CHALMERS
UNIVERSITY OF TECHNOLOGY

Real-time filtering of gaze direction signals by means of a switching Kalman filter

Master's thesis in Communication Engineering

Sebastian Adolfsson
Oskar Lindblad

Department of Electrical Engineering or Technology
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2020

Master's thesis 2020

Real-time filtering of gaze direction signals by means of a switching Kalman filter

Sebastian Adolfsson
Oskar Lindblad

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Electrical Engineering
Chalmers University of Technology
Gothenburg, Sweden 2020

Real-time filtering of gaze direction signals by means of a switching Kalman filter
SEBASTIAN ADOLFSSON
OSKAR LINDBLAD

Supervisor: Mohammad Hossein Moghaddam, Communication Systems
Supervisor: John Finér, Smart Eye AB
Supervisor: Markus Lindelöw, Smart Eye AB
Examiner: Erik Agrell, Communication Systems

Master's Thesis 2020
Department of Electrical Engineering
Division of Communication Engineering
Communication Systems
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in LaTeX
Gothenburg, Sweden 2020

Real-time filtering of gaze direction signals by means of a switching Kalman filter
SEBASTIAN ADOLFSSON & OSKAR LINDBLAD
Department of Electrical Engineering
Chalmers University of Technology

Abstract

In eye-tracking applications, the goal is to estimate the gaze direction and eye motion of the subject at hand. The motion of the human eye has different modes, such as fixation, saccades, glances, and smooth pursuits. The goal of this thesis is to find a technique that improves the gaze signal in the system as well as classifying the current mode for real-time applications. Therefore, this thesis explores different techniques for filtering the various modes of the eye, including using the head as a basis for the gaze direction.

The thesis was carried out at Smart Eye AB in Gothenburg, Sweden. The eye-tracking system is built upon estimating the eye-state from near infrared cameras. The raw data was then extracted from videos recorded with Smart Eye's system. From the extracted data, this thesis will try to improve the gaze data with filtering. The main filtering method evaluated in this report is based on a switching Kalman filter consisting of two modes. One mode describes fixations and smooth pursuits and the other one describes saccades and glances. Both of the modes consist of constant velocity models with different uncertainty levels. This is to keep one reactive to changes in the signal such as saccades and glances, and the other for filtering during fixation and smooth pursuits. The filter switches between these modes by using a velocity based threshold test and by synchronizing the thresholds test information for both eyes to get a more robust classification.

The resulting filter successfully reduces the noise levels in the gaze direction signal, hence improving the precision of the system. By using the head orientation as a gaze estimate, the filter also increases the availability of the gaze direction signal. Combining the information derived from each eye improved the classification accuracy of the filter. By combining the information from each respective eye the precision of the estimated consensus gaze was also improved. What could be further developed from this report is a better process model for saccades in real-time. A more suitable process model for saccades could improve the accuracy of the classification as well as providing the filter with better filtering capabilities during saccades.

Keywords: switching Kalman filter, Kalman filter, gaze direction, eye-tracking, fixation, smooth pursuit, saccades, glances, classification.

Acknowledgements

A special thanks to our supervisors John Finér and Markus Lindelöw, at Smart Eye AB, for believing in us and contributing with their expertise and support during the implementation of the algorithms for this report.

We would also take the time to thank our examiner and supervisor at Chalmers University of Technology, Erik Agrell and Mohammad Hossein Moghaddam for their guidelines in directing the report to its current state.

Finally, we would like to thank our families for their support and encouragement throughout all these years at Chalmers and during this thesis project.

Sebastian Adolfsson, Oskar Lindblad, Gothenburg, May 2020

Contents

List of Figures **xi**

List of Tables **xiii**

1 Introduction **1**
1.1 Background . . . 1
1.2 Purpose . . . 2
1.3 Objectives . . . 2
1.4 Scope . . . 2

2 Theory **5**
2.1 Eye motion . . . 5
2.1.1 Fixation . . . 5
2.1.2 Smooth pursuit . . . 5
2.1.3 Saccade . . . 6
2.1.4 Glances . . . 6
2.2 Kalman filter . . . 6
2.3 Switching Kalman filter . . . 8
2.4 Head pose . . . 9
2.5 Eye origin . . . 9

3 Methods **11**
3.1 Software . . . 11
3.2 Data gathering . . . 11
3.2.1 Manually annotated data for classification of motions . . . 11
3.2.2 Prerecorded data with ground truth for fixations . . . 12
3.2.3 Synthetic data . . . 12
3.2.3.1 Synthetic data for consensus gaze estimation . . . 13
3.3 Real-time eye motion and gaze filtering . . . 13
3.3.1 Constant velocity Kalman filter . . . 13
3.3.2 Noise estimation . . . 15
3.3.2.1 Flush buffer during saccade . . . 15
3.3.3 Mode switching . . . 15
3.3.3.1 Velocity test . . . 16
3.3.3.2 Combine information from both eyes . . . 16
3.3.4 Head orientation for gaze estimation . . . 16
3.3.5 Flow chart of the filter . . . 17

3.4 Estimation of the consensus gaze . . . 18
3.4.1 Estimation of the consensus origin . . . 19

4 Results **21**
4.1 Filtering . . . 21
4.1.1 Filtering during fixation . . . 22
4.1.1.1 Head orientation estimates during fixations . . . 25
4.1.2 Filtering during smooth pursuit . . . 26
4.1.3 Filtering during saccades and glances . . . 27
4.1.4 Consensus filtering . . . 27
4.1.5 Origin filtering . . . 28
4.2 Classification . . . 30
4.2.1 Responsivity . . . 32
4.3 Comparison with the previous filtering method . . . 32

5 Discussion **35**
5.1 Filtering . . . 35
5.1.1 Fixations . . . 35
5.1.2 Smooth pursuit . . . 35
5.1.3 Saccade . . . 36
5.1.4 Consensus filtering . . . 37
5.2 Origin filtering . . . 37
5.3 Classification . . . 38
5.3.1 Responsivity . . . 38
5.4 Combining information from both eyes . . . 39
5.5 Offset in ground truth data . . . 39
5.6 Comparision with other filter methods . . . 40
5.7 Future work and general filter performance discussion . . . 40

6 Conclusion **41**

Bibliography **43**

A Conversion from 3D direction vector to heading and pitch **I**

List of Figures

2.1 A schematic of the different steps performed in a switching Kalman filter [9]. 8

3.1 Schematic view of the different steps in the filtering of the gaze direction. 18

4.1 Filterered heading together with raw gaze signal during a sequence with fixations and saccades for the left eye. 21
4.2 Filterered pitch together with raw gaze signal during a sequence with fixations and saccades for the left eye. 22
4.3 Filtering in the heading rotation during a fixation for left eye. 22
4.4 Filtered and raw pitch signal together with ground truth value for the right eye. 24
4.5 The reduction of the standard deviation evaluated for different noise levels during fixations. The reduction of the standard deviation was evaluated using synthetic data. 25
4.6 Filtered heading rotation for the right eye using the head rotation for estimation when no gaze data is available. 26
4.7 Filtered gaze for the left eye during a smooth pursuit. 26
4.8 Filtered gaze direction versus raw gaze data for the left eye heading rotation during two saccades. 27
4.9 Filtered consensus signal together with raw signal for consensus derived from Smart Eye's system. 28
4.10 Filtered origin signal together with raw signal in the x-direction for the right eye. 29
4.11 Filtered origin signal together with raw signal in the x-direction for consensus derived from Smart Eye's system. 29
4.12 Saccade classification when only using information from one eye. . . . 31
4.13 Saccade classification when using information from both eyes. 31
4.14 Comparison between the implemented Kalman filter with the previous gaze filter. 33

A.1 The figure shows a 3D-view of heading, pitch, and roll. Roll, the rotation around the z-axis, is often not used in eye tracking because it is an unnatural movement in eye-tracking applications. II

List of Tables

4.1 Trueness and precision for the left eye without head pose correction. . 23
4.2 Trueness and precision for the right eye without head pose correction. 23
4.3 Trueness and precision for the consensus eye without head pose correction. 24
4.4 Confusion matrix for classification accuracy. A positive outcome indicates a saccade. 32

1
Introduction

This chapter will firstly cover an introduction of the relevant background for eye-tracking. Then the purpose, objective, and scope of this thesis are explained, respectively. The objective of this thesis is to improve the quality, in terms of trueness and precision, of the gaze direction and classify the motion of the human eye in real-time from image-based sensor data. Further will this report look at methods for estimating the gaze direction when no direct gaze signal is available. For this, methods for estimating the gaze direction by using the head orientation has been evaluated during this project. Methods for achieving this are based on filtering techniques and physical models of the human eye motion.

1.1 Background

The human eye, also called "the world's worst camera", is the subject investigated in eye-tracking - the technique of tracking the motion of the eye [1]. A person's visual attention span can for the most part be determined by the direction of the gaze. Hence, by tracking the motion of the eye and its gaze direction, information about the visual attention can be quantified and extracted. This information could thereafter be deployed in a wide variety of areas such as:

- assistance in driving: determine if the driver's focus is on the road or elsewhere;
- advertising: determine if the subject is viewing the advertisements directed towards him or her;
- human-computer interaction: information derived from the eye could be used as a control signal for computer programs;
- medical research: track eye-motion in medical application areas.

Smart Eye AB, in Gothenburg, is a company at the technological edge in eye-tracking. The company provides both medical research with eye-tracking systems as well as driver monitoring systems. Their algorithms are, as of right now, deriving gaze direction data from eye features extracted from image sensors. By filtering the gaze signal derived by their system, they are hoping to increase the accuracy of the signal as well as using techniques for improving the availability of the gaze direction signal, i.e., estimating the gaze direction signal even when no data is coming directly from the eye features. The filtering of the signal should handle all types of motion of the human eye, such as:

- fixations;
- smooth pursuit;
- saccades;

- glances.

These motions describe most of the visual focus of the human eye [1].

1.2 Purpose

The purpose of this thesis is to improve the gaze direction signal, in terms of trueness and precision, in the current system at Smart Eye AB to be used in driver monitoring systems. By improving both trueness and precision of the system means, by definition, that the accuracy of the system has increased. This master thesis is building upon a previous master thesis at Smart Eye AB [2] for gaze direction filtering. Previously only raw gaze direction from one eye has been filtered and handled. In this thesis, both eyes will be taken into account in terms of reducing the noise of the system and classifying the mode. Furthermore, raw, gaze direction signals from both eyes will be used to estimate a combined gaze direction for both eyes, also called the consensus gaze direction[1].

Parameters as head pose and rotation will be included in the filtering process to try to estimate how these parameters could improve the availability of the gaze signal when no direct gaze direction signal is available.

The algorithms that are to be implemented are supposed to run in real-time on the current system at Smart Eye AB. To achieve the goal of improving the gaze direction quality, techniques based on filtering theory will be applied to the system.

1.3 Objectives

This section will highlight the most important questions in this thesis. The objectives of this thesis are the following:

- Finding feasible models for human eye motion and applying them in the filtering process;
- Combining the motion of the eye with the head rotation for improving the availability of the gaze signal;
- Combining the information from both eyes to improve classification of motion modes and to reduce noise;
- Implementing a consensus gaze direction from the raw, gaze direction signals from both eyes.

These bullet points will be investigated and evaluated throughout this report.

1.4 Scope

The scope of this thesis is to produce a filter that can run in real-time and reduce the noise effects from sensors such that the precision of the signal increases. The unfiltered signal is extracted from the system given at Smart Eye AB and how the system extracts the gaze direction data will not be taken into account. The classification will only consider classifying saccadic events. When presented with

[1] The focus point of the eyes.

situations where the gaze direction signal is unavailable, the developed algorithm will try to estimate the gaze direction based on the head orientation of the subject at hand to improve the availability of the gaze signal.

The following will not be investigated in this report:

- Compensation for calibration errors occurred in the extraction of gaze direction data of the system.
- Optimization of the algorithm for run-time execution.

The items stated in the list above will not be the focus of the report or investigated but could potentially affect the overall filter, so their implications will not be completely disregarded.

2

Theory

In the following sections the theoretical background for the filtering techniques applied in this report will be presented together with descriptions of how the human eye motion considered in the filtering process.

2.1 Eye motion

The following sections describe the four most common types of eye movement which will be addressed in this report and evaluated in the filtering process.

2.1.1 Fixation

Fixation is the state of focusing on a single stationary point and is the most recorded motion of the eye [3]. Fixations have durability from tens of milliseconds up to seconds [3]. There are many different definitions of what defines fixation, one of them is that the velocity of the eye does not exceed $100°/s$ [4].

During fixations, the eye is subjected to small involuntary micro-movements. These small movements can be divided into three sub-classes:

- tremor;
- drifts.
- microsaccades;

The purpose of tremor is unclear, but it can be thought of as small involuntary muscle reactions; drifts are small movements that move the focus point of the eye, and micro-saccades are small rapid movements meant to move the visual focus back to the desired object during fixations [3]. During fixations, these small involuntary movements are unwanted in a filtering perspective and are often filtered out [1].

2.1.2 Smooth pursuit

Smooth pursuit is the motion of following moving targets in a continuous motion while still preserving the focus point on the target [5]. The velocity of the smooth pursuit can be described as the medium velocity of the human eye and usually does not exceed $200°/s$ [5][4].

Smooth pursuit differs from saccades, described in the coming section, by that the subject during a smooth pursuit has full visual perception, whereas during saccades the subject has limited or non-visual perception [3]. During smooth pursuits

the subject can be subjected to catch-up saccades, which are small saccades occurring when the subject under investigation momentary loses track of the visual target point and does a small saccade to "catch up" to the target [6].

2.1.3 Saccade

Saccades, which means "flick of a sail" in French [1], are described as the high-velocity motion of the human eye where the focus of the eye shifts from one fixation point to another. Saccades can be described as ballistic motions - the term ballistic comes from the fact that the eye position appears to get thrown to a predetermined position [1]. Saccades are often followed by a correction saccade to compensate for overshooting the target position [5]. These compensations are often referred to as post-saccadic oscillations, or just simply PSO.

The duration of a saccade varies between $10 - 100$ ms. During a saccade the visual perception of the subject is severely reduced [1]. Common modeling for this is to model the duration of a saccade linearly where the peak velocity increases exponentially with the amplitude of the saccade [5]. The duration of a saccade is most commonly described by:

$$D = D_1 + D_0\theta_{sac}, \tag{2.1}$$

where θ_{sac} is the amplitude of the saccade in (°) and D_1 and D_0 are parameters that are age-dependent. The best-fitted values for D_1 and D_0 in adult's eye motion are 28.70 ms and 2.59 ms/° according to [7]. Further, they suggest a model for the peak velocity of the saccade motion, which is also dependent on the saccade amplitude, θ_{sac}, according to:

$$V_{sac} = V_0(1 - e^{-\theta_{sac}/\theta_0}), \tag{2.2}$$

where $V_0 = 488.22\ ^\circ/s$ and $\theta_0 = 9.17\ ^\circ$ which were also found to be best-fitted to adults.[7]

2.1.4 Glances

Glances and saccades are in literature often not separated, since the motion pattern is similar. Glances can be described as two consecutive saccades, where the fixation-point shifts to another point then quickly return to the initial fixation-point.

2.2 Kalman filter

The Kalman filter is a filtering technique that bases its output on a predicted motion of the object together with measurements of the object's state. The Kalman filter is the closed-form solution to the Bayesian filtering equation where the dynamics are linear [8].

The principles of Kalman filtering are based upon a process, which describes the predicted motion of the state of the object, and a measurement model, which

describes how the state translates to measurements. The two models are described by the following equations:

$$u_k = A_{k-1}u_{k-1} + q_{k-1}, \tag{2.3}$$

and

$$w_k = H_k u_k + r_k, \tag{2.4}$$

where u_k is the state vector of the system at time instance k, A_{k-1} the process matrix which translates the previous state to the current predicted state, w_k is the measured state, H_k the measurement matrix which maps the state to measurement, and r_k and q_{k-1} are the measurement noise and process noise, respectively. The noise processes are often viewed as independent additive noise processes. For the Kalman filter to be the theoretical closed-form solution to the filtering problem the noise needs to be an independent zero-mean additive white Gaussian process [8].

Kalman filtering can be divided into two steps, *prediction* and *update*. In the prediction step, the filter tries to predict the current state u_k, from the previous state u_{k-1}, according to:

$$\hat{u}_k = A_{k-1}u_{k-1} \tag{2.5}$$

and the prediction covariance, $\hat{P}_k$, which describes the uncertainty in the predicted state, is calculated according to:

$$\hat{P}_k = A_{k-1}P_{k-1}{A_{k-1}}^T + Q_{k-1} \tag{2.6}$$

where Q_{k-1} is the process noise covariance. The purpose of the process matrix, A_{k-1}, is to emulate the physical motion of the state space vector from the previous time-instance, and produce a prediction of the state space motion at the current time instance.

In the update step the following equations are used:

$$V_k = w_k - H_k\hat{u}_k, \tag{2.7}$$

$$S_k = H_k\hat{P}_k{H_k}^T + R_k, \tag{2.8}$$

$$K_k = \hat{P}_k{H_k}^T{S_k}^{-1}, \tag{2.9}$$

$$u_k = \hat{u}_k + K_kV_k, \tag{2.10}$$

$$P_k = \hat{P}_k - K_kS_k{K_k}^T. \tag{2.11}$$

The innovation, V_k, is the difference between the measurement and the predicted measurement, S_k is the corresponding covariance matrix for the innovation, R_k is the measurement noise covariance, and K_k is the Kalman gain.[8]

The output of the filter, u_k[1], together with an updated covariance, P_k, is the weighted output after the update step. The posterior state vector goes into as the prior state vector in the next iteration of the filter. Hence, the output of the filter is weighted between the measurement in the system together with previous estimates of the filter.[8]

[1] also called the posterior

2.3 Switching Kalman filter

A switching Kalman filter is basically a set of Kalman filters that run in parallel, where the output of the set is determined by the filter which produces the most probable state. In figure 2.1, a schematic of the process for a switching Kalman filter is shown for two parallel processes. The two Kalman filters run in parallel and the outputs from both filters are then merged together. Depending on a switching condition in the filter, one of the posterior output from the parallel filters is chosen as the output of the switching Kalman filter, together with its corresponding covariance matrix. The posterior is then fed back as prior to both of the parallel filters in the next iteration. This process is continuously repeated over time.[9]

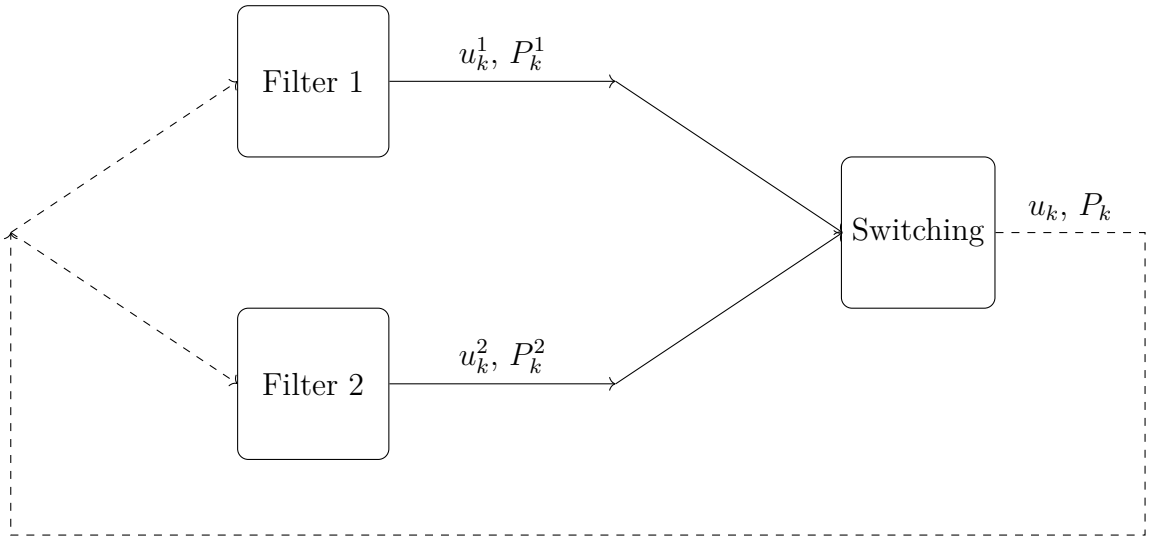

Figure 2.1: A schematic of the different steps performed in a switching Kalman filter [9].

Switching Kalman filters are developed for a process that changes behaviour over time, hence cannot be described by a single process model. The idea of a switching Kalman filter is that the process is able to switch between modes and thus react to changes in the state of the process. In turn it is more adaptable to changes than an ordinary Kalman filter.

Since no process model has been found that can capture each motion of the eye, there is a need for multiple motion models. Identification is needed so that the filter knows which mode the eyes are currently in. Therefore, identifying and switching between different motion models is important for filtering purposes.

Due to the fact that the eye-motions switch between the different motion modes there is a need for a quickly adapting filter, such as a switching Kalman filter, which can adapt quickly to sudden changes in the signal [10].

2.4 Head pose

It has been found that the head orientation contributes 68.9%, on average, to the overall gaze direction. The head orientation also moves towards the same direction as the gaze direction in 87.0%, on average.[11]

In [11] the focus of attention was studied using only head orientation and the results showed that using only head orientation in the estimation of visual attention focus gave an accuracy of 88.7%.

According to a study by Land [12] the gaze and head motion during saccades gets the same input signal, from the brain, but the motion is executed in a different manner. Land states that there is a high correlation between the head and gaze motion during saccades. Although the head motion is a slower process than the gaze motion during saccades, a correlation has been established by Land [12] which could be used for predicting gaze motion using head rotation.

2.5 Eye origin

The eye origin is defined as the origin point of the gaze direction signal. The eye origin is calculated from the center of the eye, in this case defined as the midpoint of the iris.

3
Methods

This chapter describes how the work of this thesis has been carried out. It includes the details of how the filter functions were developed and implemented. It also covers how data for evaluation of the filter was gathered and how it was used for evaluating the performance of the filter functionality.

3.1 Software

The main programming language used in this thesis is C++. All of the filtering and detection algorithms have been developed in C++. To create the plots and test functions for evaluating the filter performance Python was used.

3.2 Data gathering

Three sets of data were used to develop and evaluate the performance of the filter. Each data set had its own specific purpose and was used to evaluate different aspects and functionality in the filter development. For each data set, the data was measured in the x-, y-, and z-direction. The following subsections will describe how each data set was gathered and which functionality each dataset was used to evaluate.

3.2.1 Manually annotated data for classification of motions

Gathering of data was done at Smart Eye with their recording software and hardware. One camera was used to collect raw gaze data. The gathering procedure was executed in such a way that the test subjects sat in front of the camera system and was given a predetermined sequence of points to look at.

In total, eleven videos were recorded with two different subjects performing the test sequences. Both of the test subjects were male in their 20's with normal vision.

The points, at which the test subject was looking at, was predetermined and their location is known. Each sequence was set up so that the subject could perform all the eye movements described in section 2.1. The recorded data was thereafter manually annotated to indicate between which frames the specified motion occurred.

The recorded data was used throughout the project for the development of the filter. The data is also used to evaluate the classification performance of the filter, the responsiveness of the filter at mode switching together with filtering performance during smooth pursuits and glances.

3.2.2 Prerecorded data with ground truth for fixations

To evaluate the filter on its performance during fixations, a prerecorded data set containing ground truth was used. The ground truth values were recorded by measuring points in 3D-space and then mapping them to the coordinate system used by the eye-tracking system. When the subject fixated at a target point, it indicated this by pressing a key corresponding to that target. The duration of the key press indicated how long the subject fixated at the target. This data set was prerecorded at Smart Eye using their measuring lab.

The data was also used to compare the filter performance against the raw data derived from Smart Eye's system as well as against the previous filter implemented in the system. The functionality of using head orientation as a gaze direction estimator was also evaluated using this data set.

The metrics that were used to evaluate the filter were trueness, the absolute deviation from ground truth value, and precision, the spread of the sampled values, as well as the availability which is defined as the percentage of recorded frames containing gaze direction data.

3.2.3 Synthetic data

Synthetic data was generated using a parametric model proposed by [13]. This proposed model estimates the saccadic waveform as the sum of a soft ramp function and a negated, time-shifted soft ramp function. The model is described by the equation:

$$s(t;\eta,c,\tau) = cf(\eta t/c) - cf(\eta(t-\tau)/c), \tag{3.1}$$

where $s(t)$ is the gaze direction in degrees,

$$f(t) = \begin{cases} t + 0.25e^{-2t} & t \geq 0 \\ 0.25e^{2t} & t < 0 \end{cases}$$

and the parameters, η, c, τ, are chosen based on the physiological properties of the human eye. The amplitude of the saccade is determined by the relationship between η and τ according to the equation:

$$A = \lim_{t\to\infty} s(t) - \lim_{t\to-\infty} s(t) \tag{3.2}$$

$$= \eta t - \eta(t-\tau) \tag{3.3}$$

$$= \eta\tau. \tag{3.4}$$

Hence, the amplitude of a saccade is the product of η and τ. The peak velocity is determined by the maximum value of $s^{'}(t)$ and occurs at $t = \tau/2$. This means that the parameter τ is chosen such that the duration of a saccade mimics that of a real saccade, which depends on the amplitude of the saccade and is calculated using equation (2.1). The parameter η is then chosen to control the amplitude, often chosen around $\eta = 750°/s$ [14].

The first metric that was evaluated, using synthetic data, was the performance of the filter during saccades. Measures, which were considered, were the responsiveness of the filter, i.e. how fast the filter could react to sudden changes and the classification performance, i.e. how well can the filter separate the different movements from each other. The responsiveness was measured by counting the number of frames the filtered signal lagged behind the true signal when mode switching occurred between the states: fixation and saccade. Further, in classification, the filter was evaluated on how it performs in terms of correctly classifying the events of fixations and saccades.

Further the filter was evaluated on its performance on reducing the noise in the signal. To evaluate the filter on its noise reduction capabilities, white Gaussian zero-mean noise processes with varying standard deviation were added to the generated signal in equation 3.1.The noise process in the real signal has not been determined to be zero mean and Gaussian. However, assumptions of the noise has been taken to meet these criterion since the Kalman filter assumes that the noise has these properties [8]. Therefore, the noise was assumed to be zero-mean and Gaussian when generating the synthetic signal.

3.2.3.1 Synthetic data for consensus gaze estimation

The performance of the consensus gaze filter was also evaluated using equation (3.1). Since in a real application both eyes do not always look in the same direction, the synthetic data for validation of the consensus filter was created with a small offset between the measured gaze direction for left and right eye.

3.3 Real-time eye motion and gaze filtering

In this section, details of the implemented real-time eye motion and gaze filter is described. The filter is based on a switching Kalman filter with two parallel filters. A constant velocity filter was used in both the filters to estimate the state of the gaze. What sets the filters apart is the process noise is much higher for one of the filters. This, in turn, makes one of the filters more reactive to measurements while the other filter has more trust in the process model. The switching functionality in the filter is based on a velocity threshold test together with a combination of the eye data for left and right eye. The filter has been developed to contain two states: fixations and saccades.

3.3.1 Constant velocity Kalman filter

For a constant velocity Kalman filter, both the direction and the velocity of the state is required for prediction and update of the filter. The state-space vector, described in section 2.2, is therefore stated accordingly:

$$u = \begin{bmatrix} x & y & z & \dot{x} & \dot{y} & \dot{z} \end{bmatrix}^T \tag{3.5}$$

where x, y and z defines the direction of the gaze and $\dot{x}$, $\dot{y}$ and $\dot{z}$ defines the velocity in each direction.

The measurement derived from the system consists of a 3D-direction vector. The measurement vector, described in section 2.2, is therefore:

$$w = \begin{bmatrix} \hat{x} & \hat{y} & \hat{z} \end{bmatrix}^T \tag{3.6}$$

and the corresponding measurement matrix is defined as follows:

$$H_k = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}. \tag{3.7}$$

The measurement noise for the filter is continuously estimated for the filter and will be described in the following chapter.

Both filters are using a constant velocity model. For fixations and smooth pursuit this has been shown to be a viable model in previous work [15]. For estimating the gaze during saccades in real-time it is problematic to model the motion according to equation (2.2) since it is dependent on knowing the amplitude of the saccade before it has occurred. Therefore the saccade motion has been approximated with a constant velocity filter which is more reactive to changes in the measured data. Since the visual focus during saccades is limited and therefore of less importance, the main goal of the filter is not to reduce the noise during saccades. However, what is important is that, during saccades, the filter is able to respond to these changes and follow closely to the changes in the signal.

The process model matrix used in the filter is defined as:

$$A_{k-1} = \begin{bmatrix} 1 & 0 & 0 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & 0 & \Delta t & 0 \\ 0 & 0 & 1 & 0 & 0 & \Delta t \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{3.8}$$

and the corresponding covariance matrix for this process model used is:

$$Q_{k-1} = \begin{bmatrix} \frac{\Delta t}{3}^3 & 0 & 0 & \frac{\Delta t}{2}^2 & 0 & 0 \\ 0 & \frac{\Delta t}{3}^3 & 0 & 0 & \frac{\Delta t}{2}^2 & 0 \\ 0 & 0 & \frac{\Delta t}{3}^3 & 0 & 0 & \frac{\Delta t}{2}^2 \\ \frac{\Delta t}{2}^2 & 0 & 0 & \Delta t & 0 & 0 \\ 0 & \frac{\Delta t}{2}^2 & 0 & 0 & \Delta t & 0 \\ 0 & 0 & \frac{\Delta t}{2}^2 & 0 & 0 & \Delta t \end{bmatrix} \tag{3.9}$$

where Δt is the time difference between the current frame and the previous. Due to that frames can be lost and/or measurements can be lost, the value for Δt is continuously updated. Hence both A_{k-1} and Q_{k-1} are continuously updated.

3.3.2 Noise estimation

As the noise of the system can change depending on several factors, including the position of the subject about the camera, the motion of the subject, contrasts in the eye for individual people etc., the noise levels are continuously changing. Therefore, to be able to use the filter for any subject independently of these factors, the noise level in the system is continuously estimated. The noise variance in the x-direction is estimated according to the following equation:

$$\sigma_{x,k}^2 = \frac{1}{N} \sum_{i=k-N}^{k} (\hat{x}_i - \bar{x}_i)^2, \tag{3.10}$$

where k is the current frame, $\hat{x}$ is the measured direction, $\bar{x}$ is the mean of the measured direction during the last N frames and N is the current number of samples in the noise buffer. The noise variance for the y- and z-direction is derived in same manner. The measurement noise covariance matrix is then continuously updated with the values from equation 3.10 according to:

$$R_k = \begin{bmatrix} \sigma_{x,k}^2 & 0 & 0 \\ 0 & \sigma_{y,k}^2 & 0 \\ 0 & 0 & \sigma_{z,k}^2 \end{bmatrix}. \tag{3.11}$$

The maximum number of samples used in the buffer is 30, which for a 60 Hz system is equal to 0.5 seconds. The re-estimation of the noise levels occurs during fixations. To prevent the filter from trying to estimate the noise levels during saccades or glances, a velocity threshold has been deployed on when to estimate the noise levels. If the velocity threshold is exceeded, the noise levels are not updated. This reassures that the re-estimation of the noise levels only occurs during fixations.

3.3.2.1 Flush buffer during saccade

When a saccade is detected, the noise queue buffer is flushed, which means that all values in the buffer are thrown away. This allows the filter to re-estimate the noise after a saccade event has ended. This, in turn, will give the filter a better estimation of the noise since the noise level might be different compared to the level before the saccade, and also because the average of the signal will be different compared to its value before the saccade.

Furthermore, when a saccade is detected, the covariance matrix of the low uncertainty filter is set to have the same covariance matrix as the high uncertainty filter. This allows the filter to be more responsive to changes during and immediately after a saccade, which helps when filtering the signal during post-saccadic oscillations.

3.3.3 Mode switching

To establish if a mode switch should occur, a switching test has been applied. This test is based on a velocity threshold test together with an eye data combination test. These two tests are executed sequentially, with the velocity test being the first one. For each eye, the velocity test is firstly executed and if that test is passed the

next one, which is the eye combination test, is executed. Only after both tests are executed and have established that a mode switch has occurred, the filter changes the mode. These two tests are implemented in order to reduce the number of falsely activated mode switches, hence making the switching more reliable and robust from false detections.

3.3.3.1 Velocity test

The velocity test was used for determining the start of a saccade or glance for each eye. As the start of a saccade is characterised by a velocity increase, this information can be used as a switch between the different modes.

The velocity test will detect a saccade if the velocity in one of the directions, from the low uncertainty filter, exceeds the 3σ-level, i.e. if the output of the filter is outside the 99.7% confidence interval of its posterior estimation. This test is applied for both eyes and later combined for accepting or rejecting the mode switching.

3.3.3.2 Combine information from both eyes

Information from each eye is derived separately and further filtered separately. The noise process for each eye has also been shown to be independent of each other. This information has been used in further reduction of false activation of the mode switching. As stated before, noise spikes in the data from each eye can trigger the velocity threshold in activating the saccade filter. By using the independence of the noise for each eye, in such that it is less likely that single noise spikes occur at the same time in both eyes, and the fact that the eyes move in a synchronous manner, noise spikes which can trigger the saccade filter are reduced by comparing the data for each eye and only go into the saccade mode if both eyes indicate that a saccade is present. This uses the principle of conjugate eye movement, which can be affected by different diseases [16]. However, this is not part of the scope of this project and will not be considered in this report.

3.3.4 Head orientation for gaze estimation

In situations where there is no gaze data available, such as when the person blinks, the person is not facing the camera or image frames have been dropped during processing, the filter should still be able to produce estimations of the gaze direction. When presented with a situation where data from the eyes is not available, the filter will continue to perform filtering by using the head orientation to estimate the gaze direction. To estimate the gaze direction by using the head orientation, a gain factor is continuously estimated for each direction. The gain factor contribution in the x-direction is estimated according to:

$$H_c = \frac{\sum Gaze_x}{\sum Head_x}. \tag{3.12}$$

The gain factor is estimated for the other directions in the same manner. This gain factor, H_c, then shows how much the subject turns its head in relation to how much the gaze changes. This factor could then be incorporated into the Kalman

filter for estimating the direction of the gaze using the head orientation. By continuously estimating a gain factor, the gaze direction can be estimated using the head orientation. The gaze direction will be estimated solely on head orientation by calculating this gain factor when faced with situations where gaze data is not available.

The procedure is that the estimated gain factor for the head is combined with the head orientation in the update step as a measurement to the Kalman filter and, in such, still produce estimates of the gaze even when presented with a situation where no gaze data is available.

This also ensures that the Kalman filter does not need to be reinitialised after it has gone a long time without input data.

3.3.5 Flow chart of the filter

Figure 3.1 shows a schematic view of the entire filter for each eye. The figure shows that the filter makes a prediction based on the output from the previous iteration, like a normal Kalman filter. The filter then checks if there is a measurement available. If not, then the update step is done using the gain factor from the head. If a measurement is available, then the gain factor is updated and the filter moves on to performing the velocity test. Based on the outcome of the test, the update step is either performed using the low uncertainty filter or the high uncertainty filter. The process is then repeated for the next iteration.

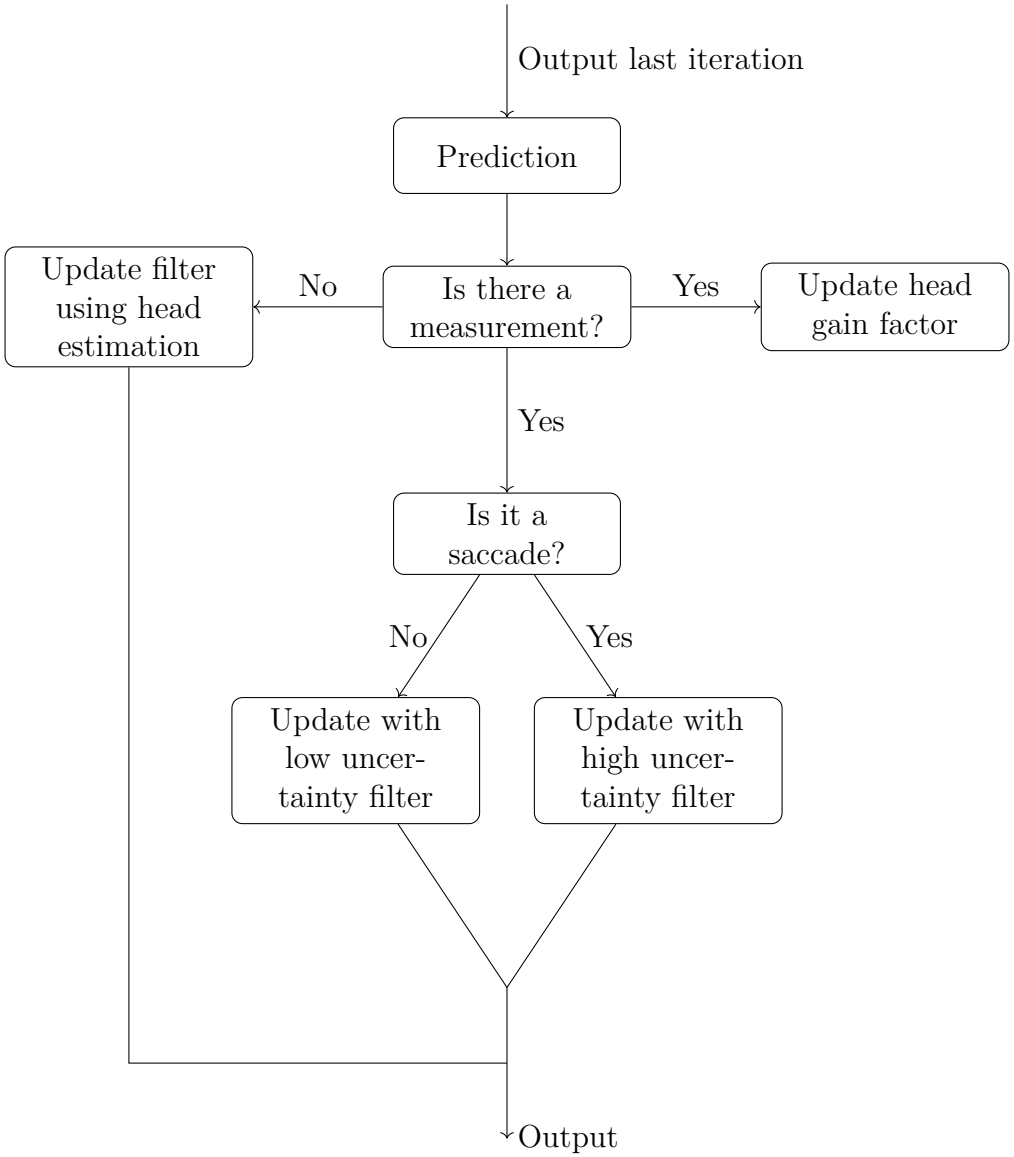

Figure 3.1: Schematic view of the different steps in the filtering of the gaze direction.

3.4 Estimation of the consensus gaze

The estimation of the consensus gaze, the combined gaze direction for both eyes, is based on the same type of Kalman filter as for the individual eyes, i.e. a switching Kalman filter which consists of two constant velocity Kalman filters. The first filter has a fairly low process noise while the second filter has a much higher process noise. The filters developed are tuned so that, just as for each eye, the implemented filter can reduce the noise in the signal as well as react to sudden changes in the signal. The high process noise filter, which is used to detect and follow fast changes in the signal, is triggered when a saccade is detected. The switching of filter occurs when each eye states that a saccade has occurred simultaneously, as described in section 3.3.3.1. Hence, this means that the consensus filter never detects saccades by itself,

but rather makes use of the information provided by the individual eyes.

Estimation of the consensus gaze was done by combining the measured gaze for the left and right eye. The two measurements are combined by using a Kalman filter with two inputs instead of just one input. This means that the update of the Kalman filter includes twice as many measurements as that of the individual eyes. The measurement noise for the consensus filter was set by copying the estimated noise for the individual eyes, i.e., the estimated noise levels for each individual eye are used in the consensus filter.

If one of the eyes is missing data, then the measurement noise is increased for the corresponding measurements, which means that the consensus filter will practically disregard those measurements. If data is lost for both eyes then the consensus gaze will be estimated using the position and rotation of the head, which is described in section 3.3.4.

3.4.1 Estimation of the consensus origin

Unlike the estimation of the consensus gaze, estimating the origin of the consensus gaze is done by finding the midpoint between the center of the eye, described in section 2.5, for each individual eye in 3D-space. The signal is then filtered using only a low process noise constant velocity Kalman filter.

4

Results

The results from the implemented filter will be evaluated and analysed in this chapter. The performance of the filter will be both analysed on measured and synthetic data, to visually illustrate its performance, and on ground truth data to evaluate the accuracy of the filter quantitatively.

The plots shown in the coming section are recorded using a 60 Hz camera system at Smart Eye. All results have been translated from x-, y- and z-direction to heading and pitch for illustration purposes. For more detailed information on how heading and pitch are defined, see appendix A.

4.1 Filtering

In figures 4.1 and 4.2, a sequence of gaze direction motions including saccades, fixations, and glances are presented for heading and pitch, respectively. The figures include raw gaze data as well as the filtered signal from the developed filter.

As can be seen in the figures, the filter reduces the noise levels in the signal and is at the same time able to quickly respond to changes in the input signal.

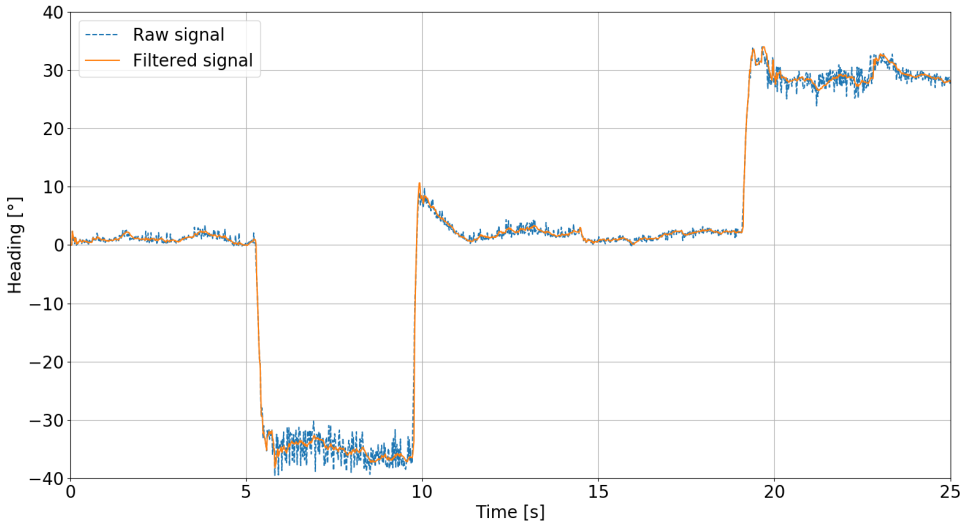

Figure 4.1: Filterered heading together with raw gaze signal during a sequence with fixations and saccades for the left eye.

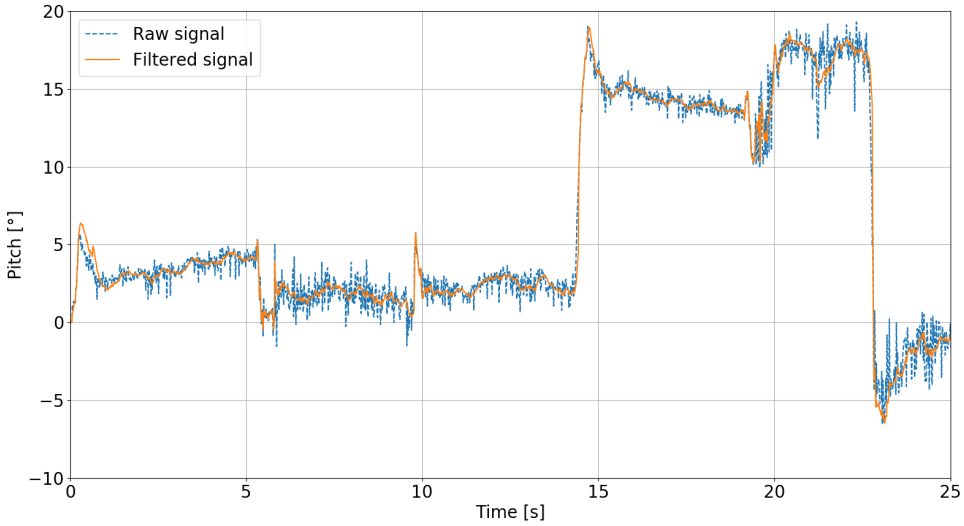

Figure 4.2: Filterered pitch together with raw gaze signal during a sequence with fixations and saccades for the left eye.

4.1.1 Filtering during fixation

Figure 4.3 shows the performance of the filter during a fixation in heading for the left eye. The figure is illustrative of how the filter reduces the noise of the signal. The figure illustrates how the variation in the raw signal has been reduced by filtering the signal. Although not shown, the noise has been reduced with similar results for pitch.

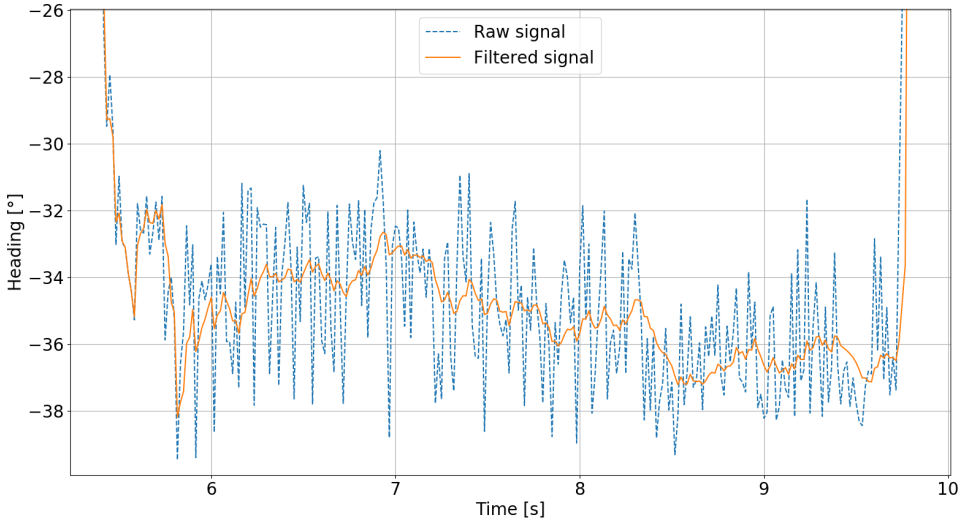

Figure 4.3: Filtering in the heading rotation during a fixation for left eye.

Because of the simplicity of tracking the gaze direction during fixations, the

performance of the filter has been evaluated on its trueness and precision. The evaluation was done by measuring the distance between the filtered signal and the ground truth state, see section 3.2.2. The results were then compared to the trueness and precision of the raw gaze signal. In tables 4.1, 4.2, and 4.3, the results are summarized for each respective eye, without any compensation from the head orientation. The values in the tables depend on the dataset used. This particular dataset is from a driving application and is therefore particularly difficult to extract data from. It is therefore more interesting to study the relative values between the raw and filtered data, and not the absolute numbers.

What can be seen in the tables below is that the filter increases the precision of the signal for each eye compared with the raw signal. The precision increase is around 36%, 36%, and 24% for the left, right and consensus gaze, respectively. On the other hand, there seems to be a small decrease in trueness performance when filtering the signal. For the right and consensus gaze, the offset from the true value has increased, but for the left eye, the offset from the true value has decreased. The deviation in trueness in tables 4.1, 4.2, and 4.3 can partly be explained by an offset in the measurement data from the ground truth values.

Table 4.1: Trueness and precision for the left eye without head pose correction.

Signal	Left Eye Raw	Left Eye Filtered
Trueness [°]	5.012	4.982
Trueness Heading [°]	3.378	3.311
Trueness Pitch [°]	2.902	2.908
Precision [°]	1.974	1.265
Precision Heading [°]	1.370	0.8664
Precision Pitch [°]	1.342	0.8448
Availability [%]	81.0	83.9

Table 4.2: Trueness and precision for the right eye without head pose correction.

Signal	Right Eye Raw	Right Eye Filtered
Trueness [°]	4.222	4.901
Trueness Heading [°]	2.407	3.059
Trueness Pitch [°]	2.947	3.147
Precision [°]	1.985	1.274
Precision Heading [°]	1.379	0.8658
Precision Pitch [°]	1.352	0.8617
Availability [%]	83.0	85.9

Table 4.3: Trueness and precision for the consensus eye without head pose correction.

Signal	Consensus Eye Raw	Consensus Eye Filtered
Trueness [°]	3.813	3.919
Trueness Heading [°]	2.159	2.252
Trueness Pitch [°]	2.646	2.695
Precision [°]	1.587	1.208
Precision Heading [°]	1.108	0.8798
Precision Pitch [°]	1.077	0.7350
Availability [%]	94.3	96.6

In figure 4.4, pitch rotation for the filtered signal together with the raw signal and ground truth value has been plotted. As can be seen, the raw signal has an offset from the ground truth. Since the filtered signal should follow the mean of the raw gaze signal, this results in an error when filtering the signal which could explain the worsened values in trueness for the filtered signal.

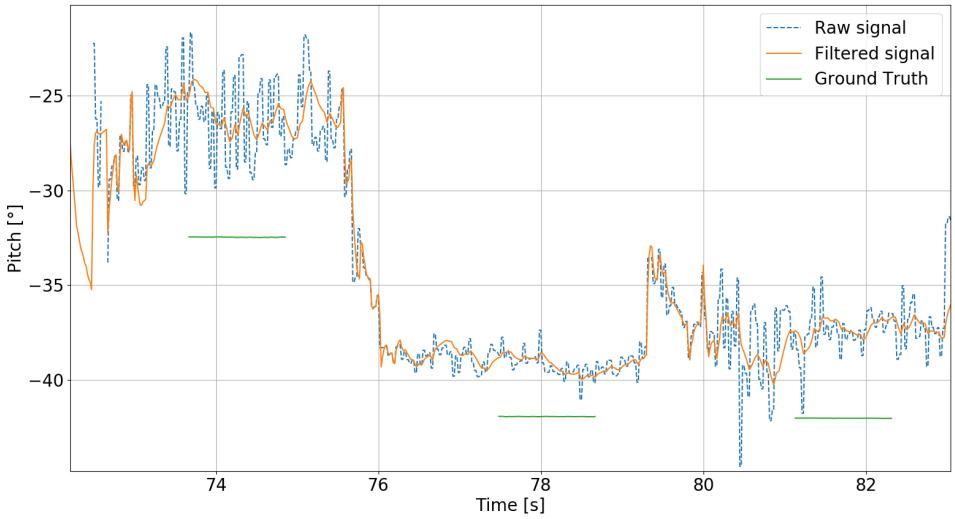

Figure 4.4: Filtered and raw pitch signal together with ground truth value for the right eye.

In figure 4.5, the reduction of the standard deviation compared to the theoretically added noise. The reduction was evaluated using synthetic data during fixations. The figure illustrates that the filter has a smaller effect on low noise levels, which is expected, but the filtering has a dramatic impact as the noise levels grow larger. For instance, it can be seen from the figure that for a theoretical standard deviation of close to 1, the filter reduces the noise level by around 85%.

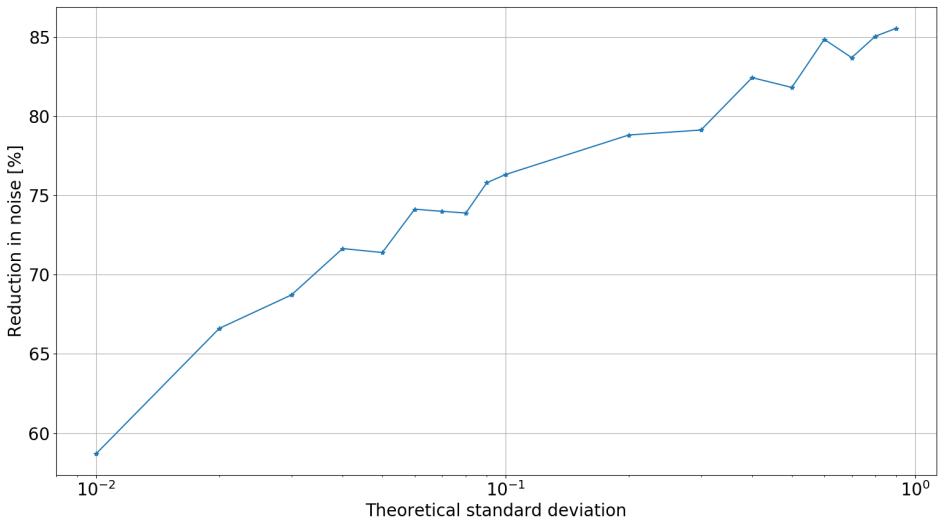

Figure 4.5: The reduction of the standard deviation evaluated for different noise levels during fixations. The reduction of the standard deviation was evaluated using synthetic data.

4.1.1.1 Head orientation estimates during fixations

By adding head orientation as indicator for gaze direction, the availability of the gaze direction can be increased substantially. Since the head orientation is almost always available, the filter reaches close to 100% availability when including the head orientation during fixations. Which is an increase from the values obtained in tables 4.2 and 4.3.

Figure 4.6 shows the result for the gaze filtering of the right eye for when no measurement of gaze signal is available and the filter uses the head orientation for estimation of the gaze direction. As can be seen, using the head orientation is feasible for the estimation of the gaze direction as described by [11]. The estimations work properly, the head turns mostly in the same direction as the gaze, but it is subject to an offset. The offset is estimated to be around 5°, but it can vary up to 25°. This offset has yet another uncertainty associated with it: the raw signal from the gaze direction has already an offset from the ground truth which in turn influences the gain factor associated with the head estimation.

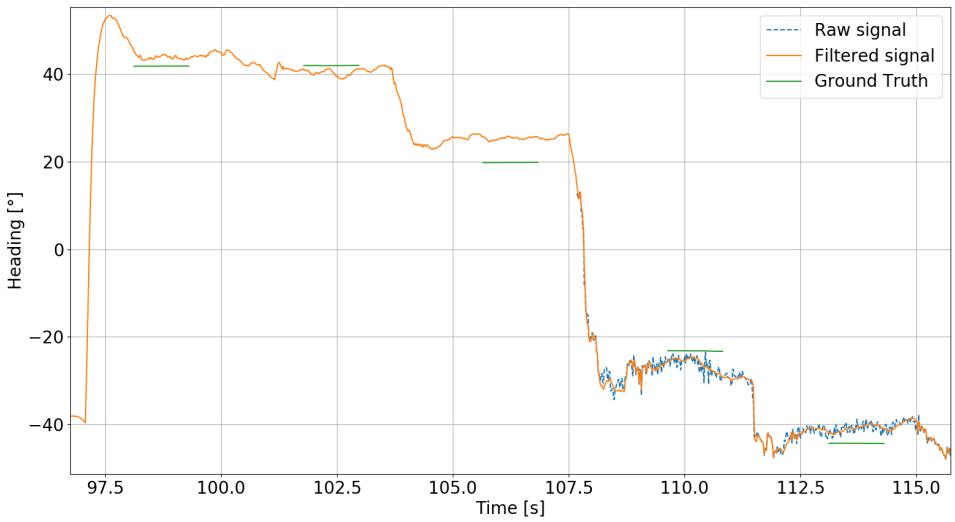

Figure 4.6: Filtered heading rotation for the right eye using the head rotation for estimation when no gaze data is available.

4.1.2 Filtering during smooth pursuit

During smooth pursuit, the filter behaves as expected and as shown in previous work [15]. The constant velocity Kalman filter filters the data, reduces the noise levels, and is at the same time able to follow the signal accurately. The filtered signal can be seen in figure 4.7 together with the raw gaze signal.

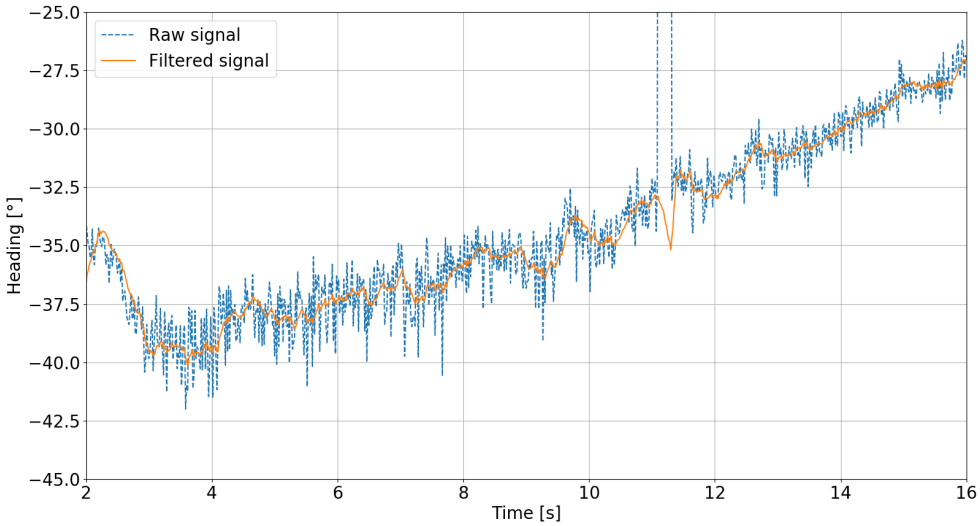

Figure 4.7: Filtered gaze for the left eye during a smooth pursuit.

Unfortunately, smooth pursuit is very difficult to be evaluated in terms of trueness and precision in the same way as for fixation. This is due to the fact that it

is hard to determine where the true state lies during a smooth pursuit. But for precision, it can be visually seen that the filter reduces the variance of the raw gaze signal.

4.1.3 Filtering during saccades and glances

In figure 4.8, the performance of the filter is shown during two saccades. As can be seen, the filter follows the raw gaze signal closely during both saccades and is highly reactive at the beginning of both saccades. During a saccade, the higher uncertainty filter is applied to the signal which makes the filtered signal follow more closely to the measured signal. Since the visual perception during saccades is limited, discussed in section 2.1.3, filtering during saccades are of less importance. Instead, the importance lies in following changes in the signal.

In figure 4.8, it can also be seen that the filter can quickly adapt to the fixation occurring directly after the first saccade sequence and start filtering with the low uncertainty filter once again. After the second saccade, a small overshot of the filtered signal can be observed, this occurs when the filter fails to classify the ongoing saccade properly and starts to filter with the low uncertainty filter too early after a saccade.

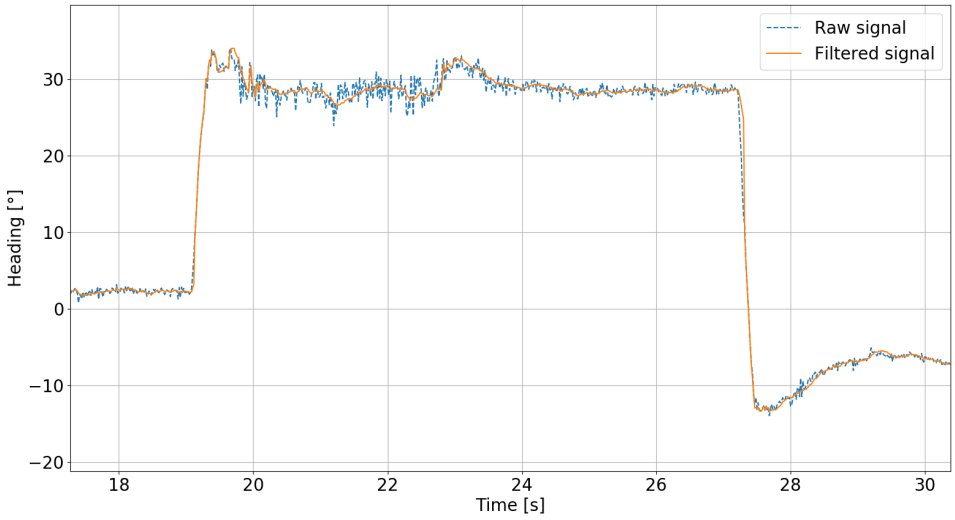

Figure 4.8: Filtered gaze direction versus raw gaze data for the left eye heading rotation during two saccades.

4.1.4 Consensus filtering

In figure 4.9, the estimated consensus heading rotation, using the implemented Kalman filter, is shown together with the derived consensus gaze heading implemented in Smart Eye's system. As can be seen, the filtered signal is better in terms of precision than the derived raw signal, coinciding with the results in section 4.1.1. However, the way how the consensus direction is derived, differs for the respective

signals and thus resulting in different direction outputs. As can be seen in figure 4.9, around frame 1100, the filtered consensus signal does not follow the raw signal for consensus. In this plot, the filtered consensus signal lies closer to the ground truth value but the offset varies and the filtered consensus could just as well lie further away.

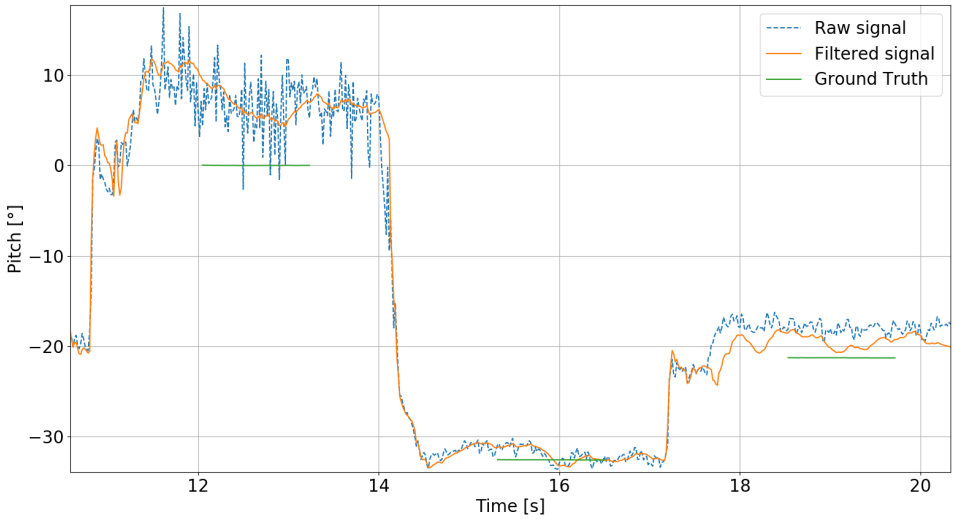

Figure 4.9: Filtered consensus signal together with raw signal for consensus derived from Smart Eye's system.

4.1.5 Origin filtering

The results for the origin filtering, in figures 4.10 and 4.11, are shown in the x-direction because both the heading and pitch rotations experienced discontinuities and were difficult to interpret.

The origin for the left and right eye was filtered in the same manner as their respective gaze signal. This was done in order to not create any latency between the two signals since they are both needed in other applications. The results, which are shown in figure 4.10, showed that filtering the origin signal is quite well described by a constant velocity Kalman filter since the signal changes relatively slowly. The origin filter also contained a higher uncertainty filter but its presence is not as prominent since the raw signal for the origin had much less noise, to begin with.

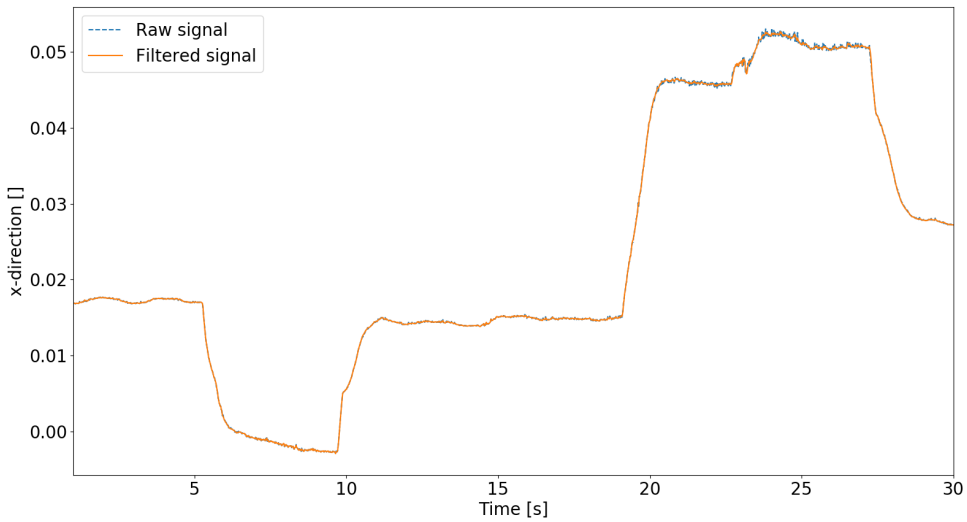

Figure 4.10: Filtered origin signal together with raw signal in the x-direction for the right eye.

As described in section 3.4.1, the method used for the respective eyes was not feasible to apply when filtering the origin for the consensus gaze. Instead, the filter is based on using a constant velocity Kalman filter to filter the midpoint between the two eye centers in 3D-space. The results are shown in figure 4.11.

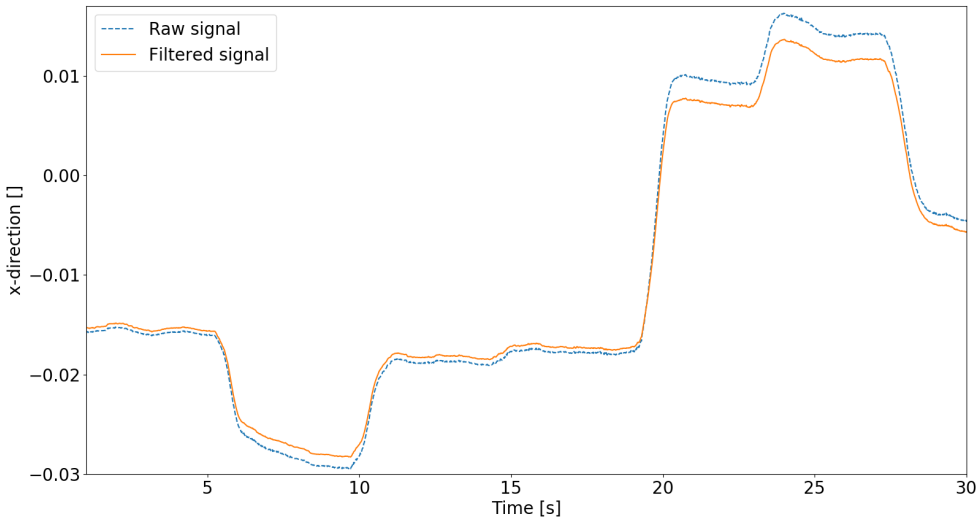

Figure 4.11: Filtered origin signal together with raw signal in the x-direction for consensus derived from Smart Eye's system.

Figure 4.11 shows that the filter follows the overall trend of the raw signal for the consensus origin. As can be seen in the figure, there tends to be a small offset

between the raw and the filtered signal. However, it should be noted that the raw origin signal is calculated by weighing together the origin signals from the left and the right eye, so this means that it could potentially be the raw signal which contains the offset.

4.2 Classification

By using information from both eyes, falsely triggered saccades have been reduced. Noise spikes in the data can trigger the velocity threshold to initiate the high uncertainty filter to operate on the data, which in turn lowers the precision of the filtered signal. By making use of the fact that each eye has independent noise, the number of falsely triggered saccades has been reduced. Since noise spikes happen randomly in each eye they are less likely to occur simultaneously in both eyes at the same instance. So for the high uncertainty filter to be triggered, both eyes has to declare that a saccade has occurred simultaneously, otherwise the saccade is rejected.

In figure 4.12, the activation points, marked with purple stars, of the high uncertainty filter is shown when only using information from one eye for classification of a saccade. The vertical lines in figure 4.12 indicate the start and the end of the manually annotated saccadic sequence. As can be seen during the saccade, the classification is correct but during fixations, the noise can sometimes be wrongly interpreted as the start of a saccade. In figure 4.13, the same sequence is plotted and the classification of saccades is marked with purple star. Unlike the first plot, this plot combines information from both eyes in determining if a saccade has occurred. As can be seen, the number of instances where falsely activated saccades have occurred has been reduced during fixations. Hence, the classification algorithm can be improved by combining information from both eyes.

Figure 4.13 shows that the high uncertainty filter has been activated a few samples after the saccade has ended. This happens due to the fact that the noise buffer, which estimates the noise in the system, has been emptied after a saccade. The noise buffer is emptied because the noise levels after a saccade can be vastly different from the values before the occurrence of the saccade. By emptying the noise buffer after a saccade the filter becomes more reactive, which in turn can handle PSOs and at the same time is able to quickly estimate the new noise level.

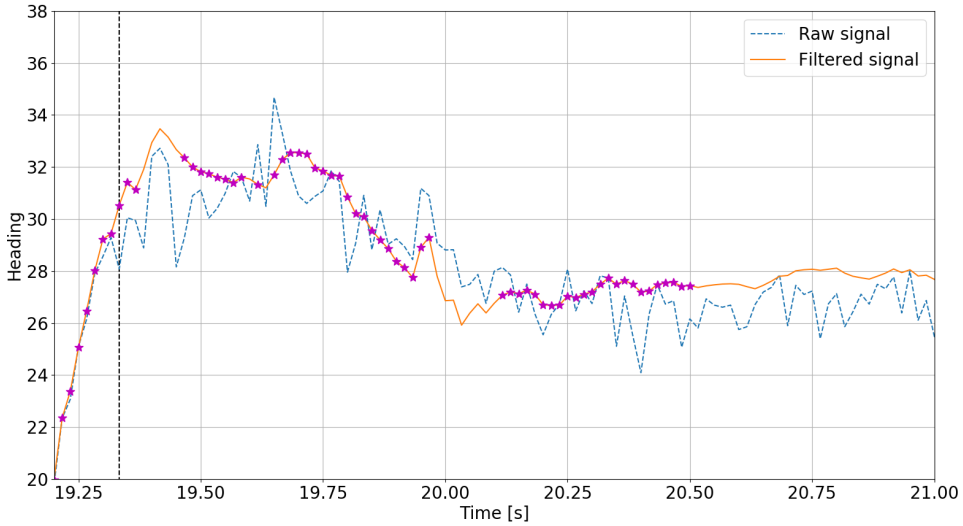

Figure 4.12: Saccade classification when only using information from one eye.

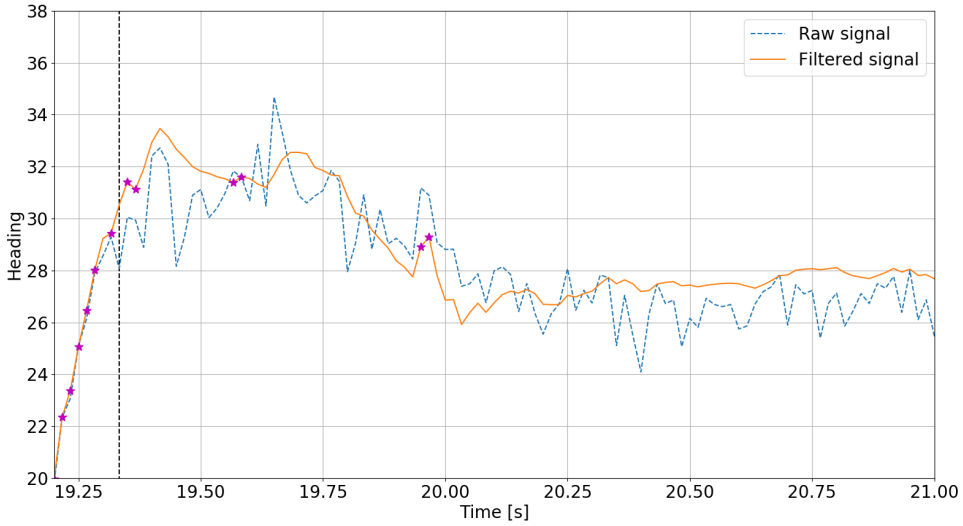

Figure 4.13: Saccade classification when using information from both eyes.

In table 4.4, the classification accuracy of the filter when using information from both eyes is shown. As seen in the table, the classification of fixations, which is indicated by the fraction of true negatives, is quite accurate, with 97.8% correct classifications. For saccades, the classification accuracy is much lower, approximately 59.3%. The lower accuracy during saccades is due to many contributing factors, which will be discussed more in-depth in the following chapter.

Table 4.4: Confusion matrix for classification accuracy. A positive outcome indicates a saccade.

	Actually positive	Actually negative
Predicted positive	59.3%	2.2%
Predicted negative	40.7%	97.8%

4.2.1 Responsivity

The responsivity of the filter can be quantified by measuring the latency of the filter during a saccade. The latency can be calculated by measuring how many frames the filter lags behind the true state. Firstly, this was measured using synthetic data with no added noise. The test showed that the filter lagged behind the true state by approximately 4.1 frames, when simulating saccades for a 60 Hz system. Secondly, the responsivity was measured using one manually annotated recording and the result showed that the lag in this case was approximately 3.6 frames.

4.3 Comparison with the previous filtering method

Figure 4.14 shows the pitch rotations for the implemented filter together with the previously implemented gaze filter in Smart Eyes system. The previously implemented filter is based on spherical linear interpolation [17]. As can be seen, the filters perform approximately equal in terms of precision and trueness. Both filters successfully reduce the noise levels of the raw gaze signal by roughly the same amount and have approximately the same mean.

Due to the usage of head orientation, the availability has been increased for the implemented Kalman filter compared to the previous filter. The availability of the implemented Kalman filter, with head orientation, lies around 100.0%, for the left eye. For the previous filter, the availability is the same as for the raw gaze signal stated in table 4.1.

Apart from the increase in availability by using the Kalman filter, another benefit comes from that the implemented Kalman filter also can classify the ongoing eye movements, which the previous filter does not.

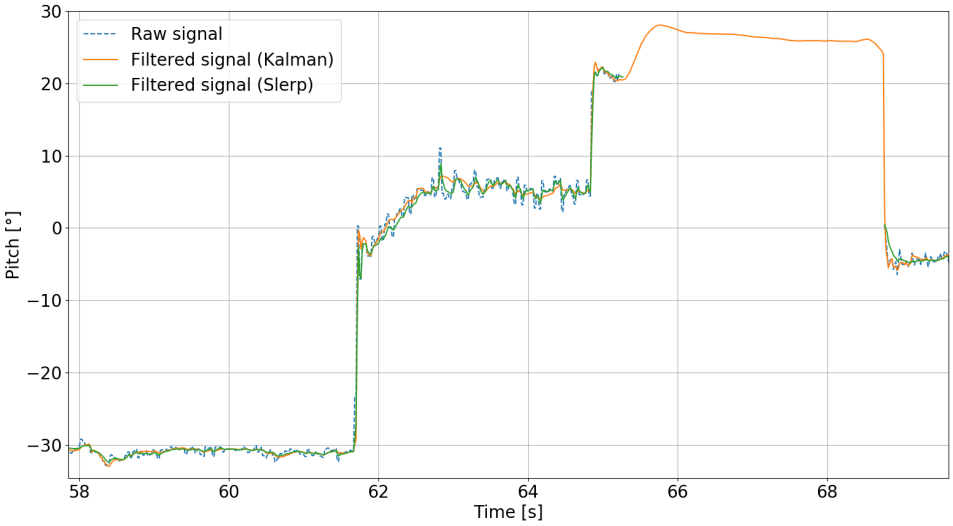

Figure 4.14: Comparison between the implemented Kalman filter with the previous gaze filter.

5
Discussion

The following chapter will discuss the results from the implemented filter. The chapter will include a discussion of other potential methods for filtering and classification. Further, future improvements or modifications of the filter will be addressed.

5.1 Filtering

The following section will discuss and interpret the results shown in section 4.1.

5.1.1 Fixations

The results in section 4.1.1 show that the switching Kalman filter reduces the noise variance significantly from the measurements during fixations. As shown by the results and stated by Komogortsev [15], a constant velocity model is a viable model to use during fixations, which can be oberserved visually in figure 4.3 and numerically in tables 4.1, 4.2, and 4.3.

Using the head orientation has also been seen to be a viable estimate for the gaze direction since the result against ground truth data gives reasonable estimates. By using the head orientation for gaze estimation, when no gaze direction signal is available, the availability of the signal has been increased. This signal is on the other hand associated with a higher uncertainty and as stated by Stiefel-Hagen and Zhu [11] there are motions where the head orientation and gaze direction moves in non correlated motion, which would lead to bad estimates of the gaze direction. The result shown from Land [12] and Stiefel-Hagen and Zhu [11] shows that there is a correlation between gaze direction and head orientation and their respective motions.

Although the noise in the gaze direction has not been investigated to be additive white Gaussian noise, which is a condition for the Kalman filter according to [8], the filter still shows promising results in reducing the noise levels. Further investigation of the noise process in the gaze direction could lead to better performance in filtering the signal and is something that could be done in future work.

5.1.2 Smooth pursuit

Filtering during smooth pursuit is very important when trying to estimate the gaze direction of a person. Smooth pursuit makes up a large part of “ocular activity” [18] and can occur frequently in a driving application.

In figure 4.7, we can see that the implemented filter reduces the noise and is able to follow the underlying trend of the signal fairly well. There are of course some instances where the filter classifies the signal as a saccade when the actual movement is a smooth pursuit. But as catch-up saccades can occur during smooth pursuit these classifications are not always incorrect.

As stated in section 4.1.2, it is difficult to give a quantitative number on the filter performance during smooth pursuit, but from a visual inspection, it appears that using a constant velocity Kalman filter works well, which was also established by Komogortsev [15]. This technique works well since the movement, during a smooth pursuit, is relatively slow and does not involve any sudden changes.

5.1.3 Saccade

Filtering during saccades has been limited in the implementation of the filter. Since the visual ability during saccades is restricted during saccades, this is of less importance, and instead, the importance lies in following the changes in the signal.

During saccades, the high uncertainty filter is applied to the signal, which reduces the filtering ability of the filter but increases the reactivity of the filter. This is to ensure that the filtered signal follows changes in the measured data. To increase the reactivity even further, the noise buffer - containing the estimated measurement noise - is emptied. The emptying of the noise buffer also leads to better estimates of the noise levels after a saccade, since the noise can be vastly different. Further benefits by emptying the noise buffer is that PSOs, often occuring after saccades, are more easily filtered. The downside with emptying the noise buffer is that the filter performs worse in reducing the noise directly after a saccade, which lowers the precision for a few samples after a saccade.

The saccade motion filtering could be further improved by having a better process model for saccades. The models looked at in this project were dependent on knowing the amplitude of the saccade in advance, which leaves them not suitable for real-time filtering.

Previous attempts at estimating a saccade motion model has been made by [2]. Their attempt was built on a smoothing process, where they recalculated the state of the signal a few frames back in time when a saccade was detected. The result of their filter during saccades was comparable to the results in this report by using a high uncertainty, constant velocity filter for saccades. The conclusion from this is that since the filtering during saccades is of less importance, using a constant velocity filter with high uncertainty is sufficient during saccades.

The velocity threshold test used in this project functioned as intended. The filter is fast in detecting saccades and adapting to changes in the measured signal. The lag for the filtered signal, compared with the measured signal, was shown in section 4.2.1 to be 3.6 frames for recorded data and 4.1 frames on synthetic data. The threshold could have been lowered to increase the responsiveness of the filter, but this would lead to more falsely classified saccades and reduced quality in filtering.

5.1.4 Consensus filtering

The filtering of the consensus gaze direction differs from how the gaze direction is filtered for the individual eyes. The consensus gaze direction uses the raw signal for each eye as its measurements. This is built on the fact that both eyes are looking at the same target. If that is not the case, the consensus gaze filter could potentially experience difficulties in estimating reliable data for the gaze direction. This has not been the scope of this report and would have to be studied further if the filter were to be applied to any subject.

An illustration of the filtering done by the consensus gaze filter is shown in figure 4.9. This shows a sequence containing both fixations and saccades, and the filter does a fairly good job of estimating both of them. This is further strengthened by the results shown in table 4.3. However, from the figure we learn that an offset could appear between the raw and filtered consensus gaze signal. This is something that was not seen for the individual eyes and arises from the fact that there is no direct estimation of the consensus gaze. Rather it is based on the signal from the individual eyes and if the raw and filtered signals are processed differently, these offsets can occur.

The saccade detection for the consensus gaze is solely built upon the detections made by the left and right eye. Hence, the consensus gaze does not make any detections of saccades of its own. It uses information already provided from data processed by the individual eyes in determining if a saccade has occurred. Therefore, the consensus-filtered gaze does not provide any new information about the state of the filter. The reason for this is that the consensus filter would not make a better estimation of when a saccade occurs since it uses the same inputs as the individual eyes, but rather mixes it which means it might lose some of the independence of the eyes' movement. The other reason is that it would seem unreasonable if the filtered signal for the individual eyes performed a saccade while the filtered signal for the consensus gaze did not, or the other way around. Therefore, the most logical assumption would be to use the saccade detection done by the individual eyes.

5.2 Origin filtering

The origin, for the individual eyes, was filtered the same way as the gaze direction - a switching Kalman filter with two models with different process noises. Although the origin's movement much more resembles that of a constant velocity process and does not contain any quick saccades compared to the gaze direction, the reason for using the same type of filter as for the gaze is that the origin and gaze are correlated. This means that if the origin is filtered more slowly than the gaze, then the origin runs the risk of lagging, which could lead to an incorrect estimation of where the eyes are looking.

The result of the filtering in the x-direction for the right eye is shown in figure 4.10. The figure shows that the origin signal is well described by the implemented filter. The raw signal is fairly well-behaved, to begin with, but as was mentioned before, it is important that the origin and gaze are filtered in the same way.

The result from filtering the consensus origin, shown in figure 4.11, indicates

that the signal is well described by the implemented filter. Here it can be seen that a small offset sometimes occurs between the raw and filtered signals. As mentioned in section 4.1.5, this offset arises from the fact that the two signals are processed differently. Since it is not possible to directly estimate the origin of the consensus gaze, it is very difficult to say which one is correct. However, the filtered signal uses the center of the eyes as its basis for calculating the origin consensus. This increases the availability of the signal because the eye centers have a higher availability because they are anchored to the head position. Even if the filtered signal has an offset, this could probably be tolerated since it is a very small error and at the same time the availability is increased to almost 100 %.

5.3 Classification

The classification of the different modes in the system is based upon the switching between the two parallel filters implemented. The classification can separate fixations from saccades. However, the implemented filter is not able to classify events such as smooth pursuit. This was a deliberate limitation of the filter because of the inconclusive definition of a smooth pursuit [3]. Smooth pursuit is therefore included in the fixation classification, since they are described by the same motion model in the filter.

During fixations, the algorithm gives satisfactory results in terms of classification. A 97.8% correct classification of fixations was observed. For saccades, the classification is somewhat less accurate. The accuracy of correct classifications during saccades has been established as 59.3%. The lower accuracy during saccades has some implementation dependencies and restrictions due to the nature of the filtering process. Firstly, the saccade detection is based on a velocity threshold. As soon as the filter has detected a saccade, the velocity estimate of the filter is increased together with its uncertainty. By increasing the uncertainty of the estimate during saccades this threshold becomes increasingly harder to exceed, thus leading to fewer saccade detections. Furthermore, since the saccades usually occur during a small number of frames, the responsiveness of the filter has a big impact on the classification accuracy during saccades. This in turn leads to lower accuracy of the classification during saccades.

Since there is no clear definition where the start and stop of a saccade occur there is also some biasing in the annotated data which affects the outcome of the classification accuracy. The annotation of the data has been done manually from video sequences to define where the saccades occur. Since there is no clear definition of how to annotate the data, the classification results should be interpreted with caution.

5.3.1 Responsivity

In this project, responsivity measures the latency of the filtered signal compared to the raw signal at the beginning of a saccade. For the implemented filter, the latency was measured to be around 4.1 frames and 3.6 frames for synthetic and measured data, respectively. Since the duration of most saccades is longer than the measured

lag, this is not a problem in terms of filtering, because, as was discussed before, a person's visual perception is virtually zero during saccades and the high uncertainty filter quickly adapts to the new measurements. However, the responsivity of the filter has an impact on classification during saccades. If the filter were to react faster on saccadic changes in the signal, the classification accuracy of the filter during saccades could have been increased. To make the filter more reactive the threshold for activating saccades has to be decreased. This, on the other hand, would lead to more falsely classified saccades during fixations. So there is a trade-off between fast responsivity and correct classification in the implemented filter.

5.4 Combining information from both eyes

Combining the signals from the left and right eye has proven to be useful both for noise reduction and classification. Since a saccade is triggered simultaneously for each eye and the noise is estimated to be independent, the filter was able to reduce the sensitivity to independent noise spikes by comparing them to each other.

Combining information from both eyes also has a direct impact on the classification accuracy. Since the filter only concludes that a saccade has occurred if both eyes have declared a saccade simultaneously, the number of falsely classified saccades has been reduced. The downside of having a more stringent test for saccades is that it could increase the latency during actual saccades. However, this has not been observed to have a big impact and the filter performs comparably to each individual eye in terms of classification during saccades.

Using information from both eyes can be limited if the test subject has some sort of illness which reduces the simultaneous motion of the eyes, as stated by [16]. This would reduce the functionality of combining information for the eyes and is something that would have to be further investigated.

Further, the head orientation data could be used in the detection of saccades. As stated in [11] and [12] there is a correlation between head orientation and gaze direction. The head orientation could be combined with the information derived from the eyes to determine and classify the motion of the gaze.

5.5 Offset in ground truth data

As mentioned in section 4.1, there is an offset from the ground truth data in the raw gaze signal. This phenomenon has affected the final result of the filter. This has the most apparent effect on the trueness of the filtered signal as the filtered signal gets worse performance than the unfiltered signal. Since this offset exists, it is difficult to draw any conclusions about the trueness values in the filtered signal. It could be that the filtered signal performs worse in trueness, but since the filtered signal, by visual inspection, lies close to the mean value of the raw gaze signal this is very unlikely. Another factor that could affect the trueness is how it is calculated. The trueness value used in this report is measured as the absolute distance from the ground truth. This means that if the filtered signal has the same mean as the raw signal, then they would get the same trueness value, even though one signal

contains much more noise than the other. A more precise measure of the trueness would be the root mean square error, which would benefit the less noisy signal.

5.6 Comparision with other filter methods

Compared to the previously implemented filter in Smart Eye's system, the filter implemented in this project has comparable results in terms of precision and trueness. Both filters successfully reduces the noise levels in the signal and has at the same time roughly the same mean, as seen in figure 4.14. What speaks in the advantage of the implemented Kalman filters is its ability to predict the state of the gaze direction, even when no measurements are present. The Kalman filter can do this by simply continuing with the prediction step, when no measurements are available. In this project the head orientation has been used as an estimate for gaze direction when the gaze is not available. This functionality could also be implemented in the previous filter implemented and is something that could be investigated in future work.

5.7 Future work and general filter performance discussion

The implemented filter successfully reduces the noise levels in the raw gaze signal and is at the same time able to classify the states for fixation and saccades. What could be done in the future is to make the filter able to classify smooth pursuits as well. For this to function, a clear definition between the different motion of the eye has to be established.

As the head orientation has been seen to be a viable estimate for the gaze direction, further investigation into developing a gaze estimation from head orientation and how this could be included in a filtering process is of interest.

Future work which also is of interest is to see a better process model for saccades developed. A better saccade filter could be implemented in the method developed in this project but more interesting would be to implement it in an IMM filter. IMM stands for "Interacting Multiple Models" and is based on the same idea as the switching Kalman filter, but weighs the output from the different filters instead of choosing one of them [19]. The benefit of using an IMM filter is that there is no need to have a threshold-based test since the weighting of the filters is done automatically. If a better saccade model is applied to an IMM filter, then potentially, the filter could separate the different modes from each other, leading to a reduced latency during the transition between different modes.

6
Conclusion

During this project, a filter has been developed for the filtering of eye-movements and gaze directions. The final result shows a fully functioning filter that can operate in real-time. The filter works for the test data recorded and provided by Smart Eye AB. The filter performs and fulfills its initial purpose of filtering the data to reduce the noise in the output as well as improving the precision of the signal.

Due to the offset in the ground truth value, no concrete conclusion can be drawn about the trueness of the filtered signal, but from visual inspection of the filtered signal, it can be seen that the filtered signal follows the overall trend of the gaze direction signal, which can be seen as an indication that the trueness has been improved.

In terms of classification, the filter has a high accuracy in determining fixations correctly. For saccades, the classification accuracy is lower, which depends on implementation and some restrictions in the filtering process, but is still shown to be reliable. Classifying smooth pursuits still depends on the vague definition regarding its transition into saccades or fixations, and in order to incorporate this into the filter, further work has to be done in defining the state more precisely.

Further work that could improve the filtering process is to develop a better process model for saccades. By developing a process model that is more suitable for saccades, methods that do not rely on thresholds, such as an IMM filter, would then be of interest for future work.

References

[1] Andrew T. Duchowski. *Eye Tracking Methodology*. Vol. 3rd. Springer International Publishing, 2017.

[2] M. Nilsson Rabbaa and R. Borre. *An advanced 3D eye state filtering and classification algorithm based on physiological constraints of the human eye*. 2016.

[3] Kenneth Holmqvist et al. *Eye Tracking: A comprehensive guide to methods and measures*. Oxford University Press, 2011.

[4] Oleg Komogortsev. *Eye Movement Prediction by Oculomotor Plant Modeling With Kalman Filter 2003-2007*. 2007. URL: https://etd.ohiolink.edu/!etd.send_file?accession=kent1190386786&disposition=inline.

[5] Jordan Pola. "Models of the Saccadic and Smooth Pursuit Systems". In: *Models of the Visual System* (2002).

[6] O. Nachmani et al. "Predicted Position Error Triggers Catch-Up Saccades during Sustained Smooth Pursuit". In: *Society of Neuroscience* (Jan. 2020), pp. 569–591.

[7] F. Fioravanti et al. "Saccadic Eye Movement Conjugation in Children". In: *Vision Research* 35 (Dec. 1995).

[8] Simo Särkkä. *Bayesian Filtering and Smoothing*. Vol. 3rd. Cambridge University Press, 2014.

[9] Kevin P. Murphy. *Switching Kalman Filters*. 1998. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.49.5703&rep=rep1&type=pdf.

[10] Manu Kumar et al. "Improving the Accuracy of Gaze Input for Interaction". In: vol. Vol. 1. Jan. 2008, pp. 65–68. DOI: 10.1145/1344471.1344488.

[11] Rainer Stiefelhagen and Jie Zhu. "Head Orientation and Gaze Direction in Meetings". In: (June 2004). DOI: 10.1145/506443.506634.

[12] Michael F. Land. "Predictable eye-head coordination during driving". In: *Nature* 359 (1992).

[13] W. Dai et al. "A Parametric Model for Saccadic Eye Movement". In: *IEEE Signal Processing in Medicine and Biology Symposium* (Dec. 2016).

[14] Andrew T. Duchowski. "A breadth-first survey of eye-tracking applications". In: Nov. 2002, pp. 455–470. DOI: https://doi.org/10.3758/BF03195475.

[15] O. Komogortsev and J. Khan. "Kalman Filtering in the Design of Eye-Gaze-Guided Computer Interfaces". In: *Human-Computer Interaction* 3 (July 2007).

[16] John Philip Patten. "Conjugate Eye Movements and Nystagmus". In: *Neurological Differential Diagnosis*. Berlin, Heidelberg: Springer Berlin Heidelberg,

1996, pp. 91–103. ISBN: 978-3-642-58981-2. DOI: 10.1007/978-3-642-58981-2_7. URL: https://doi.org/10.1007/978-3-642-58981-2_7.

[17] E. Dam, M. Koch, and M. Lillholm. *Quaternions, Interpolation and Animation*. 1998. URL: http://web.mit.edu/2.998/www/QuaternionReport1.pdf.

[18] D. A. Robinson. "The Mechanics of Human Smooth Pursuit Eye Movement". In: *The Journal of Physiology* (Mar. 1965), pp. 569–591.

[19] *Derivation of the IMM filter*. URL: http://www.control.isy.liu.se/student/graduate/TargetTracking/IMMderivation.pdf.

A

Conversion from 3D direction vector to heading and pitch

The conversion from x-, y-, and z-direction to heading and pitch was done using a preexisting function in Smart Eye's system. The used coordinate system is shown in figure A.1. The origin, of the coordinate system used in the conversion, is located in the center of the head and is locked onto the current pose of the head of the subject.

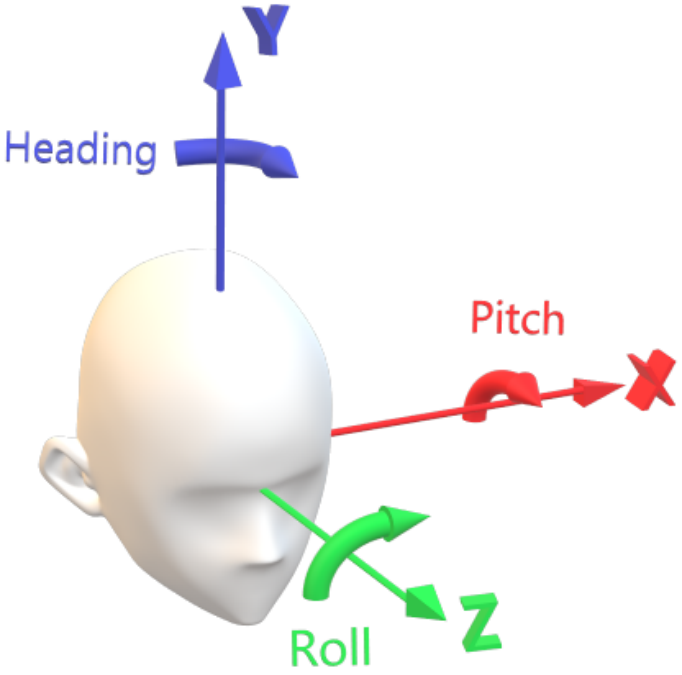

Figure A.1: The figure shows a 3D-view of heading, pitch, and roll. Roll, the rotation around the z-axis, is often not used in eye tracking because it is an unnatural movement in eye-tracking applications.

As shown in figure A.1, heading is the rotation around the y-axis, and pitch is the rotation around the x-axis in the used coordinate system. Roll, the rotation around the z-axis, has not been considered since this rotation is quite an unnatural motion in eye-tracking applications.