectively. The corresponding maximum angles of the camera in the x and y directions are α_{max} and β_{max}, respectively. A and B are compensatory factors that are chosen to optimize the subtraction of the head movements, (A = 1.15 and B = 1.19). By subtracting the estimated head movements, $\hat{h}_x$ and $\hat{h}_y$, from Equation 1, the resulting signals, $\hat{s}_x(n)$ and $\hat{s}_y(n)$, can be expressed by

$$\hat{s}_x(n) = s_x(n) - \hat{h}_x(n) = \hat{e}_x(n) + \eta_x(n) \tag{4}$$

$$\hat{s}_y(n) = s_y(n) - \hat{h}_y(n) = \hat{e}_y(n) + \eta_y(n) \tag{5}$$

which consist of the estimated eye movement signal and a noise component.

Experiment and apparatus

In this pilot study, the eye-tracking signals were recorded using the eye-tracking glasses from SensoMotoric Instruments (SMI), with a sampling frequency of 60 Hz. On the forehead of the test person, an Inertial Measurement Unit (IMU) from x-io Technologies was placed, see Figure 2. The IMU consisted of a three-axial gyro, a three-axial accelerometer, and a three-axial magnetometer, and had a sampling frequency of 512 Hz.

The stimulus consisted of black dots on a white background, presented with a video projector on a large white wall with dimensions (1.4 x 1.9 m). The test person was placed in front of the screen at a distance of 2.5 m, with the eyes aligned with the center of the screen. In order to be able to synchronize the eye-tracking signals with the signals recorded with the IMU, the experiment started with a short synchronization period, where the test person was asked to fixate on a dot and move the head horizontally. The second part of the experiment consisted of only movements of the eyes. A dot was shown in each of the four corners of the wall, and the task was to move only the eyes to the next corner and fixate on the dot while having the head as still as possible. The next task was to move only the head. The eyes were fixating on a dot in the center of the wall while the head was moved horizontally, vertically, and a mix of both. The last part of the experiment consisted of combined movements of the eyes and the head.

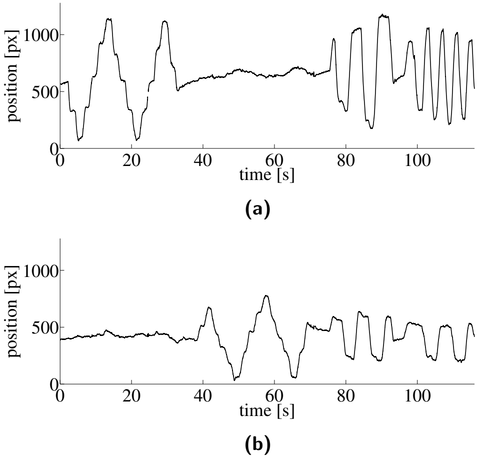


Figure 3: Example of recorded eye-tracking signal when fixating a stationary target and moving the head, (a) in the x-direction and (b) in the y-direction.

Results

An example of the recorded eye-tracking signal, when the eyes are fixating a stationary target and the head is moving first in the horizontal and then in the vertical direction, is shown in Figure 3. The estimated orientation of the head is shown in Figure 4 and the resulting eye signal after subtraction of the estimated head movements is shown in Figure 5.

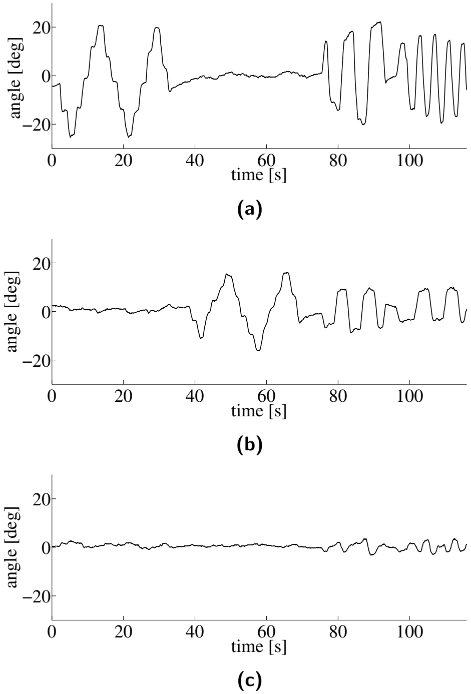


Figure 4: Example of estimated head orientation when fixating a stationary target and moving the head, in (a) ϕ, (b) ψ, and (c) θ.

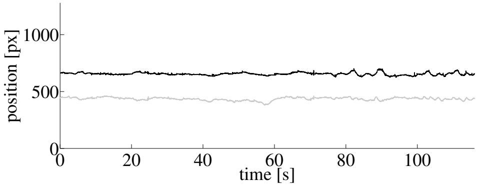


Figure 5: Example of the resulting eye signal when the compensation of the head movements is preformed. Black line corresponds to the x-direction and grey line to the y-direction.

In order to evaluate the effect when compensating for head movements in the eye-tracking data, the standard deviation of the data during fixating on a stable target was calculated for four different cases.

1. Fixating on a stationary target with the head still and no compensation of head movements.
2. Fixating on a stationary target with head moving and no compensation of head movements.
3. Fixating on a stationary target with the head still and compensation of head movements.
4. Fixating on a stationary target with head moving and compensation of head movements.

The standard deviation of the data for the four cases was calculated as

$$\sigma = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(\hat{s}_x(n) - \bar{\hat{s}}_x)^2} \tag{6}$$

for the horizontal and the vertical direction separately, where $\hat{s}_x(n)$ is the resulting eye signal, $\bar{\hat{s}}_x$ is its mean and N is the length of the signal.

	Not compensated		**Compensated**	
σ (°)	x	y	x	y
Head still	0.46	0.42	0.21	0.26
Head moving	12.89	7.99	0.61	0.85

Table 1: Standard deviation of the data for a stationary target with and without head movements, for compensated and not compensated data, in both the horizontal and vertical directions.

The results for the four cases are shown in Table 1. Even when the test person is asked to not move the head, the compensation reduces the standard deviation of the data from around 0.5° without compensation to around 0.2° when using head movement compensation. When the head is intentionally moving, compensation of the head movement reduces the standard deviation from 12.89° to 0.61° in the horizontal direction and from 7.99° to 0.85° in the vertical direction.

Discussion

In this work, the orientation of the head has been estimated using an IMU and has been subtracted from the recorded eye-tracking signal. The results show that by compensating for the head movements, the standard deviation of the eye-tracking data is significantly reduced, also for data when the test person is asked to keep the head as still as possible. The reduction of the standard deviation makes the signal recorded with a mobile eye-tracker more similar to the eye-tracking signals recorded with a remote-, or tower-mounted eye-tracker, which in turn makes it easier to develop algorithms that are able to separate between the eye movements saccades, fixations, smooth pursuits, and VOR.

Previous work where an accelerometer has been used to measure the movements of the head, has suffered from drift in the accelerometer. By estimating the head orientation with the AHRS-algorithm, which combines the signals from the gyro, the accelerometer, and the magnetometer, this drift can be compensated for [9].

The present work demonstrates that, in a controlled environment, head movements can accurately be compensated for in eye-tracking data by estimating the orientation of the head using an IMU. Future work will show whether the method can be extended to more complex situations, where both the head, the body, and the environment may be moving. Such work requires the use of the third estimated head orientation as well.

Future work will involve performing a larger study where a greater number of participants is included and where the proposed method is compared to previously presented ones.

In relation to other head movement estimation techniques, an IMU is light and relatively easy to integrate into a mobile eye-tracking device.

Acknowledgements

This work was supported by the Strategic Research Project eSSENCE, founded by the Swedish Research Council.

References

[1] Ahlström, C., Victor, T., Wege, C., and Steinmetz, E. Processing of eye/head-tracking data in large-scale naturalistic driving data sets. *IEEE Transactions on intelligent transportation system 13*, 2 (2012), 553–564.

[2] Allison, R. S., Eizenman, M., and Cheung, B. S. K. Combined head and eye tracking system for dynamic

testing of the vestibular system. *IEEE Transactions on Biomedical Engineering 41*, 11 (1996), 1073–1082.

[3] Essig, K., Sand, N., Schack, T., Künsemöller, J., Weigelt, M., and Ritter, H. Fully-automatic annotation of scene videos. In *SICE Annual Conference* (2010), 3304– 3307.

[4] Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., and van de Weijer, J. *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press, 2011.

[5] Kinsman, T., Evans, K., Sweeney, G., Keane, T., and Pelz, J. B. Ego-motion compensation improves fixations detection in wearable eye tracking. In *ETRA '12 Proceedings of the Symposium on Eye Tracking Research and Applications*, ACM (2012), 221– 224.

[6] Madgwick, S. O. H., Harrison, A. J. L., and Vaidyanathan, R. Estimation of imu and marg orientation using a gradient descent algorithm. In *IEEE International Conference of Rehabilitation Robotics* (2011).

[7] Rothkopf, C. A., and Pelz, J. B. Head movement estimation for wearable eye tracker. In *ETRA '04 Proceedings of the Symposium on Eye Tracking Research and Applications*, ACM (2004), 123– 130.

[8] Tafaj, E., Kübler, T., Kasneci, G., Rosenstiel, W., and Bogdan, M. Online classification of eye tracking data for automated analysis of traffic hazard perception. In *Artificial Neural Networks and Machine Learning – ICANN 2013*, Springer (2013), 442–450.

[9] x ioTechnologies. x-IMU User Manual 5.2, x-io Technologies, November 2013. http://www.x-io.co.uk/downloads/.

Bibliography

[1] C. Hendry, A. Farley, and E. McLafferty, “Anatomy and physiology of the senses,” *Nursing Standard*, vol. 27, no. 5, pp. 35–42, 2012.

[2] Wikimedia Commons, “Schematic diagram of the human eye en,” 2007.

[3] C. Garhart and V. Lakshminarayanan, “Anatomy of the eye,” in *Handbook of Visual Display Technology.* Springer, 2012, pp. 73–83.

[4] E. P. Widmaier, H. Raff, and K. T. Strang, *Vander’s human physiology: the mechanisms of body function.* McGraw-Hill Higher Education, 2011.

[5] J. D. Enderle and J. D. Bronzino, *Introduction to biomedical engineering.* Academic Press, 2012.

[6] R. J. Leigh and D. S. Zee, *The neurology of eye movements.* Oxford University Press New York, 1999, vol. 90.

[7] A. Duchowski, *Eye tracking methodology: Theory and practice.* Springer, 2007, vol. 373.

[8] L. Larsson, M. Nystrom, and M. Stridh, “Detection of saccades and postsaccadic oscillations in the presence of smooth pursuit,” *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 9, pp. 2484–2493, 2013.

[9] S. Martinez-Conde, S. L. Macknik, and D. H. Hubel, “The role of fixational eye movements in visual perception,” *Nature Reviews Neuroscience*, vol. 5, no. 3, pp. 229–240, 2004.

[10] K. Rayner, “Eye movements in reading and information processing: 20 years of research.” *Psychological bulletin*, vol. 124, no. 3, p. 372, 1998.

[11] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. Van de Weijer, *Eye tracking: A comprehensive guide to methods and measures.* Oxford University Press, 2011.

[12] S. Van der Stigchel, "Recent advances in the study of saccade trajectory deviations," *Vision research*, vol. 50, no. 17, pp. 1619–1627, 2010.

[13] E. Kowler, "Eye movements: The past 25years," *Vision research*, vol. 51, no. 13, pp. 1457–1483, 2011.

[14] C. H. Meyer, A. G. Lasker, and D. A. Robinson, "The upper limit of human smooth pursuit velocity," *Vision research*, vol. 25, no. 4, pp. 561–563, 1985.

[15] C. Ahlstrom, T. Victor, C. Wege, and E. Steinmetz, "Processing of eye/head-tracking data in large-scale naturalistic driving data sets," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 2, pp. 553–564, 2012.

[16] H. Collewijn and J. Smeets, "Early components of the human vestibulo-ocular response to head rotation: latency and gain," *Journal of Neurophysiology*, 2000.

[17] R. Gellman, J. Carl, and F. Miles, "Short latency ocular-following responses in man," *Visual neuroscience*, vol. 5, no. 02, pp. 107–122, 1990.

[18] L. R. Young and D. Sheena, "Eye-movement measurement techniques." *American Psychologist*, vol. 30, no. 3, p. 315, 1975.

[19] N. Wade and B. Tatler, "The moving tablet of the eye: The origins of modern eye movement research," 2005.

[20] D. W. Hansen and Q. Ji, "In the eye of the beholder: A survey of models for eyes and gaze," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 3, pp. 478–500, 2010.

[21] A. Al-Rahayfeh and M. Faezipour, "Eye tracking and head movement detection: A state-of-art survey," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 1, pp. 11–22, 2013.

[22] R. S. Allison, E. Eizenman, and B. S. Cheung, "Combined head and eye tracking system for dynamic testing of the vestibular system," *IEEE Transactions on Biomedical Engineering*, vol. 43, no. 11, pp. 1073–1082, 1996.

[23] E. M. Foxlin, M. Harrington, and Y. Altshuler, “Miniature six-dof inertial system for tracking hmds,” in *Aerospace/Defense Sensing and Controls*. International Society for Optics and Photonics, 1998, pp. 214–228.

[24] E. Tafaj, T. C. Kübler, G. Kasneci, W. Rosenstiel, and M. Bogdan, “Online classification of eye tracking data for automated analysis of traffic hazard perception,” in *Artificial Neural Networks and Machine Learning–ICANN 2013*. Springer, 2013, pp. 442–450.

[25] K. Satoh, S. Uchiyama, and H. Yamamoto, “A head tracking method using bird’s-eye view camera and gyroscope,” in *Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality*. IEEE Computer Society, 2004, pp. 202–211.

[26] C. A. Rothkopf and J. B. Pelz, “Head movement estimation for wearable eye tracker,” in *Proceedings of the 2004 symposium on Eye tracking research & applications*. ACM, 2004, pp. 123–130.

[27] T. Kinsman, K. Evans, G. Sweeney, T. Keane, and J. Pelz, “Ego-motion compensation improves fixation detection in wearable eye tracking,” in *Proceedings of the Symposium on Eye Tracking Research and Applications*. ACM, 2012, pp. 221–224.

[28] A. Sasou, “Acoustic head orientation estimation applied to powered wheelchair control,” in *Robot Communication and Coordination, 2009. ROBOCOMM’09. Second International Conference on*. IEEE, 2009, pp. 1–6.

[29] F. Li, J. B. Pelz, and S. J. Daly, “Measuring hand, head, and vehicle motions in commuting environments,” in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2009, pp. 72 401I–72 401I.

[30] D. D. Salvucci and J. H. Goldberg, “Identifying fixations and saccades in eye-tracking protocols,” in *Proceedings of the 2000 symposium on Eye tracking research & applications*. ACM, 2000, pp. 71–78.

[31] O. V. Komogortsev, D. V. Gobert, S. Jayarathna, D. H. Koh, and S. M. Gowda, “Standardization of automated analyses of oculomotor fixation and saccadic behaviors,” *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 11, pp. 2635–2645, 2010.

[32] P. Blignaut, "Fixation identification: The optimum threshold for a dispersion algorithm," *Attention, Perception, & Psychophysics*, vol. 71, no. 4, pp. 881–895, 2009.

[33] F. Shic, B. Scassellati, and K. Chawarska, "The incomplete fixation measure," in *Proceedings of the 2008 symposium on Eye tracking research & applications*. ACM, 2008, pp. 111–114.

[34] R. Engbert and R. Kliegl, "Microsaccades uncover the orientation of covert attention," *Vision research*, vol. 43, no. 9, pp. 1035–1045, 2003.

[35] M. Nyström and K. Holmqvist, "An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data," *Behavior research methods*, vol. 42, no. 1, pp. 188–204, 2010.

[36] R. van der Lans, M. Wedel, and R. Pieters, "Defining eye-fixation sequences across individuals and tasks: the binocular-individual threshold (bit) algorithm," *Behavior research methods*, vol. 43, no. 1, pp. 239–257, 2011.

[37] F. Behrens, M. MacKeben, and W. Schröder-Preikschat, "An improved algorithm for automatic detection of saccades in eye movement data and for calculating saccade parameters," *Behavior research methods*, vol. 42, no. 3, pp. 701–708, 2010.

[38] M. Dorr, T. Martinetz, K. R. Gegenfurtner, and E. Barth, "Variability of eye movements when viewing dynamic natural scenes," *Journal of vision*, vol. 10, no. 10, p. 28, 2010.

[39] O. V. Komogortsev and A. Karpov, "Automated classification and scoring of smooth pursuit eye movements in the presence of fixations and saccades," *Behavior research methods*, vol. 45, no. 1, pp. 203–215, 2013.

[40] D. J. Berg, S. E. Boehnke, R. A. Marino, D. P. Munoz, and L. Itti, "Free viewing of dynamic stimuli by humans and monkeys," *Journal of Vision*, vol. 9, no. 5, p. 19, 2009.

[41] L. Larsson, M. Nystrom, and M. Stridh, "Discrimination of fixations and smooth pursuit movements in high-speed eye-tracking data," in *Proc. 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC '14*. IEEE, 2014, p. 3797–3800.

[42] M. Vidal, A. Bulling, and H. Gellersen, “Detection of smooth pursuits using eye movement shape features,” in *Proceedings of the Symposium on Eye Tracking Research and Applications.* ACM, 2012, pp. 177–180.

[43] E. Tafaj, G. Kasneci, W. Rosenstiel, and M. Bogdan, “Bayesian online clustering of eye movement data,” in *Proceedings of the Symposium on Eye Tracking Research and Applications.* ACM, 2012, pp. 285–288.

[44] SensoMotoric Instruments, *iViewETG Manual. Version 2.0*, September 2013.

[45] S. O. Madgwick, A. J. Harrison, and R. Vaidyanathan, “Estimation of imu and marg orientation using a gradient descent algorithm,” in *2011 IEEE International Conference on Rehabilitation Robotics (ICORR).* IEEE, 2011, pp. 1–7.

[46] x-io Technologies, *x-IMU User Manual 5.2*, November 2013.

[47] C. Blakemore and M. Donaghy, “Co-ordination of head and eyes in the gaze changing behaviour of cats,” *The Journal of physiology*, vol. 300, p. 317, 1980.

[48] E. Viirre, D. Tweed, K. Milner, and T. Vilis, “A reexamination of the gain of the vestibuloocular reflex,” *J Neurophysiol*, vol. 56, no. 2, pp. 439–450, 1986.

[49] iSolver Software Solutions, *ioHub Event Monitoring Framework*, http://www.isolver-solutions.com/iohubdocs/, March 2014.

[50] M. Vidal, A. Bulling, and H. Gellersen, “Analysing eog signal features for the discrimination of eye movements with wearable devices,” in *Proceedings of the 1st international workshop on pervasive eye tracking & mobile eye-based interaction.* ACM, 2011, pp. 15–20.

[51] C. Paladini, K. Mergenthaler, R. Kliegl, R. Engbert, and M. Holschneider, “Microsaccade characterization using the continuous wavelet transform and principal component analysis,” *Journal of Eye Movement Research*, vol. 3, no. 5, p. 1, 2010.

[52] L. Sörnmo and P. Laguna, *Bioelectrical signal processing in cardiac and neurological applications.* Academic Press, 2005.

[53] L. Larsson, A. Schwaller, K. Holmqvist, M. Nyström, and M. Stridh, “Compensation of head movements in mobile eye-tracking data using

an inertial measurement unit," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication.* ACM, 2014, pp. 1161–1167.