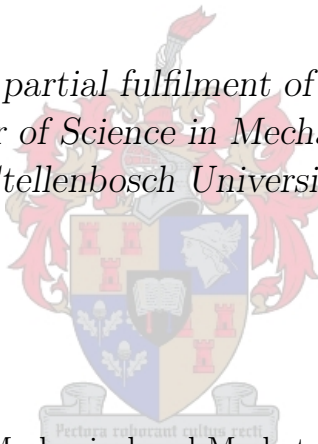


Nystagmus and Eye Reflex Sensor

by

Wayne Swart

*Thesis presented in partial fulfilment of the requirements for
the degree of Master of Science in Mechanical Engineering at
Stellenbosch University*



Department of Mechanical and Mechatronic Engineering,
University of Stellenbosch,
Private Bag X1, Matieland 7602, South Africa.

Supervisors:

Prof. C. Scheffer

Prof. K. Schreve

December 2011

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the owner of the copyright thereof (unless to the extent explicitly otherwise stated) and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date:

Copyright © 2011 Stellenbosch University
All rights reserved.

Abstract

Nystagmus and Eye Reflex Sensor

W. Swart

*Department of Mechanical and Mechatronic Engineering,
University of Stellenbosch,
Private Bag X1, Matieland 7602, South Africa.*

Thesis: MScEng (Mech)

December 2011

Nystagmus is an eye movement pattern that consists of a drifting gaze component, known as the slow phase, followed by a corrective quick phase component. The presence of nystagmus or the lack thereof under certain conditions can be used for various diagnostic purposes including the diagnosis of physiological, pathological and neurological conditions. The angular velocity of the quick phase can make the detection of nystagmus a challenging task for the untrained eye, since the quick phases are usually comparable with saccadic eye motions. The goal is thus to develop a fully automated diagnostic tool that can identify the presents of nystagmus in a patient's eye motions.

In this thesis, an appropriate eye tracking method was selected from a number of eye tracking methods that are commonly implemented in the literature. A video-oculography goggle concept was chosen based on criteria such as invasiveness, sampling rate, accuracy and telemedicine capability, amongst other nystagmus related necessities. A binocular video-oculography concept was chosen that satisfied the technical requirements and provided a cost-effective design. An automated analysis algorithm was developed for automatic nystagmus identification from eye motion data. The algorithm was validated by testing the performance of the algorithm on an optokinetic nystagmus signal. It proved to provide a reliable automatic identification of nystagmus beats, even in signals that contained nystagmus as well as random motion components. A statistical analysis showed that the algorithm provided a sensitivity of 91.8% and a specificity of 96.5% for pure nystagmus signals, and a sensitivity and specificity of 87.8% and 91.1% respectively for mixed signals.

Uittreksel

Nystagmus en Oog Refleks Sensor

(“*Nystagmus and Eye Reflex Sensor*”)

W. Swart

*Departement Meganiese en Megatroniese Ingenieurswese,
Universiteit van Stellenbosch,
Privaatsak X1, Matieland 7602, Suid Afrika.*

Tesis: MScIng (Meg)

Desember 2011

Nystagmus is ’n oogbewegingspatroon wat bestaan uit ’n dwalende tuurkomponent, wat die stadige fase genoem word, gevolg deur ’n vinnige korrigeringsbeweging wat bekend staan as die vinnige fase. Die teenwoordigheid van nystagmus, of afwesigheid daarvan in sekere gevalle, kan gebruik word in ’n verskeidenheid diagnostiese toepassings, onder andere die diagnose van fisiologiese-, patologiese- en neurologiese kwale. Die hoeksnelheid van die vinnige fase lei daartoe dat nystagmus dikwels moeilik is om te bespeur vir ongeöefende oë, aangesien dit vergelykbaar is met *saccade* bewegings. Die doel van hierdie navorsing is dus die ontwikkeling van ’n stelsel wat ’n volledige outomatiese identifisering van nystagmus kan behartig.

’n Gepaste oogvolgtegniek was gekies vanuit ’n aantal verskillende oogvolgmetodes wat dikwels in die praktyk gebruik word. Die finale keuse was ’n skermbril, video-oogvolgmetode wat gekies was op grond van kriteria soos onder andere, invallendheid, meetfrekwensie, akkuraatheid en geskiktheid vir telemedisyne toepassings. Die ontwikkelde brilkonsep bied ’n koste-effektiewe oplossing, met die moontlikheid om albei oë te volg en bevredig al die bogenoemde tegniese spesifikasies. ’n Geoutomatiseerde nystagmus identifiseringsalgoritme is ontwikkel. Die algoritme se effektiwiteit is getoets op optokinetiese nystagmusseine. Betroubare resultate is verkry vanaf die algoritme, selfs in die geval van gemengde seine wat bestaan uit arbitrêre- en nystagmus komponente. Statistiese analiese het gewys dat die algoritme ’n sensitiwiteit van 91.8% en ’n spesifisiteit van 96.5% kon behaal vir seine met slegs nystagmus inhoud. Vir gemengde inhoud seine het die algoritme ’n sensitiwiteit van 87.8% en spesifisiteit van 91.1% behaal.

Acknowledgements

I would like to express my sincerest gratitude to the following people:

- My study leaders, Prof. Cornie Scheffer and Prof. Kristiaan Schreve for their invaluable advice and guidance throughout the thesis.
- Dr. D. Koekemoer and Dr. A. Ziskind for additional clinical advice regarding the project.
- My colleagues at BERG that participated in the nystagmus recording survey.



Dedications

This thesis is dedicated to my friends and family whose continuous support and encouragement carried me through the last two years.

Contents

Declaration	i
Abstract	iii
Uittreksel	iv
Acknowledgements	v
Dedications	vi
Contents	vii
List of Figures	ix
List of Tables	xii
Nomenclature	xiii
List of Abbreviations	xiv
1 Introduction	1
1.1 Contents Overview	1
1.2 Research Objectives	2
2 Literature Review	3
2.1 Clinical Literature	3
2.2 Engineering Literature	12
2.3 Safety Study	21
3 Video-Oculography Device	24
3.1 Camera Selection	24
3.2 Lighting Source	27
3.3 Filter Selection	28
3.4 Lens Selection	30
3.5 VOG Goggle Design	31
3.6 Software Interface	32
4 Pupil Tracking	33
4.1 Image Processing Fundamentals	33
4.2 Tracking Algorithms	42

DEDICATIONS

4.3	Proposed Tracking Algorithm	47
5	Signal Processing	56
5.1	Parametric Data Analysis	56
5.2	Wavelet Approach	57
5.3	Proposed Signal Analysis Algorithm	58
5.4	Test Method and Results	71
6	Conclusion	79
	Appendices	81
A	Safety Calculations	82
A.1	LED Specifications	82
A.2	Safety Evaluation	85
B	Frequency Domain Theory	87
B.1	Fourier Transform Theory	87
B.2	Discrete Fourier Transform	88
B.3	Fast Fourier Transform	89
B.4	Convolution in the Frequency Domain	91
C	Manufactured Device and Selected Components	93
D	Feature Tracking and Related Theory	97
D.1	Bresenham Algorithm	97
D.2	Ellipse Fitting Through Least Squares Minimization	98
E	Test Data	101
E.1	Parameter Definitions	101
E.2	Collected Data	103
F	Software Development	114
F.1	Image Capture and User Interface	114
F.2	Tracking Software	116
F.3	Signal Analysis Software	117
	List of References	118

List of Figures

2.1	Anatomy of the Eye	4
2.2	Nystagmus Waveforms	6
2.3	Electrode Placement for EOG Recordings (Barea <i>et al.</i> , 2003)	13
2.4	EOG Signal Generated by Eye Movement	13
2.5	Scleral Search Coil Design (Robinson, 1963)	15
2.6	Scleral Search Coil Applied in Practice (Shelhamer, 1998)	15
2.7	Reflection as Captured by Remote VOG System: (a.) Dark Pupil, (b.) Bright Pupil and (c.) Corneal Reflex	17
2.8	A Commercial Remote Eye Tracking System by <i>Cogain</i> (Böhme <i>et al.</i> , 2006)	18
2.9	Monocular VOG System Developed by the Universität Zürich (Daunys <i>et al.</i> , 2007)	18
2.10	Top View of Camera-Mirror System with Light Paths	19
2.11	VO425 VOG System Developed by the <i>Interacoustics</i> ® (Stowe, 2007)	20
2.12	CIE Photobiological Spectral Bands (Sloney, 2007)	22
3.1	Saccade Data: (a.) Saccade Velocity vs. Saccade Magnitude (Original Data) and (b.) Saccade Period vs. Saccade Magnitude	25
3.2	PlayStation® 3 Eye Camera	27
3.3	IR Lighting Source PCB	28
3.4	LED Voltage Regulator for USB Power	28
3.5	Transmittance Curve for 940FIW25 Bandpass Filter	29
3.6	Reflectance Curve for 730FHQ5050 Dichroic Short-Pass Filter	29
3.7	Resolution Comparison: (a.) Standard PS3 Eye Lens, and (b.) M56-533 Lens	31
3.8	Exploded View of VOG Device	32
4.1	Basic Image Formation	34
4.2	1-D Correlation Method	36
4.3	Morphological Transformation Demonstration Body	38
4.4	Erosion of Demonstration Body	38
4.5	Dilation of Demonstration Body	39
4.6	Opening and Closing Operations on 1-D Intensity Array	40
4.7	Closed Demonstration Body	40
4.8	Sobel Masks: (a.) Vertical Gradient Component and (b.) Horizontal Gradient Component	41
4.9	Sobel Edge Detection: (a.) Original Image and (b.) Sobel Edge Image	42
4.10	Glint Shifted Over Pupil	43
4.11	Error Produced by Glints Present in Pupil Region	43
4.12	Glinted Pupil Test Image	44
4.13	Effects of Proposed Intensity Transformation Algorithm	45

DEDICATIONS

4.14	Circular Hough Transform	46
4.15	First Proposed Hybrid Pupil Tracking Method	49
4.16	Real World VOG Image	50
4.17	Histogram Processing of VOG Image	50
4.18	Threshold Image	51
4.19	Removal of Standalone Points: (a.) Thresholded Image Before Stand Alone Point Removal and, (b.) Image After Stand Alone Point Removal	51
4.20	Result of Closing Operation	52
4.21	Estimated Pupil Boundary	53
4.22	Estimated Pupil Boundary from Ellipse Fit	54
4.23	Lateral Gaze Tracking	55
4.24	Estimated Pupil Boundary from Ellipse Fit for no Glints on Boundary	55
5.1	Mallat Tree with Three Level Wavelet Expansion	60
5.2	Daubechies 2 Wavelet: (a.) Scaling Function, φ and (b.) Wavelet Function, ψ	61
5.3	Haar Wavelet: (a.) Scaling Function, φ and (b.) Wavelet Function, ψ	62
5.4	Demonstration Nystagmus Signal	62
5.5	Five Level Wavelet Decomposition Coefficients	63
5.6	Detail Coefficients at Local Extrema	64
5.7	Noise Reduction by Threshold: (a.) Noisy Signal, (b.) First Level Detail Coefficients, and (c.) Second Level Detail Coefficients with Threshold Values	66
5.8	Noise Reduction on Real VOG Signal: (a.) VOG Signal, (b.) First Level Detail Coefficients, and (c.) Second Level Detail Coefficients with Threshold Values	67
5.9	Extrema from Detail Coefficients	68
5.10	Reconstructed Signal	69
5.11	Parametric Definitions	70
5.12	Nystagmus Diagnosis: (a.) Original Signal, and (b.) Processed Signal	74
5.13	Nystagmus Diagnosis in Random Signal: (a.) Original Signal, and (b.) Processed Signal	77
A.1	Visual Representation of a Conical Section Solid Angle	83
A.2	Solid Angle Definition	83
A.3	Comparison of Solid Angles as Calculated by IEC and Solid Angle Definition	86
C.1	VOG Device Face Contact Side	93
C.2	VOG Device Front Side and Visual Shield	94
C.3	IR Bandpass Filter	94
C.4	Dichroic Filter	95
C.5	Custom Mounting Socket: (a.) Front View, and (b.) Rear View	95
C.6	LED PCB: (a.) Front View, and (b.) Rear View	96
C.7	USB LED Power Circuit	96
D.1	Circle Rounding Errors	97
D.2	Digital Circle	98
D.3	Ellipse	99
D.4	Coordinate Transformation (Coetzee, 2006)	100
E.1	Parameter Definitions for Pure Nystagmus	102
E.2	Parameter Definitions for Mixed Signals	103

DEDICATIONS

E.3	Signal 1	104
E.4	Signal 2	104
E.5	Signal 3	104
E.6	Signal 4	105
E.7	Signal 5	105
E.8	Signal 6	105
E.9	Signal 7	106
E.10	Signal 8	106
E.11	Signal 9	106
E.12	Signal 10	107
E.13	Signal 11	107
E.14	Signal 12	107
E.15	Signal 13	108
E.16	Signal 14	108
E.17	Mixed Signal 1	108
E.18	Mixed Signal 2	109
E.19	Mixed Signal 3	109
E.20	Mixed Signal 4	109
E.21	Mixed Signal 5	110
E.22	Mixed Signal 6	110
E.23	Mixed Signal 7	110
E.24	Mixed Signal 8	111
E.25	Mixed Signal 9	111
E.26	Mixed Signal 10	111
E.27	Mixed Signal 11	112
E.28	Mixed Signal 12	112
E.29	Mixed Signal 13	112
E.30	Mixed Signal 14	113
F.1	Graphical User Interface with No PS3Eye Cameras Connected	115
F.2	Graphical User Interface with Two PS3Eye Cameras Detected but Unconnected	115
F.3	Graphical User Interface with Two PS3Eye Cameras Detected and Connected	116

List of Tables

2.1	Tracking Method Comparisons	20
3.1	Lens Specifications	30
5.1	Statistical Results achieved by Jansen <i>et al.</i> (2010)	73
5.2	Collected Data for Pure Nystagmus Signals	76
5.3	Collected Data for Mixed Signals	78
A.1	IEC Maximum Radiance of Class 1 Source Over 100s	86

Nomenclature

Constants

$$\pi = 3.141\,592\,654$$

Variables

g_x	Horizontal Image Gradient	[intensity/pixels]
g_y	Vertical Image Gradient	[intensity/pixels]
m	Phase Magnitude	[pixels]
v	Phase Velocity	[pixels/sample]
x_0	Pupil Centre X-Coordinate	[pixels]
x_c	Median X-Coordinate	[pixels]
x_l	Minimum X-Coordinate	[pixels]
x_u	Maximum X-Coordinate	[pixels]
y_0	Pupil Centre Y-Coordinate	[pixels]
T	Phase Period	[samples]
α	Angle	[radians]
θ	Angle	[°]
Ω	Solid Angle	[steradians]

List of Abbreviations

CFT	Continues Fourier Transform
DFT	Discrete Fourier Transform
DWT	Discrete Wavelet Transform
ECG	Electrocardiogram
EEG	Electroencephalogram
EOG	Electro-oculography
FOV	Field of View
FT	Fourier Transform
IDFT	Inverse Discrete Fourier Transform
IFT	Inverse Fourier Transform
IR	Infrared
LED	Light Emitting Diode
MAD	Mean Absolute Deviation
ms	Miliseconds
NIR	Near-Infrared
nm	Nanometers
OKN	Optokinetic Nystagmus
PCB	Printed Circuit Board
RANSAC	Random Sample Consensus
SE	Structuring Element
USB	Universal Serial Bus
VOG	Video-oculography

Chapter 1

Introduction

Nystagmus is an involuntary eye motion that is defined by a slow gaze drift in one direction, followed by a quick corrective motion in the opposite direction. These two motions are referred to as the slow phase and the quick phase respectively. The combination of the two phases that represent one nystagmus wave period is generally referred to as a nystagmus beat. Since nystagmus can be used to diagnose various physiological-, pathological- and neurological conditions, it stands to reason that an diagnostic tool that can automatically diagnose nystagmus will be of great value.

1.1 Contents Overview

Literature reviews on relevant background for both the clinical and the engineering aspects of the thesis will be provided. Clinical literature will provide sufficient background on nystagmus, the causes thereof, and the different forms in which it presents itself. Engineering literature will focus on the different eye tracking methods and the implications thereof. The advantages and disadvantages of each method will be reviewed and the systems will be compared to one another to find the optimal solution for the problem at hand.

Designing of the eye tracking device and the selection of appropriate components will also be discussed. Design choices influenced by the clinical mobility requirements and other issues, such as ergonomic design, will be discussed. It will also be shown that the final device designed for this thesis provides a sufficient low cost design, ideal for telemedicine applications. Given a laptop computer, the device does not require any additional adapters or data acquisition systems.

Finally, all the required software algorithms necessary to provide an automatic diagnosis from the captured eye motion data is discussed to provide a full understanding of the operational functions of the system. Validation and performance evaluation of the proposed signal analysis algorithm on real world nystagmus signals is presented. Test procedures that were followed to capture the nystagmus signals will also be discussed.

1.2 Research Objectives

Objectives that were set for this thesis, in order to achieve the ultimate goal of automatic diagnosis of nystagmus, include the design and manufacture of a suitable prototype eye tracking device that can accurately track and record eye rotation over time. This device should be mobile and easy to use in order to facilitate clinical testing in both urban and rural locations. Another objective is the development of the appropriate accompanying signal processing software that may be required in order to complete the goal of an automatic nystagmus diagnosis from the captured eye motion data.

An important development requirement is telemedicine capability. This requirement introduces some significant constraints on both hardware development and software development. As was stated in the previous paragraph, telemedicine devices should be mobile and user friendly. Since telemedicine is most often practiced in rural areas, supporting equipment are scarcely available, which imposes more design constraints regarding the hardware development. Furthermore, the device should be minimally invasive.

The objectives for this thesis are to develop an eye tracking device that satisfy all the above mentioned constraints and to develop the software that can interpret and analyze the signals from the developed device. As will be discussed during the thesis, the most elegant solution is a video based eye tracker which introduced another requirement, namely the development of eye tracking software. Due to the telemedicine requirement, the computational loads of the tracking and processing algorithms should be kept to a minimum. Since this thesis is an introductory phase to the development of a nystagmography device with the primary objective of nystagmus capture and identification and the secondary objective of optimizing the design for telemedicine use, the focus fell on developing a prototype that can provide sufficient proof of concept and not the development of a market ready product.

Chapter 2

Literature Review

Some core terms and concepts regarding this project will be introduced in this chapter. The purpose of this chapter is to motivate the necessity of the project and present the associated literature that covers both the clinical and technical aspects of the thesis.

2.1 Clinical Literature

This section will review topics regarding the clinical aspects of the project. The main focus will be to introduce the concept of nystagmus and the different ways in which it presents itself, as well as the diagnostic value thereof. Furthermore, the possible uses of a device that can assist with the successful diagnosis of nystagmus will be introduced.

2.1.1 Eye Anatomy and Saccadic Motion

This section is included in order to introduce some terms regarding the anatomy of the eye, as well as eye movement. These terms will be raised frequently, and knowledge thereof will help follow the subsequent discussions more clearly. It will also clarify the discussions related to problem statement and project motivation.

2.1.1.1 Anatomy of the Human Eye

Figure 2.1 shows the anatomy of the human eye. The outermost part which covers almost the entire eyeball, is a tough connective tissue called the sclera, which is the white part of the eye. A transparent window in the sclera located at the front of the eye, where the pupil is located, is called the cornea. The pupil is not so much an anatomical part as it is the lack thereof. It is an opening in the front of the eye that permits light to pass to the sensory part of the ocular system in the back of the eyeball. The size of the pupil opening is controlled by the surrounding pigmented muscle called the iris. The iris consists of two smooth muscle layers, i.e. the circular and radial muscle layers.

Contraction of the circular muscle layer constricts the pupil, while contraction of the radial muscle layer will dilate the pupil. As mentioned before, the iris is a pigmented muscle and this is what gives the eyes colour.

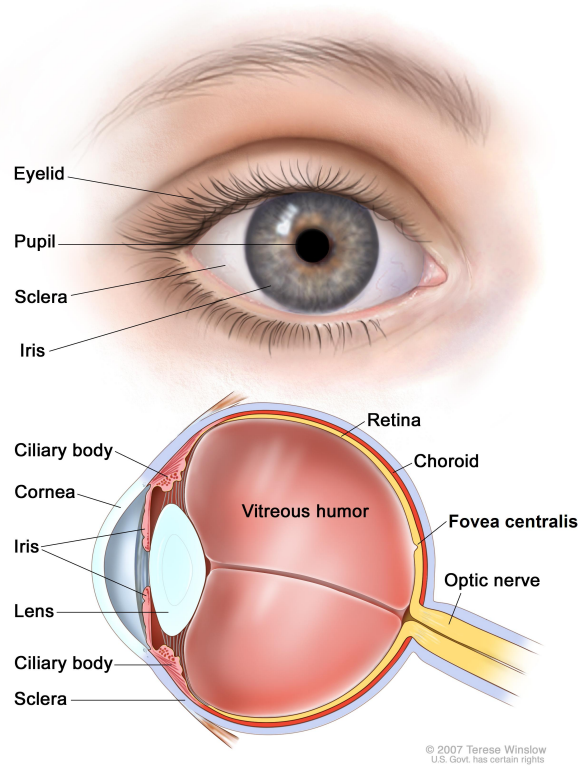


Figure 2.1: Anatomy of the Eye

Courtesy of Medical Illustrator Terese Winslow © 2007

The anatomical features discussed up until this point will mainly be referenced in the sections discussing image processing techniques. The terms of the anatomical parts that follows will predominantly relate to the literature review discussed in the sections that follow. As can be deduced from Figure 2.1, the ocular system is a vastly complex network of cellular-, muscular-, optic- and neurological subsystems that collaborate to fulfil the function of the ocular system as a visual sensor. Subsequently, a detailed discussion of the anatomy of the eye will not be given, but rather a condensed overview of the relevant parts. The retina is a neural layer containing millions of photoreceptor cells. The neuron cells, which acts as conductors from the photoreceptor cells, gather at a region on the retina called the optic disk. It is here where the neuron cells exit the retina into the optic nerve, which leads to the brain. There is another important region on the retina called the fovea or fovea centralis. The fovea is a tiny pit containing a very high density of photoreceptor cells and is the region on the retina where the target image is formed. Thus, one would constantly adjust one's gaze and focus in order to focus the image of interest on the fovea. The physiological literature for the last two paragraphs were referenced from (Fox, 2009).

2.1.1.2 Saccadic Eye Motion

A saccade can be described as a quick eye motion that allows a rapid shift in gaze from one image to another. Saccadic eye motion is one of the fastest voluntary muscle motions that can be achieved by the human body (Enderle *et al.*, 2005). Rotational velocities of up to a thousand degrees per second can be achieved during saccadic

motion. Although a saccade is generally described as a voluntary motion, it may occur during subconsciously driven eye movements.

2.1.2 Nystagmus

In 1921 a renowned neuro-ophthalmologist advised a fellow academic with the following words:

“Never write about nystagmus, it will lead you nowhere.” - Wilbrand

Following in the steps of numerous others, this author decided to ignore this advice in the hope of shedding just enough light on the subject in order to sufficiently clarify the importance of this project.

Nystagmus is a condition that presents itself in the form of an involuntary eye movement and consists of two distinct movement phases in a recursive pattern. The two phases are called the slow- and quick phases, and are thus classified by the angular velocity of the specified eye motion. The slow phase is usually related to a person’s gaze, while the quick phase is a corrective motion in the opposite direction. The velocity of the quick phase is comparable to that of a saccade, and is indeed identified as a saccade by some authors in the medical literature (Dell’Osso, 1989). In this case the saccadic motion is not a voluntary motion as stated in the previous section. The sequential combination of the above mentioned phases is called a nystagmus beat. The interested reader is encouraged to view one of the numerous video recordings of nystagmus on the internet to form a better picture of nystagmus induced eye motions, since this will be a prominent part of the thesis’ discussion¹.

Nystagmus in itself is not a disease, but rather a symptom of an underlying condition or disease. There are numerous different instances in which nystagmus can occur as a result of various different pathological-, physiological-, or neurological conditions. This produces an invaluable opportunity to make use of nystagmus as a diagnostic tool. A brief overview of some of the conditions that can cause nystagmus will be presented in the remainder of this section.

2.1.2.1 Nystagmus Waveforms

Cogan (1962) observed two basic types of waveforms for nystagmus, namely pendular nystagmus and jerk type nystagmus. He also noted that these waveforms are frequently observable for certain types of nystagmus. It would seem that he hinted towards associations of pendular- and jerk type nystagmus with sensory and motor nystagmus respectively. This topic will be discussed in detail later. For now, the discussion will focus on the types of waveforms, as shown in Figure 2.2, which shows the two main types of nystagmus waveforms and deviations thereof.

¹<http://www.mrcophth.com/movies/periodicalalternatingnystagmus.WMV>

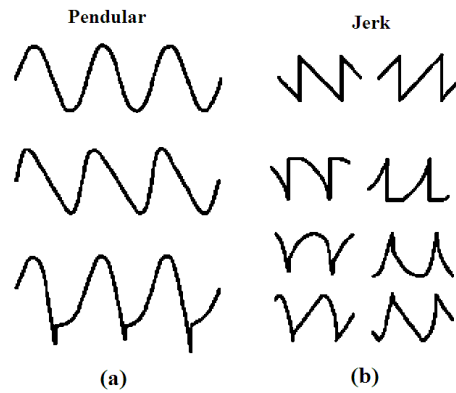


Figure 2.2: Nystagmus Waveforms

Edited from Dell'Osso and Daroff (1975)

Pendular nystagmus has a sinusoidal waveform as shown in Figure 2.2.A. The phases in the two opposing directions are thus symmetrical, contrary to the classic description of a nystagmus beat as was discussed earlier. This, by no means, makes this type of waveform less important. Cogan (1962) noted that this type of waveform is associated with congenital nystagmus caused by the lack of central vision as a result of congenital birth defects such as albinism, cataracts or corneal opacities. Furthermore, Dell'Osso (1973) found that the point of maximum visual acuity was at one of the two peaks, which would seem to reinforce the above theory of visual defect that could cause an alternating gaze around the region of low visual acuity.

Jerk type nystagmus is a waveform that adheres to the classic description of nystagmus and is shown in Figure 2.2.D. Dell'Osso *et al.* (1974) described the jerk waveform to have a slow phase consistent to foveal drift away from the target followed by a corrective fast phase in the opposite direction. Congenital nystagmus presenting itself in such a waveform is usually accompanied by a neutral point. When the patient's gaze is in the neutral point position, the amplitude of the waveform is either at its minimum or zero. When the gaze is shifted away from the neutral point, the fast phase of the jerk waveform is always in the direction of the gaze (Cogan, 1962; Dell'Osso *et al.*, 1974).

2.1.2.2 Congenital Nystagmus

Congenital nystagmus occurs within three months from birth, and can be subdivided into two principal types, namely sensory- and motor nystagmus (Cogan, 1967). As suggested by the chosen names, the type of nystagmus is identified by the inherent afferent or efferent defect respectively.

Sensory nystagmus is the result of impaired vision due to abnormalities in the visual pathways, for example retinal or optic nerve diseases (Von Noorden and Campos, 2002). There are many different abnormalities that can hinder proper image formation on the fovea and lead to sensory nystagmus. Congenital cataracts serves as a good example. According to Orssaud (2003) a great number of these conditions are genetic. This description is very similar to the nystagmus described by Cogan (1962) as ocular nystagmus. It is generally accepted that this type of nystagmus is predominantly of a horizontal nature, and it is often associated with a pendular waveform (Cogan, 1967).

Motor nystagmus is caused by a defect in the efferent pathways and might also involve some problems in the centre for ocular motor control in the brain (Von Noorden and Campos, 2002). Cogan (1967) mentioned that this type of nystagmus is frequently associated with the jerk type waveform. Motor nystagmus is thus predominantly a neural issue with no ocular defects as is the case with sensory nystagmus. Cogan (1962) described a type of nystagmus he called neural nystagmus, which is closely related to this definition of motor nystagmus. Cogan (1962) also included something called labyrinthine nystagmus, which is associated to the ocular-vestibular system and will be discussed in more detail in a following section. Neural nystagmus is associated with all directions, namely horizontal, vertical and torsional. Cogan (1962) noted that vertical nystagmus is an important symptom, as it is usually indicative of central nervous system disease, and it is the only type of nystagmus that will be present in Parkinson's disease patients. Furthermore, he also stated that torsional nystagmus can suggest neural or labyrinthine disease.

2.1.2.3 Acquired Nystagmus

Sometimes nystagmus manifests later in life, and not within the first few months after birth as is the definition of congenital nystagmus. Gresty *et al.* (1991) reported cases of congenital nystagmus manifesting in teenagers and even in adults. In this case the validity of the term comes into question (Von Noorden and Campos, 2002). Therefore, nystagmus that occurs later in life is generally referred to as acquired nystagmus. Acquired nystagmus is mostly a case of motor defect type nystagmus since sensory problems which develop later in life do not usually cause nystagmus (Cogan, 1962). Gresty *et al.* (1984) did point out, however, that congenital nystagmus can sometimes only be discovered later in life because of factors masking the nystagmus during early years. Sometimes the neutral point is more of a neutral gaze region in which nystagmus does not occur. They also report that nystagmus is frequently misinterpreted as normal eye movements and seen as a special feature rather than an abnormality. In these cases the term "congenital nystagmus" is still applicable.

2.1.2.4 Physiological Nystagmus

Physiological nystagmus is a type of nystagmus that can be evoked by means of a physiological stimulation. In this case the nystagmus is a result of a physiological process and is completely normal. In fact, the absence of nystagmus would point to an abnormality. There are three basic types of physiological nystagmus, namely end-point, optokinetic, and ocular-vestibular nystagmus.

End-point nystagmus, also generally known as gaze-evoked nystagmus, will occur in any normal person when the person's gaze is shifted to the far lateral position (Koekemoer, 2011). There are, of course, exceptions to the rule, but end-point nystagmus can be evoked in most normal individuals. This type of nystagmus usually has a jerk type waveform (Daroff *et al.*, 1979). End-point nystagmus can be divided into three basic subtypes, namely fatigue nystagmus, unsustained- and sustained nystagmus. Fatigue nystagmus occurs after a lateral eccentric gaze has been maintained for about 30-90 seconds and usually intensifies as the gaze is maintained (Abel *et al.*, 1978).

Unsustained- and sustained end-point nystagmus occurs immediately after achieving an

eccentric gaze position. The difference is that unsustained end-point nystagmus appears as a transient condition and subsides after a few seconds, while sustained end-point nystagmus persists throughout the duration of the eccentric gaze.

Optokinetic nystagmus is a jerk type nystagmus brought forth by a moving object or pattern. This is a normal response in which the slow phase is the following eye motion and the quick phase is the motion that returns the gaze to the starting point. In practice, this type of nystagmus is commonly evoked by an optokinetic drum which is a cylinder with black and white stripes parallel to the axis of rotation. When optokinetic nystagmus cannot be elicited, it may suggest some form of neural damage (Cogan, 1962).

The final type of physiological nystagmus is ocular-vestibular nystagmus or just vestibular nystagmus, which is also known as caloric nystagmus and is a result of the physiological connection between the ocular- and vestibular systems. Cogan (1962) referred to this type of nystagmus as labyrinthine nystagmus and reported that it has a jerk type waveform. The semicircular canals in the inner ear act as motion sensors in the human body and relay signals to the ocular motor system in the brainstem via the vestibular nerve. When the head is at rest, the tonic neural stimulation from the two ears to the ocular motor system are the same. When the head is in motion, it causes a difference in the neural stimulation received from the two vestibular nerves and subsequently prompt the ocular motor system to move the eyes with the same velocity in the opposite direction (Fox, 2009; Daroff *et al.*, 1979; Cogan, 1962). This is called the ocular-vestibular reflex or caloric response and can be stimulated by irrigating the ear with hot or cold water or air. According to Cogan (1962), hot water stimulates the labyrinthine, while cold water depresses it. Thus, hot water will cause eye motion to the side that is being irrigated, and cold water will cause motion to the side that is not being irrigated. Irrigation of the ear in an attempt to evoke a caloric response is often used as a test for brainstem death in comatose patients (Koekemoer, 2011). Another way to stimulate this response is by putting the patient in motion during examination. This is often done by using a rotating chair. Pathological conditions in the vestibular system can cause spontaneous vestibular nystagmus which can then be used in the clinical diagnosis of the vestibular problem. Balance issues are frequently related to vestibular abnormalities.

2.1.2.5 Miscellaneous Nystagmus

There are many forms of nystagmus of which the classification depends on, among other things, the waveform, direction of movement, time of onset and pathogenic-, neurological- or physiological origin. Dell'Osso (1989) lists 45 types of nystagmus and therefore the types of nystagmus are far too numerous for each type to be discussed in detail in this thesis. So far, this section discussed the types of nystagmus that are most common and of higher importance. Now, some types of known nystagmus that are infrequently encountered or of less clinical importance, will be reviewed.

Latent nystagmus is a form of nystagmus that occurs if one eye is blindfolded, even though the eyes are stationary when both are open. This can sometimes cause a erroneous diagnosis of visual impairment in a healthy patient when monocular testing is used. See-saw nystagmus is an interesting form of nystagmus which cause a pendular waveform with one eye moving upward, while the other one moves downward. Convergence nystagmus is the slow abduction movement of the eyes followed by a

corrective quick adduction movement of the eyes. Conversely, divergence nystagmus is characterized by a slow adduction movement of the eyes followed by a corrective quick abduction motion. Nystagmus is also frequently associated with spasmus nutans, which is a condition that causes the spasmodic nodding of the head of a patient. Lastly, alcohol/drug induced nystagmus appears in any normal person when, as the term suggests, the person is intoxicated. This phenomenon is frequently used by law enforcement personnel as a field sobriety test in some countries (Good and Augsburger, 1986). Dell’Osso (1989) objected against this test since gaze-evoked end-point nystagmus or congenital nystagmus can cause a false positive. Also, small amounts of tranquillizers that will not affect driving ability can cause nystagmus.

Since every type of nystagmus in existence cannot be discussed in detail, the interested reader is invited to consult the clinical literature dedicated to the vast variety of nystagmus. The author would suggest literature by Cogan (1962) and Dell’Osso (1989) as an adequate starting point.

2.1.2.6 An Alternative Opinion on Nystagmus

Dell’Osso and colleagues have been advocating that nystagmus should not be subdivided into the sub-categories of motor- and sensory nystagmus, as is the popular practice in the medical literature (Dell’Osso, 1989; Dell’Osso *et al.*, 1974, 2007). According to these authors, all types of nystagmus derives from the ocular motor system. In a relatively recent paper, Dell’Osso *et al.* (2007) stated that the modern paradigm of motor- and sensory nystagmus is a result of the misinterpretations of the work of the late David Cogan, one of the keenest clinical observers of nystagmus and perhaps one of the most renowned researchers on the phenomenon. Cogan (1967) used the terms “sensory” and “motor” while discussing pendular- and jerk type nystagmus respectively. Dell’Osso *et al.* (2007) referenced a hand written note as well as an article (Cogan, 1967) by David Cogan to make two arguments. First, that the waveform of the nystagmus cannot be used to make a clinical diagnosis, since both is subject to the ocular motor system. And furthermore that Cogan did not imply that the pendular- and jerk waveforms are associated with sensory- and motor nystagmus respectively. Secondly, they argued that David Cogan never claimed that nystagmus could be separated into sensory- and motor type nystagmus.

The first argument made by Dell’Osso *et al.* (2007) is understandable. Although Cogan often made connections between the waveforms and the presumed cause of the nystagmus in his work, it seems that he was rather reluctant to claim a definite relation between the subjects (Cogan, 1962, 1967). Furthermore, the main point of discussion in the hand written note to Dell’Osso *et al.* (2007) was that he (Cogan) thought he made it clear that the distinctions between the waveforms are not sufficient to make a classification of nystagmus. The second argument mentioned above, however, is rather confusing. The main argument here, was that Cogan used the words “presumed pathogenetic”, and therefore, that his work was based on his clinical impression. Subsequently, they believe that the notion of sensory- and motor nystagmus should be dismissed. Cogan’s impression of the separability of nystagmus into sensory- and motor nystagmus seemed to be rather strong though, since he made direct use of the terms “sensory” and “motor” (Cogan, 1967). Furthermore, Cogan explicitly mentioned sensory

CHAPTER 2. LITERATURE REVIEW

causes of nystagmus in sections about ocular- and congenital nystagmus (Cogan, 1962). This all points to rather thorough clinical research.

Following the above paragraph, it is not clear how Dell’Osso *et al.* (2007) can claim only a general misinterpretation of Cogan’s work without discrediting some portions of it. Perhaps it can be said that Dell’Osso *et al.* (2007) do not deny the indirect cause of nystagmus by some sensory defects, but only want to emphasize the fact that all eye motion, including nystagmus, is controlled by the ocular motor system, and subsequently feel the need to remove all confusion regarding the ocular system by avoiding the term “sensory nystagmus”. Whatever the case may be, it is evident that medical researchers do not yet fully understand or agree on the phenomenon of nystagmus, which further reinforces the motivation to develop a clinical diagnostic device that can assist in the study of nystagmus. The author would like to conclude this section on nystagmus with the following quote:

“A lot of intensely intelligent and highly dedicated workers have given their lives to this subject of nystagmus and very little has come out of it.” - Frank Elliot

2.1.3 Telemedicine

There are dozens of definitions for telemedicine, each depending on the type of the healthcare considered and the definition’s time of origin, as telemedicine is a continuously developing field. Bashshur *et al.* (1997) sums up telemedicine with a universal definition independent of time:

“Broadly, telemedicine involves the use of modern information technology, especially two way interactive audio and video telecommunications, computers and telemetry, to deliver health services to remote patients and to facilitate information exchange between primary care physicians and specialists at some distances from each other.”

Telemedicine is thus the practice of bringing healthcare services to the patient without the patient having to travel to the physician, typically via the use of various electronic communication methods like broadband landlines or satellite links. Most commonly this will refer to the act of bringing healthcare to rural areas. In this case the primary healthcare professional is not necessarily a trained physician and even if this is the case, it is likely that the person is not a specialist in the required field. Acquiring the necessary information and sending it to the appropriate specialist can facilitate a team based effort to reach the correct diagnosis. The above-mentioned information will most likely include some form of signal- or media data. Clinical signals may include ECG’s, EEG’s and blood oxygen saturation levels, etcetera. Media data can include anything from photos, videos and sound recordings to radiology scans. Although telemedicine includes rural healthcare, it is not bounded by it. Telemedicine can also refer to the remote monitoring of vitals for the elderly, for example. Another possible application of telemedicine is the practice of healthcare in correctional facilities as reported by Nacci *et al.* (2002).

One of the objectives for this thesis is to develop a device that can be used in telemedicine applications. Diagnostic tools for telemedicine applications are often used in situations where the primary healthcare professional is not a medical doctor, but also helps qualified physicians to collect data in situations where the patient/doctor ratio is

CHAPTER 2. LITERATURE REVIEW

overwhelming, as is often the case in rural practices. The implications that stem from the need for a telemedicine friendly device, are that the device should be mobile and easy to use. It was assumed for the purpose of this thesis that the only additional equipment that will be available would be a midrange laptop.

2.2 Engineering Literature

The previous section focussed on the literature review regarding the related clinical aspects of the project. The remainder of this chapter will introduce the technical aspects of the project from an engineering perspective. This will include an overview of the existing eye motion capture technologies, as well as some safety considerations associated with these methods.

2.2.1 Eye Tracking Technology

In this section a number of different eye tracking techniques will be discussed. The practice of eye tracking by the use of some technological device is widely referred to as oculography. Eye tracking technology have been applied in the medical field for more or less 80 years, mainly used in studies of eye mechanics and eye motion related illnesses. In the recent couple of decades, eye tracking technology was also successfully implemented to generate command inputs for assistive devices designed for the handicapped. Therefore, eye tracking is a well-established research field, although there are a limited number of choices of fully commercialized systems. The objective of this section is to present the most popular eye tracking methods and give a brief discussion about the advantages and disadvantages associated with each method. In this section, eye motion will be referred to several times and it is important to understand the meaning behind the terms in order to clearly follow the topic discussed. Firstly, all eye movements are of a rotational nature. Thus, the eye rolls around three axes. However, since eye motion is intuitively observed in a 2-dimensional plane, roll around the vertical and horizontal axes are typically observed as pupil translation, while roll around the anterior-posterior axis (visual axis) is observed as torsional motion. Therefore, whenever the terms "roll" or "torsion" are used without specifying the axis, the motion described will refer to the rotation of the eye around the anterior-posterior axis.

2.2.1.1 Electro-oculography

Electro-oculography (EOG) is a popular eye tracking method in devices designed for disabled persons. These designs are used as input devices for computers or electronically controlled wheelchairs as implemented by Barea *et al.* (2003), for example. This allows paralysed people to achieve some level of independence. EOG is a method that makes use of the resting retinal potential to determine the angular displacement of the eye. The resting retinal potential is a low, but detectable, voltage generated by a small current flowing from the cornea towards the retina (Shinomiya *et al.*, 2008). The resting retinal potentials have a range of approximately 0.4 to 1.0 millivolts (Liu *et al.*, 2007), although potentials of up to 3.5 millivolts have been reported by Barea *et al.* (2003). EOG voltage signals from the electrodes have linear relations with the different components of the eye rotation and range between 15 and 200 microvolts (Liu *et al.*, 2007). Recording eye rotations around two axes requires a minimum of five appropriately placed electrodes as shown in Figure 2.3.



Figure 2.3: Electrode Placement for EOG Recordings (Barea *et al.*, 2003)

In Figure 2.3, the green electrode, usually placed in the centre of the forehead, is the reference electrode. The black and white electrodes are the horizontal channel anode and cathode respectively, while the red and brown electrodes are the vertical channel anode and cathode respectively. Figure 2.4 shows an example of a recorded EOG signal from the horizontal channel electrodes to demonstrate the principle of operation.

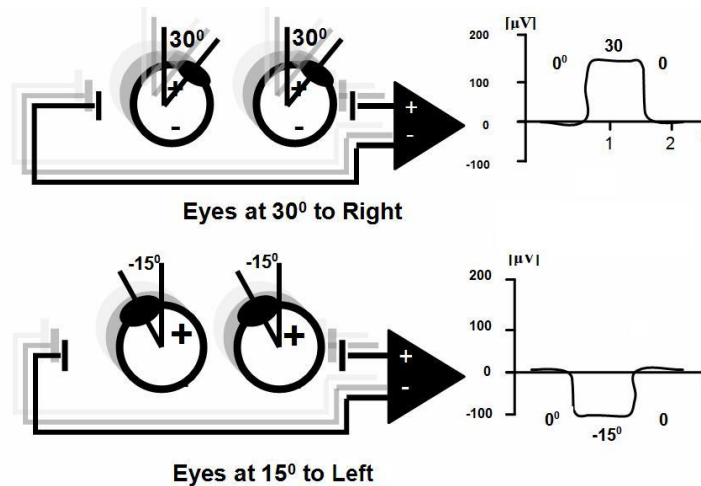


Figure 2.4: EOG Signal Generated by Eye Movement

Image edited from Trikha *et al.* (2007)

Electro-oculography has a major advantage in the sense that it is a non-invasive method. It would also seem that high accuracy measurements of the eye's rotation can be achieved by electro-oculography methods since Barea *et al.* (2003) and Liu *et al.* (2007) reported errors of 2° and less. However, Shinomiya *et al.* (2008) have reported that electro-oculography is sensitive to measurement conditions and the surrounding environment, affecting repeatability. Furthermore, they also documented that the occurrence of cross-talk potentials have a significant influence on the measurements. Cross-talk is the error produced in a measurement as a result of the resting retinal potential in the adjacent eye that is not being tracked and therefore produces error components in the signal.

2.2.1.2 Photo-Electric Oculography

Photo-electric oculography can refer to various methods depending on the choice of transducer. These methods can be named in different ways, but they all make use of the principal concept of light reflection. The idea is to illuminate the eye and capture the reflected light with photosensitive transducers such as photosensitive resistors. These transducers capture the sharp reflection gradients at the limbus, the boundary of the iris and sclera, or make use of the proportional light reflection from the sclera. In some methods, particularly used in earlier years, a contact lens containing a mirror surface around the pupil was applied to the subject prior to recording. The use of photosensitive transducers basically makes this collection of methods a simplified version of video-oculography, which will be discussed in a following section. An example of a photo-electric oculography system is the IRIS which was developed by Reulen *et al.* (1988).

2.2.1.3 Sclera Magnetic Search Coil

The scleral search coil (SSC) method, or magnetic search coil method as referred to by some researchers, is a technique that makes use of the principle of magnetic induction in a wound coil. The coil is mounted in a silicon contact lens and is applied directly to the subject's eye, usually after being anaesthetized to minimize discomfort. An alternating magnetic field is then created across the subject's head during tracking. The varying magnetic field then generates a potential difference over the coil in accordance to Faraday's law. This voltage can be amplified and measured to determine the position of the coil in the field. Three dimensional measurements can be implemented by two orthogonally mounted coils fixed in layers at different depths. One coil is of circular shape and the second coil is formed into a bent ∞ -shaped coil, as can be seen in Figure 2.5. For these multi-dimensional measurements, two magnetic fields in quadrature phase are applied across the subject's head, orientated in the vertical and horizontal directions respectively. Quadrature phase magnetic field generation entails that the fields are generated with phase separations of 90° . The measurements can then be obtained through phase coding, which is simply the principle of measuring the voltage over each coil within a certain phases (Robinson, 1963).

Figure 2.5 shows a SSC design by Robinson (1963), the original developer of the scleral search coil. Since then Collewijn and colleagues have extensively investigated the SSC and its uses (Collewijn *et al.*, 1975; Collewijn, 1977; Collewijn *et al.*, 1985). A new silicon contact lens, which was designed to lessen discomfort and increase suction to the eye, was developed by Collewijn *et al.* (1975).

CHAPTER 2. LITERATURE REVIEW

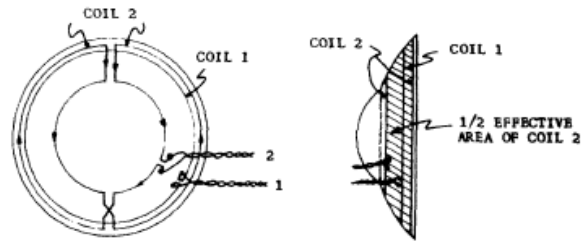


Figure 2.5: Scleral Search Coil Design (Robinson, 1963)

Scleral search coil tracking is a popular method due to the high accuracy provided by it. Robinson (1963) reported accuracy and linearity values around 2%, and in later years accuracy and linearity values of around 1%, including torsional measurements were reported by some researchers (Collewijn, 1977; Collewijn *et al.*, 1985). Subsequently, the SSC method is considered by many researchers to be the gold standard for eye tracking techniques with regards to accuracy. Another major advantage of the scleral search coil method is the ability to provide torsional measurements. Although it is possible to implement torsional tracking with 3-dimensional video-oculography, special techniques and algorithms are required. This will be discussed in more detail in following sections. The SSC method is the only eye tracking technique that inherently provides torsional measurement capability. Collewijn *et al.* (1985) successfully implemented torsional tracking in humans using the method proposed by Robinson (1963) and phase detection techniques described by Collewijn (1977).

The scleral search coil method does come with the disadvantage of being an invasive method. Apart from the fact that it is more difficult to set up, it also causes some obvious discomfort for the subject. The irritation caused to the eye when applied to a real person, is plainly visible in Figure 2.6.



Figure 2.6: Scleral Search Coil Applied in Practice (Shelhamer, 1998)

Various authors have reported that recording time is limited due to the invasive nature of the contact lens (Houben *et al.*, 2006; Moore *et al.*, 1996). Another disadvantage of the SSC method, reported by Moore *et al.* (1996), is lens slippage, particularly during torsional motion. These problems are reported despite various attempts to correct them, for example the silicon contact lens, developed by Collewijn *et al.* (1975), designed to

lessen discomfort and improve suction to the eye. Furthermore, Kenyon (1985) developed a soft contact lens in attempt to improve the design by Collewijn *et al.* (1975).

2.2.1.4 Video-Oculography

Video-oculography (VOG) is the practice of eye tracking by making a video recording of the eye and then processing each frame of the video recording individually in order to determine the direction of gaze at each discrete time-step. This is implemented by using a field of study, broadly referred to as image processing. Video-oculography is the most recent development among the eye tracking techniques, but is by no means a young technology. Young and Sheena (1975) already implemented eye tracking with video cameras in the 1970's. Researchers were somewhat reluctant to adopt this method in their studies and felt dependent on other methods like EOG and SSC during the last half of the twentieth century (Gans, 2001). Lately, VOG has become increasingly popular, a trend that can most likely be attributed to the technological advances in both photography and digital processing. There are a number of variations of VOG implemented in eye tracking systems that are currently in use. The rest of this section will discuss the different techniques, and combinations thereof, implemented in operational VOG systems.

The first important choice that needs to be considered when developing a VOG system, is the lighting source that will be used. This has a definite impact on the equipment and optics selection for the system. It also affects the development of the tracking algorithm to some extent. There are two basic options regarding lighting. The first is to make use of the ambient visible light, and the second is to make use of an infra-red (IR) lighting source. The latter has various advantages, including a controlled light environment unaffected by the flicker of electric lamps, and the fact that the IR is not within the visible spectrum and therefore does not agitate the subject. There is another advantage of IR light associated to its location in the light spectrum when considering camera placement, which will be discussed in context in a following paragraph on head mounted VOG systems.

Remote tracking is a common method of VOG, frequently implemented in situations where the tracker acts as an input device for a work station. Giving a computer command inputs by use of the eyes, as opposed to the hands, is commonly referred to as eye typing. This type of implementation is of great value as an aiding tool for the handicapped, as described by Hutchinson *et al.* (1989). Although some developers do make use of visible light in their applications (Hansen *et al.*, 2002), IR lighting is a popular choice when working with remote tracking devices. The reason for this is the opportunity for the efficient exploitation of the principle of light reflection. The corneas reflect the IR light from the source back to the camera more efficiently than any other surface on the human face. Subsequently, the pupil appears as a bright region in the image with an even brighter spot associated with the light source when the illumination is coaxial to the camera, or the pupil appear as a dark region when the illumination is not coaxial with the camera. The effect of coaxial and non-coaxial illumination is shown in Figure 2.7. The bright spot reflected from the cornea is generally known as the corneal reflex. The positions of the pupil and the corneal reflections can be determined by a number of different algorithms. The algorithm can be as simple as an adaptive threshold algorithm, or it can be more sophisticated by implementing an elegant

CHAPTER 2. LITERATURE REVIEW

template or shape fitting algorithm. The resulting positions can then be used to calculate the direction of gaze as explained by Guestrin and Eizenman (2006). Using two light sources, placed apart at a predetermined distance, simplifies the calculation of gaze direction and eliminates most problems associated with position and movement of the head (Merchant *et al.*, 1974; Sugioka *et al.*, 1996). If only one light source is used, the head needs to be kept perfectly still, or the head orientation needs to be tracked. This is usually done by tracking the face in three dimensional space from the images as is implemented by Newman *et al.* (2000); Matsumoto and Zelinsky (2000), but can also be accomplished by an alternative tracking method. The alternative method can either refer to magnetic transducers as implemented by Allison *et al.* (1996) or another camera mounted on the head to record the scene of gaze (Yu and Eizenman, 2004). Naturally, these types of methods can become tedious and complicated since more than one acquisition system is needed. Figure 2.8 shows a commercial remote eye tracking system developed by *Cogain*. As is clearly observable from the figure, the system makes use of two light sources and a single camera. Some systems use more than one camera to improve the accuracy of the eye tracking (Newman *et al.*, 2000; Matsumoto and Zelinsky, 2000).

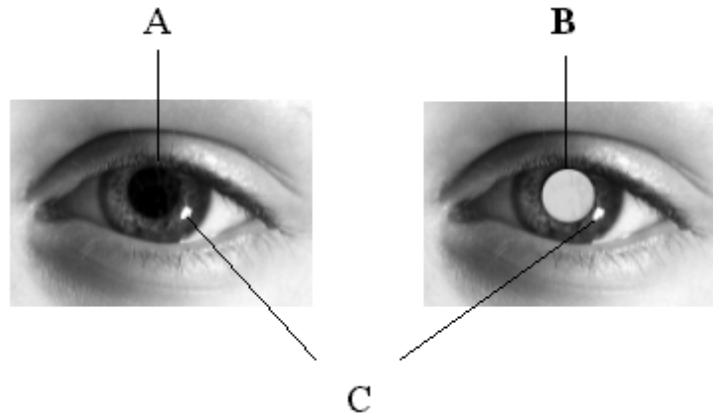


Figure 2.7: Reflection as Captured by Remote VOG System: (a.) Dark Pupil, (b.) Bright Pupil and (c.) Corneal Reflex

Edited from Milekic (2004)



Figure 2.8: A Commercial Remote Eye Tracking System by *Cogain* (Böhme *et al.*, 2006)

Head mounted video-oculography systems is the second and final class of VOG system. The design strategy for this type of VOG differs quite significantly from the remote tracking systems discussed previously. The most crucial difference that separates these methods from the remote tracking methods, is the fact that the camera is placed coaxially with the eye which inherently provides a less distorted image. Another beneficial difference to remote systems, is that the camera is much closer to the eye which results in a better spacial resolution. This also facilitates the option of a cheaper design since high resolution cameras are expensive. Monocular VOG is a design which places the camera in front of the eye of interest. It is called monocular VOG since the placement of the camera partially or completely obstructs the field of view and consequently only one eye can be recorded at a time without obstructing the entire field of view of the subject. Figure 2.9 shows an example of such a system.

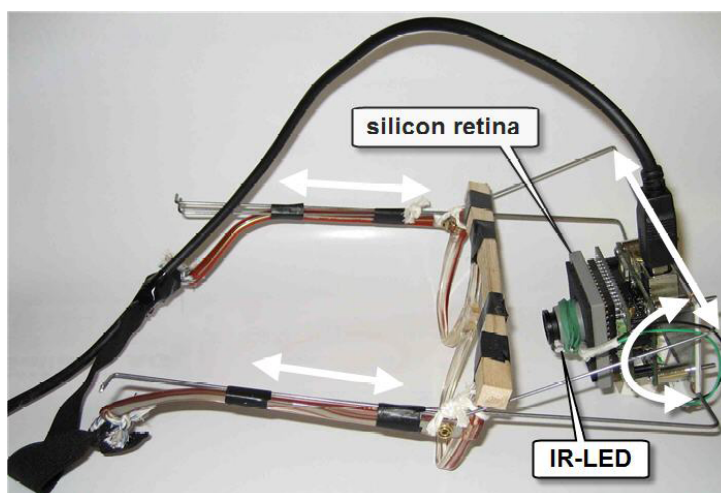


Figure 2.9: Monocular VOG System Developed by the Universität Zürich (Daunys *et al.*, 2007)

CHAPTER 2. LITERATURE REVIEW

An alternative to this method is the use of dichroic filters mounted in front of the subject's eyes. A dichroic filter is an optical filter that passes light with certain wavelengths and reflects light with wavelengths outside of the pass-band. Noting that IR light is a feasible light source for VOG purposes, the dichroic filter used in these methods are chosen to pass visible light and reflect IR light. These types of dichroic filters are often referred to as hot mirrors and provide the subject with an unobstructed field of view while allowing the cameras to record the eyes. This type of design offers the unique capability of placing the camera at a spacial angle with the eye while keeping the camera coaxial with the eye on a visual basis. Figure 2.10 graphically illustrates this concept. This ensures an optimal image formation while providing a clear field of view for the subject and create the opportunity to implement binocular tracking. An unobstructed field of view is important in order to enable the subject to respond to visual stimuli. Figure 2.11 shows the VO425, a commercial binocular VOG system used for vestibular testing, developed by *Interacoustics*². A similar system, called the Chronos Eye Tracker, developed by *Chronos Vision*, is also commercially available (Clarke, 2004). The VO425 and it's software were specifically designed with vestibular testing in mind, while the Chronos eye tracker is suitable for more general applications and includes torsional tracking.

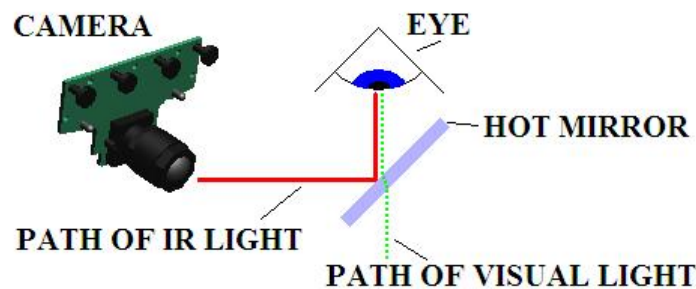


Figure 2.10: Top View of Camera-Mirror System with Light Paths

Three dimensional tracking is the final point of interest regarding video-oculography. Three dimensional tracking refers to the capability of performing torsional tracking, as well as tracking of the standard vertical and horizontal motions. As was previously mentioned in Section 2.2.1.3, torsional tracking with VOG requires the implementation of some specially formulated algorithms. According to Imai *et al.* (2005) there are three main methods that can be implemented to achieve torsional tracking with VOG. The first method is to perform a rotation vector analysis based on the location of two distinctive points in the image. These points have traditionally been placed on the sclera with a special marker prior to recordings, but Imai *et al.* (1999) have developed a method that uses the centre of the pupil and an iris freckle as the two reference points. The remaining two methods both make use of the form of the iris. The first of the methods entails tracking of distinct iral landmarks and then representing the position of the eye in Fick's coordinates. The second method performs a polar cross-correlation on iral intensities sampled on a circular path. Even though torsional tracking is cause for the development of more complicated algorithms, the possibility thereof and the fact that video-oculography is a non-invasive method, gives enough reason to consider VOG as a feasible alternative to the conventional eye tracking methods.

²Brochure available at: http://www.northeasterntech.com/pdf/VN415_VO425%20Brochure%20US.PDF



Figure 2.11: VO425 VOG System Developed by the *Interacoustics*® (Stowe, 2007)

2.2.1.5 Discussion and Method Comparisons

Up to this point, the focus was on reviewing the different eye tracking methods that are available. Now the different methods described earlier in this section will be compared to each other to weigh up the inherent pros and cons for each method, such that an appropriate method for the purpose of this thesis can be chosen. Table 2.1 shows a summary of the main criteria for selection of the required eye tracker and the performance of the different methods that are commonly implemented.

Table 2.1: Tracking Method Comparisons

Method	Sampling Rate (Hz)	Error (degrees)	Invasive?	Reference
Electro-oculography	250	2	No	(Liu <i>et al.</i> , 2007)
Magnetic Search Coil	>500	<1	Yes	(Imai <i>et al.</i> , 2005)
Video-oculography	50-200	<1	No	(Imai <i>et al.</i> , 2005) (Houben <i>et al.</i> , 2006)

Electro-oculography holds the advantage of being non-invasive while also offering a good sampling rate. The electrode placement, however, is tedious and requires some experience, since researchers in the field of EOG have found that the appropriate placement of electrodes are important. Photo-electric oculography is an outdated collection of eye tracking methods. Methods that makes use of reflective lenses are invasive and thus undesirable. Using photo-electric oculography also seems limiting, because it is basically a severely simplified version of VOG. Both these previously mentioned methods also exclude the possibility of torsional tracking, which is another limiting factor. Subsequently, neither EOG nor photo-electric oculography will be considered.

As was mentioned in Section 2.2.1.3, the scleral search coil method is widely expected as the gold standard of eye tracking methods because of its accuracy. However, Imai *et al.* (2005) concluded that there is not a significant difference between video-oculography and the scleral search coil method, since angular measurement differences between the two methods of less than 1° for all axes were achieved, and that video-oculography is therefore an acceptable alternative to the scleral search coil method. Other researchers have achieved similar results (Moore *et al.*, 1991; Houben *et al.*, 2006). However, Houben *et al.* (2006) did find that the torsional component measurement of VOG does not perform as well as the other components against the SSC method. This is in contrast with results obtained by the other above-mentioned researchers, as they all found that the torsional components of the two methods correspond well. Houben *et al.* (2006) attributes these results to the fact that other researchers either made use of an artificial eye with well defined iris landmarks, or a marked sclera. Marking the sclera with a black dot is a widely used technique for torsional tracking, and the invasive nature thereof is what prompted Houben *et al.* (2006) to avoid this method. They do, however, proceed to admit that there were some factors that might have influenced the accuracy of the VOG. This seems logical since researchers using the same commercial system have obtained better results for the VOG method. One last thing to remember is that the algorithm used to determine the 3-dimensional angular displacement of the eye plays a major role in the accuracy of the method, since Moore *et al.* (1991) achieved measurements with torsional errors of less than 0.1° with their cross-correlation algorithm.

Following the previous paragraph, it seems reasonable to argue that video-oculography is an acceptable alternative to the scleral search coil method with regards to accuracy. Perhaps the most notable disadvantage of VOG with regards to the SSC method is the sampling rate. The SSC method is capable of considerably higher sampling rates of typically 500 Hz and above, while VOG is commonly limited to 60 Hz (Imai *et al.*, 2005). However, some digital cameras that make use of firewire or ethernet connections can achieve a sampling rate of 200 Hz. It should be noted that cameras with much higher sampling rates (1000 Hz +) are commercially available, but are extremely expensive. High sampling rates simplify the capture of saccadic motions with extremely high angular velocities. The major advantage of VOG over the SSC method is that it is non-invasive. Another advantage of VOG over the SSC method is the re-usability. Unlike VOG, the SSC method requires a supply of components (scleral lenses), since the system cannot be reused for an indefinite number of times. The advantages discussed above makes VOG an attractive option as a clinical diagnostic device, especially for telemedicine environments.

The non-invasive nature of video-oculography is a decisive factor in choosing an appropriate method for a clinical diagnostic device. Therefore, it was concluded that VOG is the most suitable eye tracking technique to satisfy the goals of the project.

2.3 Safety Study

Developing a product for general use always brings about the issue of safety considerations. This is especially true for clinical devices. Even though the use of video-oculography provides a non-invasive solution for a potentially invasive procedure,

there are still potential health risks that needs to be considered. This section will discuss the foreseen risks and how to mitigate them.

2.3.1 Infrared Lighting Source

Eye safety hazards are the most prominent health risks that need to be considered for an eye tracking device. With eye tracking devices that make use of a scleral contact lens, the risks are rather obvious and physical damage on the cornea can be avoided by designing safer lenses and practising caution when applying and removing the lens from the eye. With video-oculography, however, the concern lies with retinal safety. Damage to the retina, as a result of excessive exposure to certain light conditions, is a potential health risk that needs to be considered. The type and level of retinal hazard involved are dependent in the wavelength of the light. Perhaps the most well-known example of this fact, is the hazard related to short wavelength light, like blue light and ultra violet light, which cause photochemical lesions called photoretinitis (Ham *et al.*, 1976). The different photobiological effects associated with each spectral band are shown in Figure 2.12 (Sloney, 2007). Infrared light typically used for illumination in VOG systems usually fall in the spectral band between 800 nm and 950 nm. This band is represented in Figure 2.12 as the IR-A band, and is frequently referred to as the near infrared band. As can be observed from the figure, hazards associated with this band are mainly thermal damage to the retina or cornea. Subsequently, the intensity limits of the IR illumination needs to be determined. These limits are commonly referred to as the emission limits (EL) or accessible emission limits (AEL) in safety standards.

Nonionizing Radiation Band	UV-C	UV-B	UV-A	VISIBLE	IR-A	IR-B	IR-C	
Wavelength (nm)	100	280	315	400	760	1400	3000	10 ⁶
Adverse Effects	Photokeratitis		Retinal Burns		Corneal Burns			
	Cataract		Cataracts					
	Erythema		Color Vision Night Vision Degradation					
			Thermal Skin Burns					

Figure 2.12: CIE Photobiological Spectral Bands (Sloney, 2007)

The lack of an explicit safety standard for the use of light emitting diodes (LEDs) have caused some confusion about the actual risks involved with LEDs. Safety standards for LEDs have been associated with laser safety (IEC 60825-1) and lamp safety (ANSI/IESNA 1996a,b) which causes even further confusion regarding the level of conservativeness needed to ensure a safe design. In an attempt to clear up some of this confusion, the International Commission on Non-Ionizing Radiation Protection (ICNIRP) released a statement regarding the safety of LEDs (Matthes, 2000). In this article the ICNIRP stated that a maximum radiance of $2.5 \text{ W cm}^{-2} \text{ sr}^{-1}$ and maximum average radiance of $0.18 \text{ W cm}^{-2} \text{ sr}^{-1}$ over a time of 100 seconds can be achieved by existing surface-emitting LEDs. Matthes (2000) concluded that the use of these standard surface-emitting LEDs can be considered harmless based on emission limits guidelines from various standards. Another study that was conducted over the course of five years specifically investigated the hazards of infrared light in VOG applications

(Mulvey *et al.*, 2008). This study was particularly focused on the long term effects of the infrared light, since VOG is typically used by disabled patients on a relatively permanent basis. It included evaluations based on three standards, namely the CIE lamp safety standard, the IEC/CIE lamp safety standard and the EN 60825-1 laser product safety standard. Mulvey *et al.* (2008) concluded that existing VOG systems using IR LEDs do not pose a health hazard, since the radiance of the LEDs do not nearly approach the emission limits instated by the safety standards, even though some of the systems evaluated use illumination arrays of up to 40 LEDs. This conclusion holds, even for VOG use spanning over several minutes.

Concluding from the studies discussed in the previous paragraph that the use of IR illumination sources can be considered harmless when applied with caution and in compliance with the emission limits given by appropriate safety standards, it followed that IR illumination was chosen as the active light source for the purpose of this thesis. An independent evaluation of the IR LEDs used for this thesis was conducted to ensure that the radiance levels were within safe limits. The detailed calculations are presented in Appendix A. Radiance values given by Matthes (2000) and emission limits from the IEC 60825-1 laser safety standard was used as a guideline to determine the safety of the chosen LEDs. It was found that the radiance levels of the chosen LEDs are well within safe limits, stated by both these sources, despite the fact that the use of the IEC 60825-1 laser safety standards can be considered as a conservative approach (Mulvey *et al.*, 2008).

2.3.2 Reflex Epilepsy

Epilepsy can occur as a result of certain visual stimuli. This type of epilepsy is commonly referred to as reflex epilepsy. There are two types of reflex epilepsy that are of interest in this thesis, i.e. photosensitive epilepsy and pattern-sensitive epilepsy. Photosensitive epilepsy is associated with flickering lights or moving images, while pattern-sensitive epilepsy is triggered when viewing certain patterns, and stripes in particular (Zifkin and Trenité, 2000). Both these types of stimuli are frequently used in ocular and vestibular examinations, as well as VOG system calibrations. As mentioned before, optokinetic drums, which are commonly used in ophthalmology, have vertical moving stripes. These tests can generally not be replaced in order to provide a safer test environment for patients suffering from reflex epilepsy, but precautions should be taken to raise awareness about the possible effects these tests may have on patients with these conditions. Enquiring about a history of epilepsy prior to performing examinations that makes use of the aforementioned types of stimuli should be a standard procedure.

Chapter 3

Video-Oculography Device

In section 2.2.1.5 it was concluded that video-oculography provides the most appropriate solution for the purpose of this thesis. The main advantages of VOG that attributed to the selection thereof includes the non-invasive nature thereof and the ease of use. For example, compared to electrooculography, VOG is much simpler and quicker to set up. Diagnostic devices are regularly designed to be non-invasive since it is safer and more comfortable for the patient. The simplicity and user-friendliness of such devices are also important to save time for the medical professionals. This is especially true in telemedicine applications where there is typically a high patient-physician ratio. Therefore, VOG was chosen since it best satisfies the requirements of the diagnostic device for clinical use in both urban and rural applications.

Something that was not discussed thus far, is the type of VOG that will be implemented. Recall from Section 2.2.1.4 that there are various VOG methods. The first choice is between remote VOG and head mounted VOG. Since head mounted VOG systems are generally more accurate because of increased resolution as a result of the proximity of the camera to the eye, it was decided to make use of a head mounted device. This also implies that head motion is of no consequence to the effectiveness of the system. The second choice is the placement of the camera. When the camera is placed in front of the eye, only one eye can be recorded at a time if it is desired to make use of a visual stimulus. However, to gather adequate information for some conditions, it is necessary to track both eyes at the same time. For this reason, it was decided to place the cameras at an angular offset, as shown in Figure 2.10. This method allows both eyes to be recorded simultaneously while providing a clear field of view for the subject. The rest of this chapter will discuss the detailed hardware selection and design of the VOG device.

3.1 Camera Selection

Component selection is the first step for the design process, and the most influential components of a VOG device are the cameras. As was discussed earlier, the frame capture rate, or sampling rate, is one of the most important specifications. Commercial systems, discussed in Section 2.2.1.4, provide relatively high sampling rates.

Interacoustics' VO425 VOG device have IEEE 1394 FireWire cameras which provides a sampling rate of a 105 Hz for binocular applications and 174 Hz for monocular

CHAPTER 3. VIDEO-OCULOGRAPHY DEVICE

applications. The *Chronos Vision* system can sample at 200 Hz and have a special interface PCI card which communicates with the cameras.

Even though some researchers argue that high sampling rates (150-200 Hz) are necessary to accurately capture saccades (Clarke, 1998), no decisive practical or theoretical proof for such a necessity could be found. Recall from Section 2.1.2 that the fast phase of a nystagmus beat is often regarded as a saccade. Enderle *et al.* (2005) provides a set of saccade recordings for different angular displacements. An estimated curve for the relation between saccade magnitude and saccade velocity based on the data given by Enderle *et al.* (2005) is given in Figure 3.1.

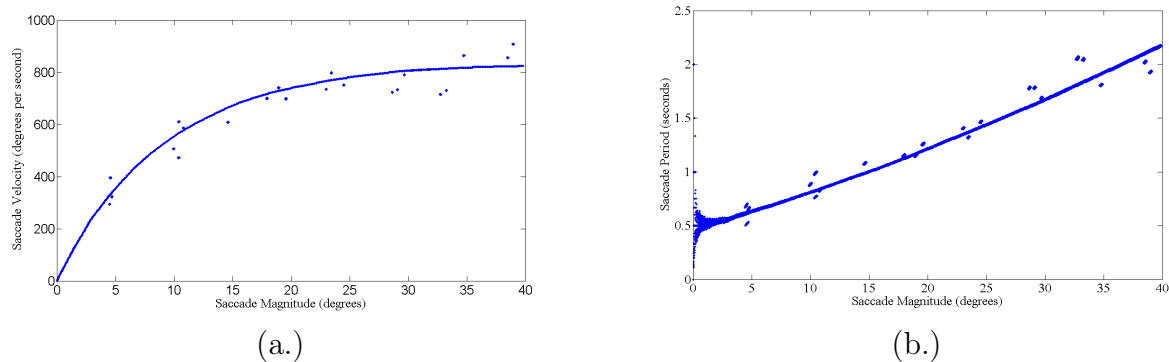


Figure 3.1: Saccade Data: (a.) Saccade Velocity vs. Saccade Magnitude (Original Data) and (b.) Saccade Period vs. Saccade Magnitude

Data used in this Figure were estimated from Enderle *et al.* (2005)

From Figure 3.1 it is clear that saccade velocity is dependent on the saccade magnitude. The worst case scenario with regards to the sampling rate, is high frequency eye motions that could possibly pass without being captured by the cameras. Gradient values of the curve in Figure 3.1 are high at lower saccade magnitudes and then gradually lowers with an increase in saccade magnitude. Thus, the more conservative approach would be to select data at lower saccade magnitudes.

Consider a 5° saccade with an approximate velocity of 350 degrees per second. One complete nystagmus beat would then consist of eye rotation through an angle of 10° . Assuming an unlikely, but more conservatively approached situation where the nystagmus beat consist of two quick phases, the total period of the beat would be 28.6 ms. The frequency of a sequence of these beats would then be 35 Hz. According to the Nyquist sampling theorem, a sampling frequency of at least twice that of the highest frequency in the signal is necessary. Therefore, for a 35 Hz signal, a sampling frequency of at least 70 Hz is required, but preferably around 175 Hz for practical implementations. Although this seems to support the previous statement that high speed cameras are necessary to capture nystagmus, it should be noted that the required sampling frequency estimation discussed above, is subject to highly conservative assumptions. In reality, nystagmus beats with amplitude values of as low as 5° rarely occur. For example, Jansen *et al.* (2010) provoked various types of physiological nystagmus, all of which had average amplitude content higher than 5° . Another conservative factor in the sampling frequency estimation above, is the assumption that the slow phase velocity is equal to the quick phase velocity of the beat, which is incorrect per definition of a nystagmus beat.

CHAPTER 3. VIDEO-OCULOGRAPHY DEVICE

Considering a more likely scenario in which a saccade beat with an amplitude of 5° has a quick phase velocity and slow phase velocity of 350 degrees per second and 175 degrees per second, respectively. This will result in a beat period of 42.9 ms and thus a 23.3 Hz signal. An acceptable sampling frequency of less than 120 Hz is thus required. To further demonstrate the adequacy of lower sampling rates typically provided by VOG systems, examples of practical research can be considered. Jansen *et al.* (2010) made use of a VOG system with a sampling frequency of 50 Hz for nystagmus, while other researchers report typical nystagmus signal content as low as 1-2 Hz, and claim that successful oculography can be achieved with a sampling rate of as low as 30 Hz (Daroff *et al.*, 1979).

It should be noted that the frequency content of the nystagmus depends on the type of nystagmus, but based on observations from the works of other researchers and the theoretical analysis according to the Nyquist sampling theorem, it can be concluded that high speed cameras that achieve sampling rates of 150 Hz or higher are rarely required to ensure the capture of nystagmus. There are various reasons to avoid high speed cameras. The first disadvantage of a high speed camera is the cost. Another drawback to high speed cameras is compatibility, regarding the interface, with the available computer system. Consider for example the commercial systems discussed above. The *Chronos Vision* system requires a PCI card slot, while each of the cameras for the *Interacoustics* device requires a FireWire port. The drawback to the PCI interface is that it requires a desktop computer to install, while clinical use will most often require a portable system. A similar problem arises with regards to FireWire communication protocol. Only high-end laptops have build-in IEEE 1394 ports, and most of those that do, only has one port. Since clinical use of diagnostic tools such as the required VOG device are frequently dependent on portability, especially in telemedicine applications, the above communication protocols pose somewhat of a drawback and impose some limits on the choice of computer that can be used for data processing purposes.

In light of the topics discussed so far, the camera that was selected for the VOG device in this thesis, is the PS3™Eye developed for Playstation® 3 games that make use of computer vision to register command inputs. An image of the PS3 Eye is shown in Figure 3.2. The PS3 Eye camera is considerably cheaper than its FireWire counterparts, but provides a sampling frequency of up to 75 Hz at a resolution of 640×480 and 125 Hz at a resolution of 320×240 . This provides a trade-off option between high sampling rate and tracking accuracy, and therefore presents the opportunity to record even the quickest nystagmus beats albeit at the cost of tracking accuracy. Furthermore, the PS3 Eye provides a field of view (FOV) of 56° at which the focal depth is between 40 and 70 mm, which makes the camera ideal for head mounted VOG systems without any lens adjustments or replacements required. Finally, the interface for the PS3 Eye is USB 2.0, which allows the simultaneous use of two cameras with almost any computer, since most computers have several USB ports.



Figure 3.2: PlayStation® 3 Eye Camera

3.2 Lighting Source

As was discussed in Section 2.2.1.4, IR light is a popular choice of light source for VOG purposes. For head mounted VOG systems, IR lighting serves two purposes. The general reason for the use of IR light in any VOG system is the fact that the light does not agitate the subject, since IR light is not within the visible spectrum. Secondly, IR light presents the opportunity to place the cameras at an angular offset by making use of dichroic filters. Near IR LEDs were selected, since standard camera image sensors are not sensitive to thermal as well as far IR spectrum wavelengths. Selection of specific IR LEDs were subject to safety evaluation as discussed in Section 2.3 and Appendix A. The selected LEDs have an emission wavelength of 940 nm. The lighting source was designed to contain twelve of the selected LEDs mounted on a circular printed circuit board (PCB) as shown in Figure 3.3 in order to provide an invariable light gradient over the eye. Furthermore, in order to eliminate the necessity for an external power supply to power the LEDs, a voltage regulator board that can draw power from a USB port was included in the design. The voltage regulator circuit is shown in Figure 3.4. Since the LEDs never operate at full power, the 500 mA supply from the USB port is sufficient to supply both IR lighting sources simultaneously. As was discussed in Section 2.2.1.4, IR light is a popular choice of light source for VOG purposes. For head mounted VOG systems, IR lighting serves two purposes. The general reason for the use of IR light in any VOG system is the fact that the light does not agitate the subject, since IR light is not within the visible spectrum. Secondly, IR light presents the opportunity to place the cameras at an angular offset by making use of dichroic filters. Near IR LEDs were selected since standard camera image sensors are not sensitive to thermal as well as far IR spectrum wavelengths. Selection of specific IR LEDs were subject to safety evaluation as discussed in Section 2.3 and Appendix A. The selected LEDs have an emission wavelength of 940 nm. The lighting source was designed to contain twelve of the selected LEDs mounted on a circular printed circuit board (PCB) as shown in Figure 3.3, in order to provide an invariable light gradient over the eye. Furthermore, in order to eliminate the necessity for an external power supply to power the LEDs, a voltage regulator board that can draw power from a USB port was included in the design. The voltage regulator circuit is shown in Figure 3.4. Since the LEDs never operate at full power, the 500 mA supply from the USB port is sufficient to supply both IR lighting sources simultaneously.

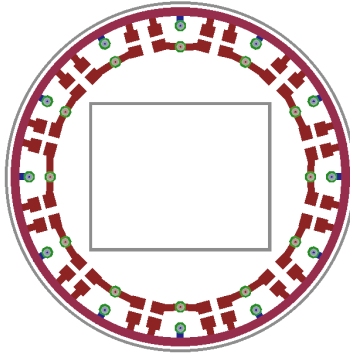


Figure 3.3: IR Lighting Source PCB

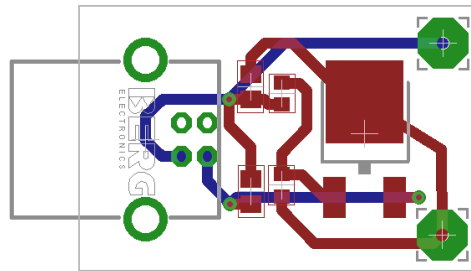


Figure 3.4: LED Voltage Regulator for USB Power

3.3 Filter Selection

In order to implement the system as shown in Figure 2.10 that allows the camera to be placed at an angular offset, it is necessary to implement a suitable imaging design. This design does not only entail the introduction of additional filters, but also the removal of the factory filters included in the camera lens system. As is the case for all standard non-specialized cameras, the PS3 Eye camera has a built-in IR notch filter included in the lens casting. The IR notch filter is included to reduce sharp ambient light intensities, or in other words, glare. Since IR light is the only spectrum of interest for the proposed VOG design, the notch filter is the exact opposite of what is needed. In order to facilitate the proposed VOG design, the IR notch filter is thus removed and an IR bandpass filter is added to the camera lens casting in order to restrict the captured spectrum to the near IR spectrum. Consequently, no visual light is passed to the image sensor, which serves two purposes: Firstly, ambient light outside the near IR band no longer affects the VOG recording, and secondly, the dichroic filter acts only as a mirror and does not appear to be transparent. This is an important property. If the image sensor could capture both IR light and visible light, the image would not only consist of the eye, but would also be distorted by visual light from the opposite side of the dichroic filter. Without an IR bandpass filter the image would thus, in effect, be two separate images superimposed on one another. The image would thus be composed out of the eye as recorded in the IR spectrum, and the opposing camera of the VOG device as recorded in the visual spectrum.

CHAPTER 3. VIDEO-OCULOGRAPHY DEVICE

As was mentioned in Section 3.2, the selected LEDs emits light at 940 nm, and the filters for the VOG system were chosen accordingly. The filters that were selected for the design are the 940FIW25 bandpass filter and the 730FHQ5050 dichroic short-pass filter from *Knight Optical*. The transmittance and reflectance curves for the bandpass filter and dichroic filter are given in Figures 3.5 and 3.6 respectively. The graphs from Figures 3.5 and 3.6 were supplied by *Knight Optical's* technical department.

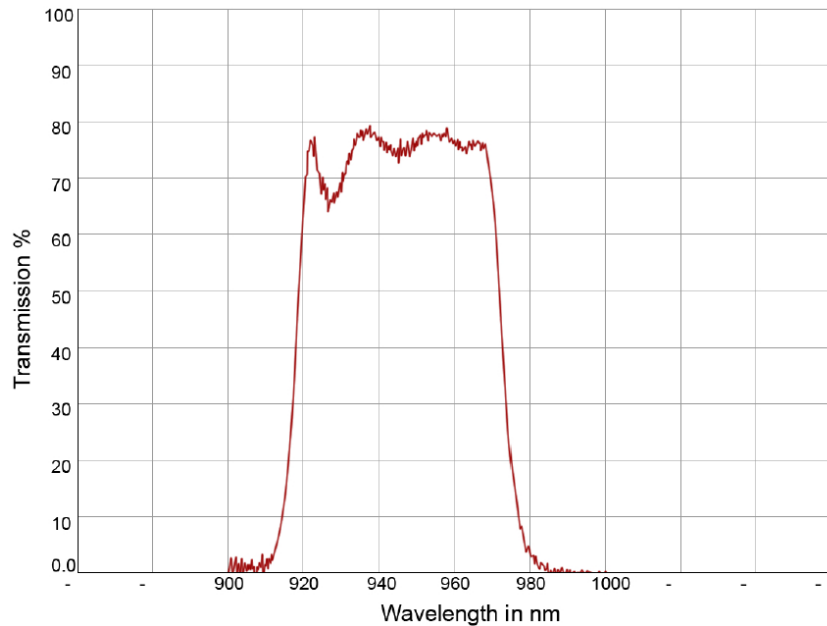


Figure 3.5: Transmittance Curve for 940FIW25 Bandpass Filter

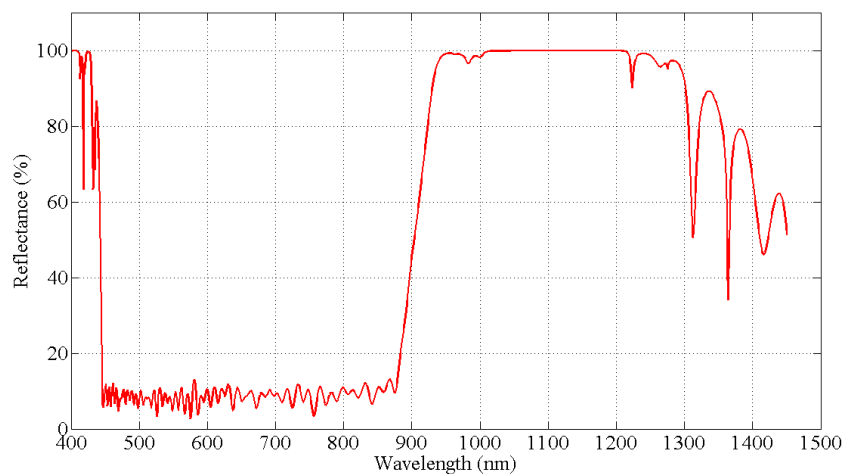


Figure 3.6: Reflectance Curve for 730FHQ5050 Dichroic Short-Pass Filter

Note from Figures 3.5 and 3.6 that the transmittance and reflectance properties are based around a 940 nm wavelength to complement the selected LEDs. The bandpass filter is classified as an extra-broadband filter and was chosen over the regular bandpass filter for two reasons. Firstly, it provides a safety factor in terms of the uncertainty of

emission wavelength of the LEDs which is dependent on the supply current. Secondly, the extra-broadband filter provides a better transmittance than the regular bandpass filter, which is especially important for this design, since it is desired to keep the light intensity to a minimum.

3.4 Lens Selection

The standard lens system of the PS3 Eye camera that was selected for the VOG device only provides limited choices of FOV and focus settings which limits the quality of the image that is captured by the camera. Although the standard lens system provides an adequate image of the eye at the required dimensions for the VOG design, the eye appears in a small window of the image, approximately half the size of the image. Therefore, a better resolution can be achieved if the image is focused such that the eye fills the entire frame. For this purpose, a new lens with the proper FOV at the required working distance was chosen to replace the standard lens built into the PS3 Eye camera. The working distance is defined as the distance between the lens and the object being filmed, in this case the eye. The lens that was chosen is the *Schneider* Compact VIS-NIR M56-533 lens. The VIS-NIR specification refers to the lens' transmittance properties for both, visible and near IR light. Table 3.1 shows the relevant information regarding the lens. The lens was selected from a catalogue based on the required working distance, FOV and physical dimension constraints. Selection of the lens to achieve the appropriate combination of the field of view and working distance parameters are critical to ensure that the optimal resolution of the eye is achieved.

Table 3.1: Lens Specifications

Image Sensor	Horizontal FOV (mm)	Angular FOV (°)	Focal Length (mm)	Working Distance (mm)
$\frac{1}{2}$ " Image Sensor (Catalogue Spec.'s)	10.7 (@ min. working distance)	43.6	8	0- ∞
PS3 Eye	40.4	40.3	custom	55

The increased resolution achieved by the M56-533 lens, as shown in Figure 3.7, results in a finer tracking sensitivity and subsequently a better signal resolution, which is essential to capture nystagmus beats with small amplitude.

The diameter of the pupil in the lower resolution image denoted by d , as shown in Figure 3.7.a, is approximately 44 pixels and the diameter of the pupil in the higher resolution image denoted by D , as shown in Figure 3.7.b, is approximately 103 pixels. Therefore, the detail of the image was increased by a factor of more than two by addition of the M56-533 lens.

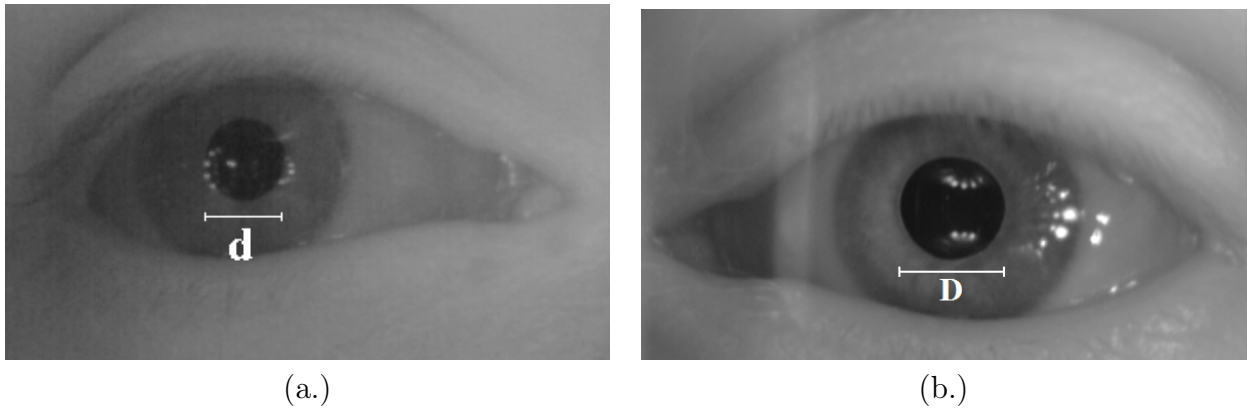


Figure 3.7: Resolution Comparison: (a.) Standard PS3 Eye Lens, and (b.) M56-533 Lens

3.5 VOG Goggle Design

In order to facilitate and integrate all the aforementioned components discussed in this chapter into a single device, a housing in the form of a goggle needs to be designed. The alternative to a goggle design is a wire frame headset device as implemented in the Chronos eye tracker (Clarke, 2004). However, this type of design excludes the possibility of selected vestibular tests that require controlled light environments. Therefore, to include the possibility of all types of tests, a goggle concept was selected.

An initial ergonomic design phase was included to determine a suitable form for the contact end of the VOG goggle that could accommodate a relatively comfortable contact plane for the subject's face. Meshed measurements and limited information from the RSA-MIL-STD-127 standard (RSA-MIL, 1994) was used to develop a low cost prototype with which the adherence comfortability of the form could be tested on various persons to ensure a suitable general form. The results obtained from this initial design phase were incorporated into the prototype design.

The prototype VOG goggle was modified from an ideal design to facilitate construction from machined aluminium stock. The design includes a linear actuator for each camera to enable focus adjustment. Other design features include the placement of the filters relative to the cameras and the eyes to facilitate image formation, as well as an entirely enclosed space via a removable lid that accommodates controlled light environment measurements. The entire goggle design was manufactured from aluminium and the lid was machined from high density polyethylene. Figure 3.8 shows an exploded view of the VOG assembly. The prototype design was subject to limited material and manufacturing constraints and therefore does not reflect a commercial prototype. The design is rather a tool for initial testing and for proof-of-concept. Photos of the manufactured device is shown in Appendix C.

The numbered components in Figure 3.8 are listed in the figure. The mounting socket serves as a assembly connection between the sourced lens and the camera PCB, since the PS3 Eye is not designed for lens exchanges. Another component that has not been discussed yet is the viewing shield. The viewing shield simply provides the option to create a darkened environment such that there is no visual distractions for the subject. This option is frequently used in ocular and vestibular tests.

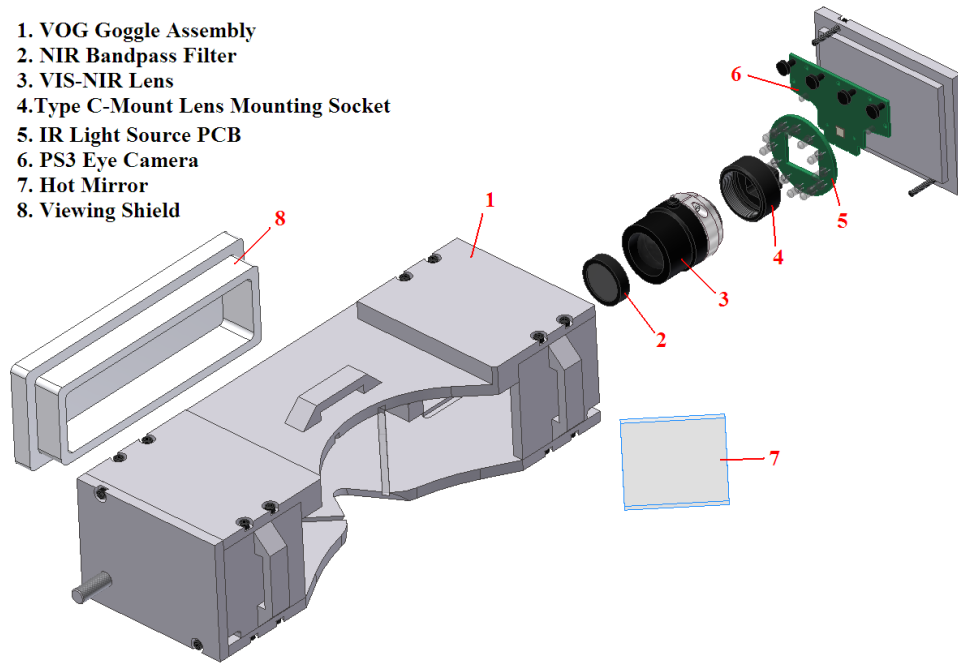


Figure 3.8: Exploded View of VOG Device

This concludes the review on hardware that was designed for this thesis. The VOG device shown in Figure 3.8 was used for all data acquisition throughout the thesis and proved to fulfill its purpose successfully. However, as was mentioned earlier, this is only a prototype design. Designing a commercial device will require a change in material and shape to decrease the weight and provide a more comfortable head mounted solution. Something else that should be taken into consideration is the relative eye placement of each individual. Although no serious placement issues were encountered with this design it was noticed that specific attributes of different individuals, such as distance between eyes, come into play. Since the camera is ideally focused only on the eye this could present problems with a static design and movable lenses might be required for a more versatile design.

3.6 Software Interface

Selection of the PS3 Eye camera for the image sensor of the VOG resulted in the necessity for specially modified software. Tracking software for this thesis was developed in C++, and the open source *OpenCV* libraries were used for image operations and processing. Since the PS3 Eye camera was not originally designed for computer use, unique software drivers for the camera is needed to operate the camera. *Code Laboratories* developed the required drivers, as well as an operation library for the PS3 Eye. This library provides a more reliable method of communication with the camera, and is also a necessity when multiple cameras are operated, since *OpenCV* does not support multiple PS3 Eye camera operations. Appendix F give more detailed discussions on the software development and presents the user interface.

Chapter 4

Pupil Tracking

Pupil tracking is the first step towards identifying nystagmus. Since the visual axis is situated at the centre of the pupil, the position of the pupil can be used to determine the direction of gaze. Recalling that the VOG recording device stays stationary with regards to the head of the subject, it is intuitive that the pupil's position will change as the gaze of the subject shifts. The centre coordinates of the pupil, recorded from a sequence of images, can then be used to perform the required signal processing needed to identify nystagmus. This chapter will review the relevant image processing theory and discuss the algorithm that was designed to track the pupil from a stream of images recorded by the camera.

4.1 Image Processing Fundamentals

Before the algorithms concerning pupil tracking can be discussed, an overview of image processing is necessary. The purpose of this section is to introduce the basic principles of images and the field of study referred to as image processing. Image processing, as the title suggests, makes use of various transformational and morphological operations to manipulate an image in order to achieve a new image that emphasize the data of interest from the original image. Image processing is thus a valuable tool to extract specific shades, features, forms or patterns from an image that can be used to achieve the specified objectives. For the purpose of this thesis, the objective will be to determine the centre of the pupil in a series of images from a video stream in order to record the orientation of the eye over time.

4.1.1 Image Terminology

This section is included to introduce the various terms associated with image processing and to state the mathematical representation of an image. An image is simply a matrix of which the dimensions are the same as the image resolution. The image resolution gives the amount of columns and rows of the image matrix respectively. Thus, a 640x480 image consists of 640 columns and 480 rows. The origin of the image is defined as the top-left corner element of the image. Each element of the matrix represents a small area of the picture/image called a pixel. For example, a single element will

CHAPTER 4. PUPIL TRACKING

represent $\frac{1}{307200}$ th of a 640x480 image where a 640x480 image has 307 200 pixels. The value of each element represents the intensity of the image at that particular position. The intensity of the image refers to the brightness of a certain colour. For grey images, there is only one colour, namely white, and thus each element only has one brightness value which represents the shade of grey displayed. For an 8-bit image, an intensity value of 0 will display a black pixel, while an intensity value of 255 will display a pure white pixel. For colour images, each element has three brightness values, one for each primary colour. A 640x480 colour image will thus be a 480x640x3 matrix, while a grey image with the same resolution will be a 480x640x1 matrix. The conventional order for accessing the different colour brightness values is Red-Green-Blue (RGB). Thus, an element of an 8-bit image with brightness values of (255,0,0) will display a bright red pixel, while an element with brightness values of (255,255,255) will display a pure white pixel. These different intensities for each colour of an element are referred to as channels. Thus, when representing an image as a three dimensional matrix, as shown earlier, the third dimension specifies the number of channels of the image. Figure 4.1 shows the graphical representation of the concepts described above. One final term with regards to an image is the image depth. Image depth refers to the range of the intensity value and thus determines the incremental change in intensity that is possible. An 8-bit image, for example, will have a range of 256 intensity values and an 16-bit image an intensity range of 65 536 values. Some applications will define colour images with 8-bit channels as 24-bit images. However, in this thesis the term ‘image depth’ will always refer to the depth of each channel.

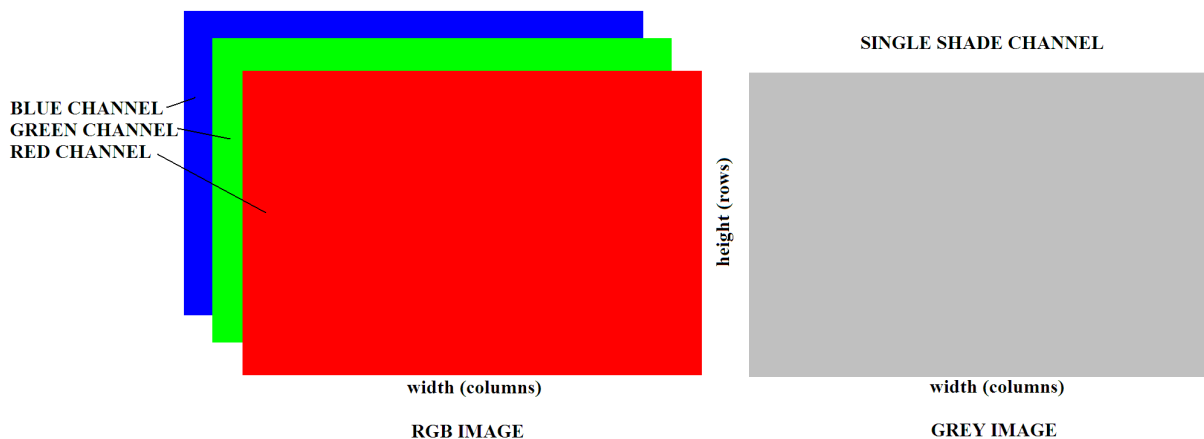


Figure 4.1: Basic Image Formation

Images captured by the VOG system’s cameras are set to grayscale which provides a single channel image. Since the combination of the lighting source and the built-in additional filters for the VOG system causes the image sensor to only capture different intensities of a single wavelength, there is no point in capturing multichannel images. This also reduces the image transfer time from the camera to the computer, since each image element is defined by only 8 bits, instead of 24 bits. The images recorded by the cameras are thus 2-D matrices, usually with a depth of eight bits. All image processing operations are performed on single channel images. Only after the tracking algorithm

determined the pupil centre, the original image is converted into a RGB image such that the estimated pupil centre and boundaries can be displayed in colour.

4.1.2 Image Transformation

Transformation of images can be subdivided into two main groups, namely geometric spacial transformations and intensity transformations. Geometric transformations can be used to scale, rotate, translate and shear or stretch an image. Using a combination of these transformations can also correct a distorted dimensional perspective. Geometrical transformation also includes the mirroring of the image around a certain row or column. Recall from Section 2.2.1.4 that the image of the eye captured by the camera of the VOG system will be mirrored, since the light captured by the camera is reflected from the hot-mirror as shown in Figure 2.10. This is a good example of where geometrical transformation is used in this thesis to mirror the image in order to retrieve the desired form, i.e. the non-mirrored view of the eye for a visual feedback purpose. Intensity transformations will be used more frequently in this thesis, however. Intensity transformation refers to colour or shade based transformation of the image which include threshold operations, brightness and contrast settings of the image and various other operations related to histogram processing. Intensity transformations frequently revolve around the histogram of the image. The histogram for any image is simply an array with the same size as the range of intensity values defined by the image depth. The value of each element in the array specifies the number of pixels in the image with intensity values equal to the element number. Thus, the 0th element of the histogram array will give the number of pure black pixels in the image, and the 255th element will give the number of pure white pixels.

4.1.3 Spacial Filtering

Spacial filtering is an important tool in image processing and the cornerstone of many algorithms. Filtering includes noise removal, sharpening, smoothing, which is a form of noise removal, and the detection of sharp gradients of intensity values. Filtering is implemented by moving a filter matrix over each pixel of the image matrix and calculating the sum of the products of the overlapping elements. These masks have uneven dimensions and are usually square. The result after each calculation is then stored in the filtered image matrix at the position corresponding to the current centre element of the mask with regards to the original image matrix. This operation is mathematically referred to as correlation. Note that a problem arises when using this algorithm. Since it is necessary that the centre element of the mask visits each element of the image matrix, some elements will fall outside the matrix when the mask is applied to border pixels. This problem can be solved by padding the image matrix with zeros. Consider an image with dimensions $M \times N$ and a filter mask with dimensions $m \times n$. Padding the image matrix with $a = \frac{(m-1)}{2}$ zero element rows at top and bottom and $b = \frac{(n-1)}{2}$ columns at the sides will then be sufficient for the filtering operation.

4.1.3.1 Correlation and Convolution

Correlation, as it was explained in the previous paragraph, is mathematically defined in equation 4.1.1, where the hollow star is the correlation operator (Gonzales and Woods, 2008). For this equation, a and b have the same definitions as given in the previous paragraph, while w and f are the mask and image matrices respectively.

$$w(x, y) \star f(x, y) = \sum_{s=-a}^a \sum_{t=-b}^b w(s, t) f(x + s, y + t) \quad (4.1.1)$$

An unwanted effect of correlation is that it gives a reflected response. Consider, for example, the one dimensional problem with a mask $w = \{1\ 2\ 3\ 4\ 5\}$ and a unit impulse function $f = \{0\ 0\ 1\ 0\ 0\}$. Figure 4.2 shows the step-by-step correlation operation of the problem, where f_p denotes the padded function of f and g is the resulting correlated function.

f_p, w	g
$\begin{array}{ccccccccc} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 2 & 3 & 4 & 5 & & & & \end{array}$	$\rightarrow 5\ 0\ 0\ 0\ 0$
$\begin{array}{ccccccccc} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ & 1 & 2 & 3 & 4 & 5 & & & \end{array}$	$\rightarrow 5\ 4\ 0\ 0\ 0$
$\begin{array}{ccccccccc} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ & & 1 & 2 & 3 & 4 & 5 & & \end{array}$	$\rightarrow 5\ 4\ 3\ 0\ 0$
$\begin{array}{ccccccccc} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ & & & 1 & 2 & 3 & 4 & 5 & \end{array}$	$\rightarrow 5\ 4\ 3\ 2\ 0$
$\begin{array}{ccccccccc} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ & & & & 1 & 2 & 3 & 4 & 5 \end{array}$	$\rightarrow 5\ 4\ 3\ 2\ 1$

Figure 4.2: 1-D Correlation Method

The same method as shown in Figure 4.2 can be followed with a two dimensional case as with images. As mentioned before, the reflected response is not desirable, and this is especially true for images. This effect can be avoided by reflecting either the mask or the image before the correlation operation. For example, if the mask w is reflected for the example problem shown above, such that $\{5\ 4\ 3\ 2\ 1\}$ is correlated with the function f , the result of the operation followed in Figure 4.2 will be $g = \{1\ 2\ 3\ 4\ 5\}$. This method is referred to as convolution. Convolution is usually implemented instead of correlation. The mathematical definition of convolution is given by Gonzales and Woods (2008) in equation 4.1.2 and the solid star is the convolution operator.

$$w(x, y) \star f(x, y) = \sum_{s=-a}^a \sum_{t=-b}^b w(s, t) f(x - s, y - t) \quad (4.1.2)$$

4.1.4 Frequency Domain Filtering

Implementing image filters in the spacial domain, as presented in Section 4.1.3, is a computationally expensive method of applying filters to images. Frequency domain

filtering offers the same end results, usually at less computing time. The exception to this rule is when smaller images are evaluated, but in general the frequency domain operations are much faster than a brute-force implementation of spacial domain filters. Frequency domain filtering therefore offers a major advantage over spacial filtering, especially in image processing applications where computations are inherently expensive. The paradigm of frequency domain filtering is based on the Fourier Transform (FT). The Fourier transform presents the possibility of converting an image to the frequency domain where filtering operations can be performed. The inverse Fourier transform (IFT) can be used to reconstruct any signal or image in the frequency domain to its original form in the spacial domain without the loss of any information. A detailed description of the Fourier transform and the derivatives that are required for practical implementation thereof, are presented in Appendix B. Convolution in the spacial domain, as discussed in Section 4.1.3.1, is simply achieved by multiplication in the frequency domain as shown in equation 4.1.3. Refer to Section B.4 for the derivation of this equation.

$$\mathcal{F}\{f(t)\star h(t)\} = H(\mu) \int_{-\infty}^{\infty} f(\tau) e^{-j2\pi\mu\tau} d\tau = H(\mu)F(\mu) \quad (4.1.3)$$

4.1.5 Morphological Image Processing

Morphological image processing is used to manipulate certain features in an image by thinning, thickening, eroding, dilating and filling of specific regions. Morphological images are usually performed on binary images where values of 0 are considered empty elements and values of 1 are considered full elements. In practice, the image will still have the same depth as the original image. Therefore, all the pixels of an 8-bit image prepared for morphological processing with a threshold operation will either be 0 or 255. Morphological operations are usually performed with something called a structuring element (SE). Structuring elements are used as a reference to determine which of the points in the image is contained within the set of interest. SEs are usually square with uneven dimensions such that the SE is symmetrical around a centre pixel. The remainder of this section will be dedicated to some relevant morphological operations and will also demonstrate the use of SEs.

4.1.5.1 Erosion

Mathematically, erosion of a set A by set B is defined as the set of all points z such that B translated by z is contained in A, where A is the set of full elements (white pixels) in the image and B is the set of full elements in the SE and both A and B are sets in \mathbb{Z}^2 (Gonzales and Woods, 2008). The symbolic mathematical definition for erosion, as given by Gonzales and Woods (2008), is given in equation 4.1.4.

$$A \ominus B = \{z | (B)_z \subseteq A\} \quad (4.1.4)$$

Therefore, an erosion algorithm simply moves the centre of the SE over each pixel of the image and if the full elements of the SE perfectly corresponds to that of the image area

CHAPTER 4. PUPIL TRACKING

with same dimensions around the pixel that is being evaluated, said pixel is included into the set that describes the eroded image. Otherwise the pixel is excluded from the set, and thus the original image is eroded. Note that for these types of algorithms where the centre of the SE is moving over each pixel in the image, the border pixels cause part of the SE to fall outside of the image. To enable evaluation of these border pixels, the image is padded with $\frac{(d-1)}{2}$ wide borders, where d is the dimension of the square SE.

Consider the solid body in Figure 4.3. The vertical gaps in the body have a width of 25 pixels and the horizontal gaps have a width of 30 pixels. Eroding the body with a 21×21 square structuring element yields the result shown in Figure 4.4.

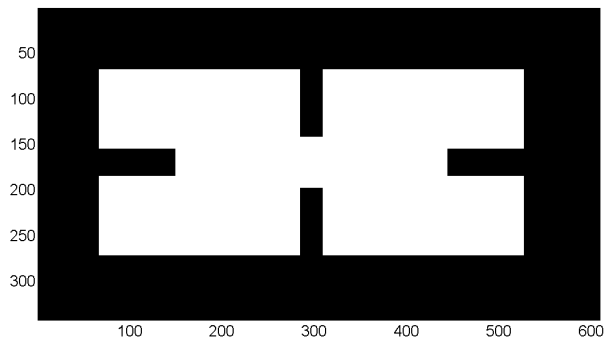


Figure 4.3: Morphological Transformation Demonstration Body

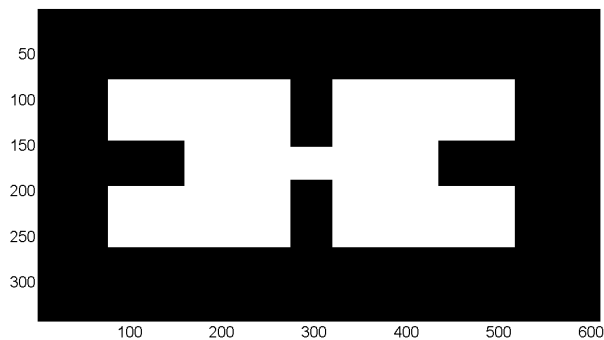


Figure 4.4: Erosion of Demonstration Body

Notice that all the edges of the body are eroded by 21 pixels in Figure 4.4. Therefore, not only does the gaps widen, but the dimensions of the body are also lowered.

4.1.5.2 Dilation

Dilating A by B entails the reflection of B around its origin and then finding the set of all points z such that when \hat{B} is translated by z , at least one element of \hat{B} corresponds to an element of A . For this definition, A is the set of points that describes the filled elements in the image, B is the set of filled elements of the SE and \hat{B} is the reflection of

B, while both sets, A and B, are sets in Z^2 (Gonzales and Woods, 2008). The symbolic mathematical definition, as given by Gonzales and Woods (2008), is given in equation 4.1.5.

$$A \oplus B = \{z | (\hat{B})_z \cap A \neq \emptyset\} \quad (4.1.5)$$

Simply stated, dilation of an image is implemented by moving the centre of the SE over each pixel of the image and filling each element of the image corresponding to the SE's filled elements if the pixel being evaluated is filled. Just as for erosion, the image is padded to avoid evaluation of non-existing pixels within the algorithm. Dilating the body in Figure 4.3 with a 25×25 square structuring element will give the result shown in Figure 4.5.

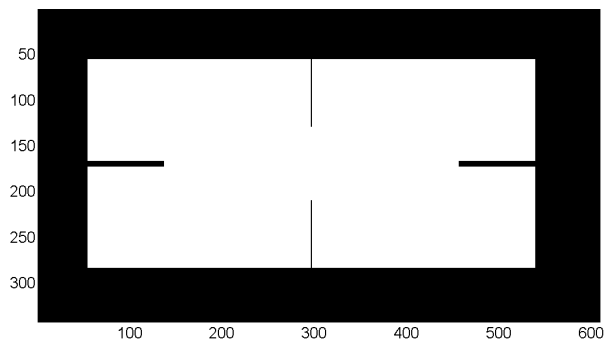


Figure 4.5: Dilation of Demonstration Body

Recall that the vertical gaps have widths of 25 pixels. The single column remaining after dilation in Figure 4.5 demonstrates how gaps can only be filled completely with SE larger than the gap itself.

4.1.6 Opening and Closing

Opening and closing operations are based on the combination of erosion and dilation. Opening is defined as eroding the image A with structuring element B and then dilating the result with B. Equation 4.1.6 gives the symbolic definition of opening for image A with structuring element B. The most intuitive way to understand opening and closing, is to visualize the intensity values of a grey scale image to form a three dimensional surface. Opening is then described as the highest intensity value of any element in the SE that can be achieved when the SE approaches the intensity surface from below. Similarly, closing is the lowest intensity value that can be achieved when the SE approaches the surface from above. Closing is defined by first dilating the image A with structuring element B and then eroding the result with B. The symbolic definition for closing of image A with structuring element B is given in equation 4.1.7. A 2-D graphical representation for the visual concept for opening and closing, as explained above, are shown in Figure 4.6.

$$A \circ B = (A \ominus B) \oplus B \quad (4.1.6)$$

$$A \bullet B = (A \oplus B) \ominus B \quad (4.1.7)$$

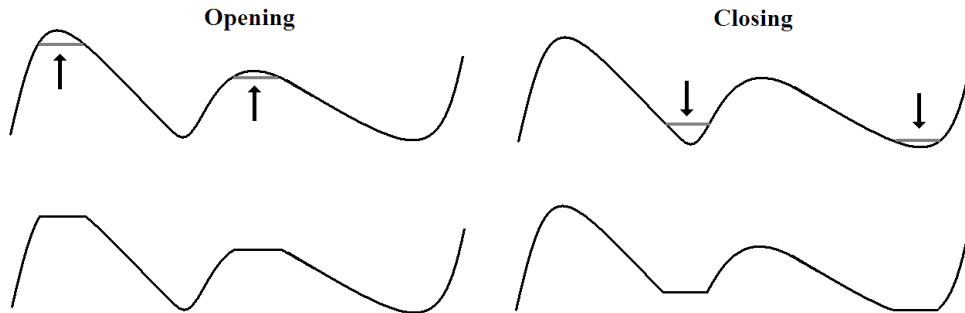


Figure 4.6: Opening and Closing Operations on 1-D Intensity Array

Performing a closing operation on the body shown in Figure 4.3 will result in a body with the same dimensions, while the gaps in the body will be thinner. Figure 4.7 shows the result of the body shown in Figure 4.3, closed with a square 27×27 structuring element.

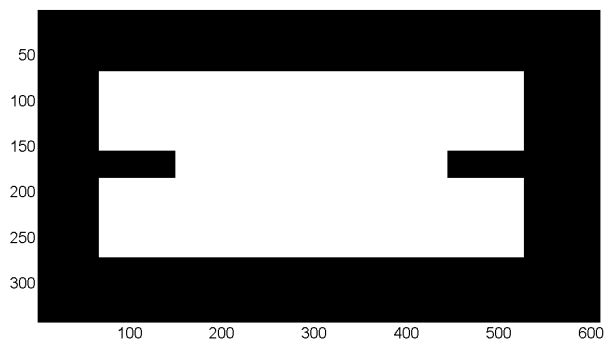


Figure 4.7: Closed Demonstration Body

Note from Figure 4.7 that the 27×27 SE is just large enough to close the thinner vertical gaps of the body shown in Figure 4.7, while the horizontal gaps stay open and the base dimensions of the body remain the same. This property of closing is useful when it is desired to remove artefacts from a feature in an image by filling up gaps. However, care should be taken when choosing a structuring element such that defining features in the image are not affected.

Consider the negative of the body in Figure 4.3. The inner part in Figure 4.3 can thus be seen as a hole and what use to be the gaps will be extrusions. Opening of this new feature with a 25×25 square SE would result in an exact copy of the feature, while opening with a 29×29 square SE would result in the disappearance of the thinner vertical extrusions.

4.1.7 Edge Detection

One last group of image processing algorithms that should be reviewed before the pupil tracking algorithm can be presented, is edge detection. Finding edges in an image have multiple purposes and are subroutines for a great number of image processing algorithms. Thus, edge detection is a well-established paradigm in the field of image processing and a number of methods exists. The most common methods implemented for edge detection are the difference of Gaussians, Canny-, Sobel- and Laplace edge detectors. First, it is necessary to define an edge. In images, edges are merely steep intensity gradients. In other words, there is a large difference in intensity between neighbouring pixels. The edge detection method that will be used throughout this thesis is the Sobel edge detector. The Sobel operator calculates the gradient of the image by implementing frequency domain convolution of two gradient masks in directions perpendicular to each other. Masks with dimensions 2×2 are perhaps the most intuitive solution, but are impractical to implement, since masks with uneven dimensions are more appropriate for convolution. Subsequently, 3×3 masks are most commonly used. The Prewitt detector is very similar to Sobel edge detector, except that Prewitt makes use of non-weighted masks. Thus, all the relevant elements concerning gradient calculations are unity. However, Sobel edge detectors make use of weighted masks that improve noise suppression (Gonzales and Woods, 2008). The sum of the elements for any of these are always zero, since a zero gradient needs to result from a smooth, constant intensity region. Figure 4.8 shows one possible mask pair for gradient calculation.

-1	-2	-1	-1	0	1
0	0	0	-2	0	2
1	2	1	-1	0	1
(a.)			(b.)		

Figure 4.8: Sobel Masks: (a.) Vertical Gradient Component and (b.) Horizontal Gradient Component

The mask shown in Figure 4.8.a is used to approximate the vertical gradient component at each position in the image and the components are collectively denoted as g_y . The horizontal gradient components are denoted by g_x and is approximated using the mask shown in Figure 4.8.b. The gradient vector of an image denoted by f is given in equation 4.1.8. The magnitude of the gradient vector is simply calculated as shown in equation 4.1.9. Some algorithms even approximate the magnitude by simply taking the sum of the absolute values of the two gradient components in order to reduce the computational load (Gonzales and Woods, 2008). The orientation of the gradient, denoted by α , is given in equation 4.1.10. The orientation of the gradient is seldom used since most algorithms are only concerned with extracting the edge lines and therefore only need the magnitude of the gradient. Reference to the gradient image will thus always refer to the magnitude of the gradient.

$$\nabla f = \begin{bmatrix} g_x \\ g_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \quad (4.1.8)$$

$$M(x,y) = \sqrt{g_x^2 + g_y^2} \quad (4.1.9)$$

$$\alpha(x,y) = \tan^{-1} \left[\frac{g_y}{g_x} \right] \quad (4.1.10)$$

For demonstration of the Sobel edge detector, Figure 4.9 shows the gradient image obtained by combining the Sobel masks shown in Figure 4.8 in the manner defined by equation 4.1.9.

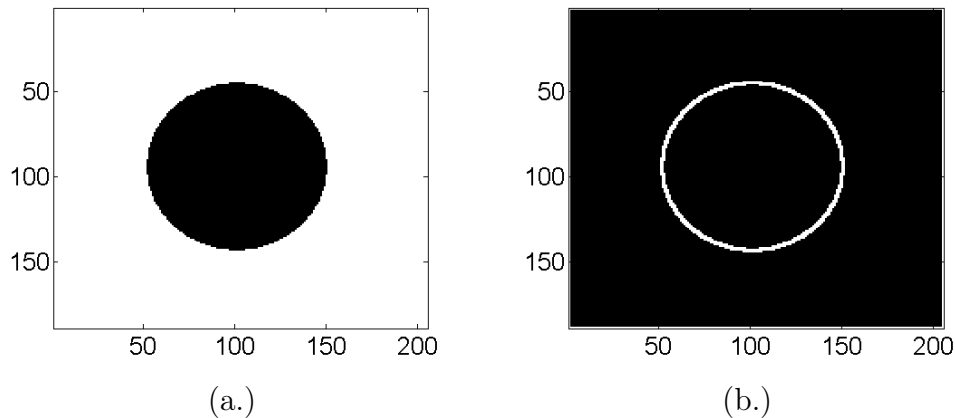


Figure 4.9: Sobel Edge Detection: (a.) Original Image and (b.) Sobel Edge Image

Figure 4.9 is a well-defined image with only two shades of colour which produces an edge image that is similarly well defined. Real world images, however, produce very little zero gradients, since constant intensity regions are almost always scarce. In most cases it is important to limit the number of non-zero pixels in a gradient image since the computational load is dependent on the number of edge pixels when the gradient image is used in subsequent subroutines as discussed in Sections 4.2.2.1 and 4.2.2.2. Therefore, it is common practice to discard gradient values that are not considered to be of any interest. There are two basic steps of achieving fewer edge pixels. The first step is to smooth the image by means of a simple smoothing filter, which results in smaller gradient values. The final step is a threshold operation that discards low gradient values.

4.2 Tracking Algorithms

All the necessary image processing tools to design an appropriate tracking algorithm have been introduced in the preceding sections of this chapter. From this point forth the chapter will include sections on the various algorithms that were considered for pupil tracking purposes in this thesis. There are two fundamental methods to identify and extract the pupil from an image. Perhaps the most intuitive method to extract the pupil is to exploit the lack of colour thereof. Recall from Section 2.1.1.1 that the pupil is merely an opening in the eyeball, and therefore reflects no light. As was discussed in Section 2.2.1.4 there are, however, reflections of the IR LEDs from the cornea. Despite this occurrence the pupil is still, on the most part, detectable as a dark region in the

CHAPTER 4. PUPIL TRACKING

image. This presents the opportunity to extract the pupil by implementing intensity transformations. The second method is to make use of feature extraction algorithms that are concerned with the circular shape of the pupil. The rest of this chapter discusses the implementation of such methods in order to enable pupil tracking.

4.2.1 Histogram Processing

Intensity transformations are simple, yet attractive due to the computational efficiency. Histogram processing in order to extract the dark region from an image is relatively simple compared to most image processing algorithms. Ultimately, these methods boil down to a simple threshold operation. The average of the coordinates for the set of points that remain after the threshold operation can then be considered to be the pupil centre. Unfortunately, as with most algorithms, the corneal reflections, or glints, cause an inaccurate approximation of the pupil centre. Consider a glint that moves across a pupil as shown in Figure 4.10. An error in approximation is produced when following the averaging operation explained above. Figure 4.11 shows the error resulting from the different sized glints moving over a pupil with a radius of 50 pixels. The vertical dotted lines denoted by x_l , x_c and x_u respectively, indicate the minimum x -coordinate, mean x -coordinate and the maximum x -coordinate of the set of coordinates that defines the pupil. For this example, the x -axis is in the same direction as the glint movement shown in Figure 4.10.

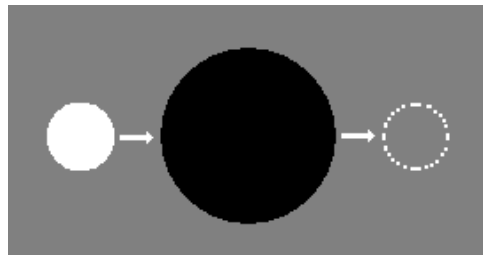


Figure 4.10: Glint Shifted Over Pupil

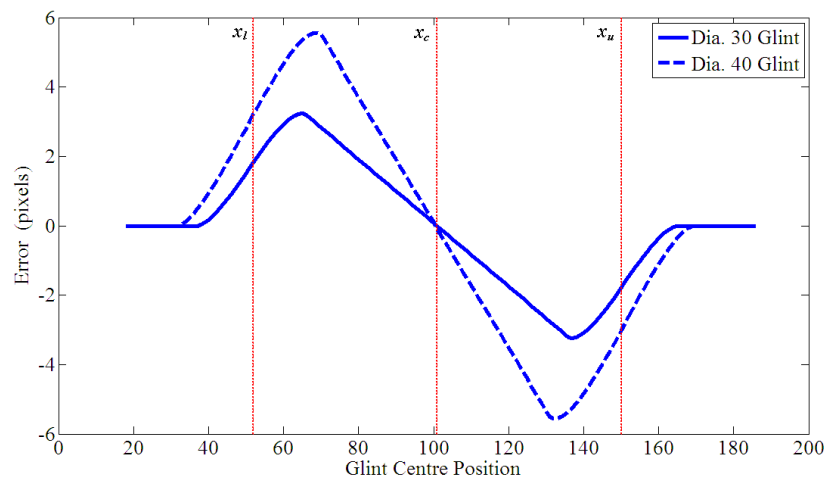


Figure 4.11: Error Produced by Glints Present in Pupil Region

CHAPTER 4. PUPIL TRACKING

The error shown in Figure 4.11 was approximated with the glint offset from only one axis. When the glint is shifted off centre on both axes, the error will be even larger. Furthermore, the IR lighting source consists of multiple LEDs that could potentially cause an even larger error. It can therefore be concluded that the effect of glints in the image cannot be ignored when approximating the centre of the pupil. The *Starburst* algorithm presented by Dongheng *et al.* (2005) makes use of an adaptive threshold technique that performs successive threshold operations at decreasing threshold values to identify the corneal reflex in an image. Removal of the corneal reflexes are then achieved by radial interpolation between the centre and the contour of the identified corneal reflex. However, Dongheng *et al.* (2005) does not present any examples of how this algorithm performs when corneal reflexes are located on the boundary of the pupil.

An intensity transformation algorithm was designed for the purpose of this thesis in an attempt to remove glints by means of bright region consumption based on the intensities of neighbouring non-glint pixels. Thus, glint pixels would undergo an intensity transformation to blend with its surrounding non-glint neighbours by means of an averaging operation. Two fundamental problems exist with this algorithm. The first is that the threshold value, that separates glint intensities from the intensities which define the rest of the image, is not easy to determine. Secondly, the effectiveness of the intensity transformation for glints located on the boundary of the pupil is dependent on the starting point and direction of the process, i.e. the start location and whether the row and column positions are incremented or decremented. Consider the image in Figure 4.12 that represents a pupil and two glints. One glint is entirely contained within the pupil, while the other is located on the boundary of the pupil. Figure 4.13 shows the effect of the proposed intensity transformation algorithm. Figure 4.13.a shows the effect of the algorithm when started at the top-left corner and moving diagonally towards the bottom-right corner, while Figure 4.13.b is achieved by starting at the bottom-right corner and moving towards the top-left corner. Figure 4.13.c shows the result of combining the other two results.



Figure 4.12: Glinted Pupil Test Image

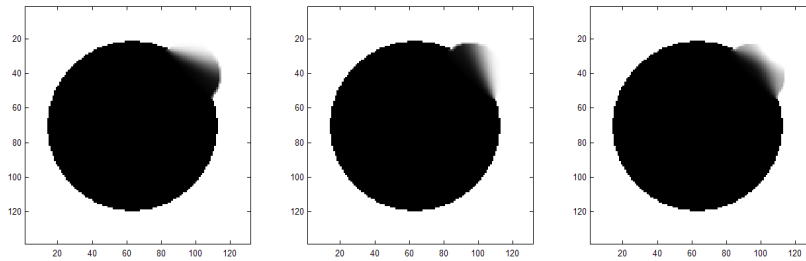


Figure 4.13: Effects of Proposed Intensity Transformation Algorithm

Following the previous discussions regarding glint removal via intensity transformation and the problems associated with glints located on the boundary of the pupil, it can be concluded that intensity transformation is not the optimal solution to achieve a robust tracking solution that includes glint removal.

4.2.2 Feature Tracking

Intensity transformations associated with histogram processing algorithms, such as those discussed in Section 4.2.1, focus on colours within an image. Feature tracking is concerned with the shapes in an image. The circular shape of the pupil presents the opportunity to locate the pupil centre by some sort of feature fitting algorithm. Most algorithms that make use of feature tracking, implement ellipse fitting to an edge image. Naturally, the edge image needs to be preprocessed to remove points that are not related to the pupil. This includes the problem of glint removal. Kumar *et al.* (2009) developed an algorithm that removes glints from the gradient image by filling up completely encircled regions and then following a series of steps that determines which regions should be deleted. Not all algorithms rely on gradient images. As discussed in Section 4.2.1, the *Starburst* algorithm first removes the glints with an intensity transformation before finding the pupil boundaries by starting at a guessed pupil centre and then moving radially outwards (Dongheng *et al.*, 2005). The methods of obtaining the appropriate set of points for feature extraction therefore differs from one algorithm to the next. However, the main objective of all feature tracking algorithms is to extract the edge of the pupil in some way or another, such that the centre of the pupil can be approximated.

4.2.2.1 Circular Hough Transform

A rather innovative method of extracting features from an image is the Hough transform. The Hough transform is based on a patent by Paul Hough (Duda and Hart, 1972). The basic concept of the Hough transform is to gather information from a gradient image into an accumulator and then extracting feature information from the

CHAPTER 4. PUPIL TRACKING

accumulator based on the elements with the highest value. This section will focus on the circular Hough transform since the pupil is the target feature.

The first step of the Hough transform is to acquire the gradient image, since only the edges of the features are of interest. After edge detection, the Hough transform algorithm visits each pixel in the gradient image and performs the subsequent subroutine for each non-zero element. Recall from Section 4.1.7 that the amount of edge pixels in the gradient image can be reduced by smoothing and a threshold operation in order to reduce the computational load of the transform. The subroutine referred to above is the accumulation routine. Each edge pixel is used as the centre point for a new circle. A digital circle is drawn around each of these centre points, and the value of each element that lies on the circle is then incremented in the accumulator which is usually of the same dimensions as the image. Please refer to Section D.1 for more information on digital circles. The radius of the circle, r , is determined by the radius of the target feature, thus the radius of the pupil in this case. After each pixel has been visited, the elements in the accumulator with the highest values can be expected to be the most probable locations for the centre of a circular feature with radius r . In the particular application of pupil tracking, it is reasonable to assume that the pupil will be the only circular feature with its specific radius and therefore only the highest value is extracted from the accumulator and is assumed to be the centre of the pupil. The accumulator matrix is generally known as the Hough transform of the image, and can be displayed as an image itself. The brighter the region in the accumulator, the more circles have intersected with that particular point, and thus the higher the possibility that the point could indicate the centre of a circle. Figure 4.14 shows the Hough transform for the gradient image in Figure 4.9.b. The radius parameter of the Hough transform was chosen to be the same as the circle in the image. Notice the bright spot in the centre of the image where the circles, drawn around each edge pixel, intersected the most.

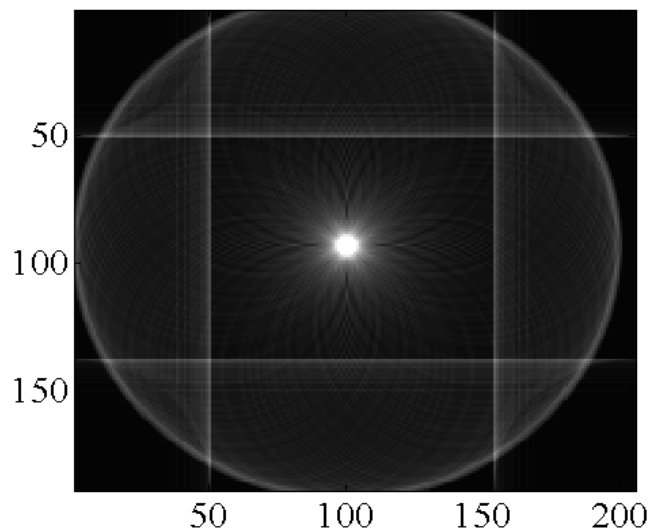


Figure 4.14: Circular Hough Transform

The Hough transform is extremely robust, since features that are unrelated to the pupil are inherently disregarded. Consider, for example, a glint located on the border of the

pupil. The edge pixels of the glint will cause accumulation of elements away from the actual pupil centre without affecting the end result. The small arc on the pupil border covered by the glint will not attribute to the accumulation of the element at the centre of the pupil and thus the element will simply have a lower value. Assuming that a significant portion of the pupil border is detectable in the gradient image, the pupil centre extracted from the Hough transform will remain accurate. Even though the Hough transform presents a robust solution for pupil tracking, it is still not an ideal method for pupil tracking. One disadvantage of the Hough transform is the fact that it is only suitable for finding circles. This poses a problem when the gaze of the subject is shifted to the far lateral position and the pupil appears as an ellipse in the image. It should be noted, however, that some researchers have succeeded in developing algorithms that can identify ellipses by means of a modified Hough transform (Goneid *et al.*, 1997). Another drawback of the Hough transform is that it is, computationally speaking, highly expensive. Finally, the Hough transform can only search for circles of a single radius at a time. Since the pupil is free to dilate and contract, the size of the pupil could change from frame to frame. Although the radius of the pupil in each frame is a good indication of the pupil radius for the next frame, it is still required to search over a range of radii to find the correct centre of the pupil. Each of these radii require a separate Hough transform evaluation which increases the computational load drastically.

4.2.2.2 Ellipse Fitting

One of the disadvantages of the Hough transform that was discussed in the previous section, is that the Hough transform can only accurately find circles and therefore the pupil cannot be tracked successfully as it becomes distorted at far lateral positions. Ellipse fitting eliminates such problems, since an ellipse can be fitted to circles, as in the case where the pupil is viewed from the front, and to elliptical shapes such when the pupil is rotated away from the visual axis and appears to deform into an ellipse. Ellipse fitting is frequently implemented by least squares minimization. Least squares minimization is a popular method for feature extraction in computer vision applications and some algorithms are modified to improve numerical stability (Halíř and Fusser, 1998). Least squares fitting algorithms are extremely sensitive to outlying points that are not part of the ellipse boundary and therefore some researchers abandon the least squares fitting entirely and implement alternative methods. As was mentioned in the previous section, modifications of the Hough transform is one option (Goneid *et al.*, 1997). Another option is the Random Sample Consensus (RANSAC) algorithm (Fischler and Bolles, 1982). An example of where the RANSAC algorithm is implemented for pupil tracking purposes, is the *Starburst* algorithm by Dongheng *et al.* (2005). These types of methods that do not implement least squares fitting usually have considerable computational loads, which motivates the use of the least squares fitting methods.

4.3 Proposed Tracking Algorithm

Throughout this chapter the required image processing techniques and various pupil tracking paradigms have been reviewed to form a sufficient foundation for the development of a suitable tracking algorithm. This section will present the proposed methods of tracking that were developed for the purpose of this thesis. As was discussed

CHAPTER 4. PUPIL TRACKING

in Section 4.2.1, intensity transformations' capability to locate the pupil centre is limited when glints are factored in, especially when one or more glints are located on the pupil boundary. However, intensity transformations are extremely useful to obtain the general location of the pupil and to isolate the pupil boundary. Therefore, a hybrid algorithm is proposed.

Before the final algorithm implemented in this thesis is presented, another hybrid algorithm initially proposed for this thesis should be reviewed. In this algorithm, an adaptive threshold operation is implemented to create two separate images from the original frame. The first image contains the collection of low intensity elements that represent the pupil and the second image contains the high intensity elements that represents the glints. An edge detection operation is performed on each of these images. The gradient image of the image containing the pupil pixels is obtained by performing edge detection with 3×3 Sobel masks, and the gradient image of the image containing the glint pixels is obtained by performing edge detection with 5×5 Sobel masks to produce thicker edge lines. The reasoning behind ensuring thicker edge lines for the glints stems from the following step. It stands to reason that the two images will share similar edge lines, since glints located within the pupil will appear as a hole in the thresholded image of the pupil. The following step is then to perform an element-wise subtraction of the glint edge image from the pupil edge image. The thicker edge lines of the glints will ensure that all traces of the edge lines caused by glints in the pupil edge image is deleted. Note that negative intensity values simply become zero for image operations. Thus, elements that are negative in the difference image as a result of filled elements in the glint edge image deducted from empty elements in the pupil edge image simply become zero. Therefore, all that is left in the difference image is the pupil edge. Glints located on the boundary of the pupil will cause the pupil edge to be partially incomplete. Assuming that the glints will not cover a significant portion of the pupil, the edge points that are left will still provide sufficient information to accurately fit an ellipse.

Consider again the test image of the glinted pupil in Figure 4.12. The resulting images of the steps explained above are shown in Figure 4.15. The images after the threshold operations to extract the pupil and glint pixels, are shown in Figures 4.15.a and 4.15.b respectively. The subsequent edge images for the pupil and glints are shown in Figures 4.15.c and 4.15.d respectively. Finally, Figure 4.15.e shows the difference image containing only the visible portion of the pupil edge.

Theoretically, this method is a perfect solution for pupil tracking. In practice, however, it is challenging to develop an adaptive threshold operation for a robust algorithm that can successfully extract the complete sets of points that define the pupil and glints in the image. Therefore, a robust practical implementation could not be achieved.

Despite the complications with the method described above, a hybrid algorithm still seems like the most elegant solution for pupil tracking. The final tracking algorithm suggested and implemented in this thesis also makes use of a hybrid algorithm, but the focus lies with the extraction of the pupil, rather than elimination of the glints in the image. The pupil is originally extracted based only on colour. For this purpose, an automatic threshold routine is implemented. The routine searches for the first significant peak in the histogram by identifying the largest group of dark pixels within the lower portion of the histogram array. This is achieved by moving a narrow band over the selected portion of the histogram array and finding the position where the sum

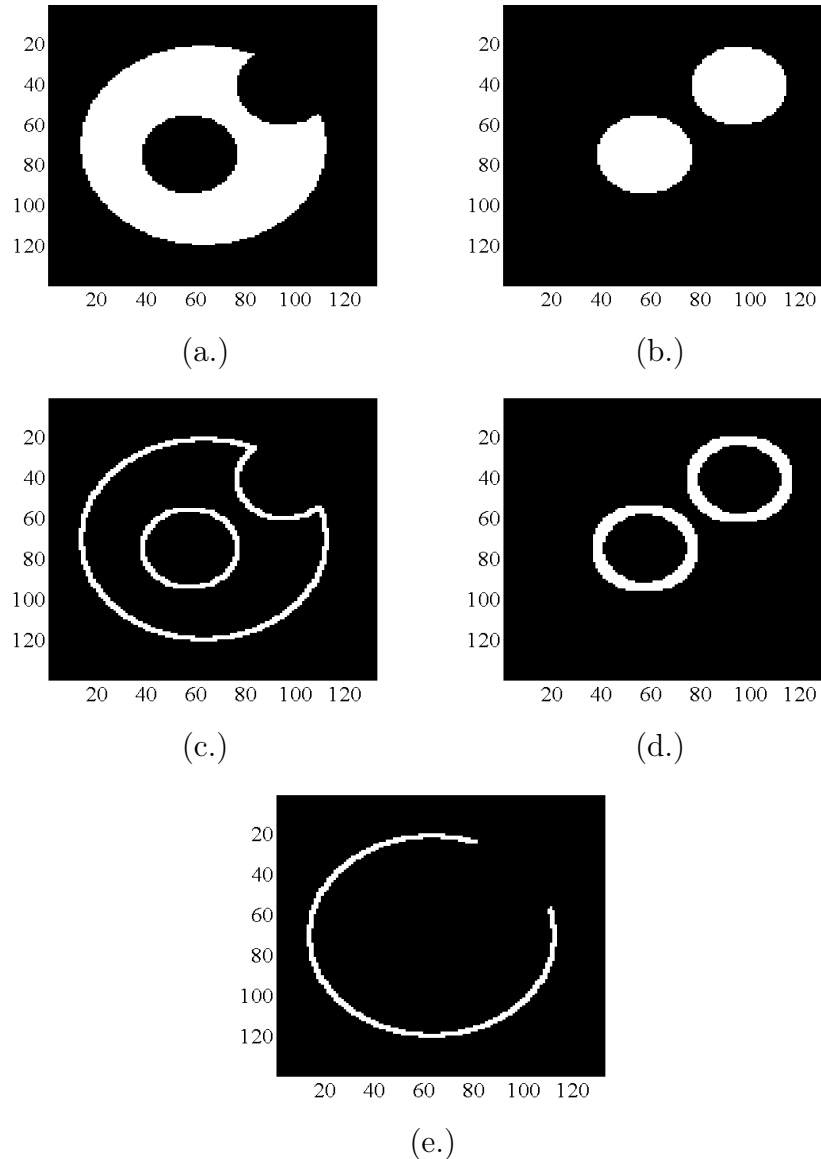


Figure 4.15: First Proposed Hybrid Pupil Tracking Method

of the histogram bin values within the band is at a maximum. From the controlled light environment it can be expected that the pupil is the largest collection of the dark pixels defined by the extracted band. Upper and lower bound threshold operations are then performed on the image in order to extract the pupil from the image.

Consider the real world VOG image in Figure 4.16. Figure 4.17 shows the histogram of the VOG image and demonstrates the implementation of the automatic threshold detection algorithm via histogram processing. The solid green line represents the position of the summation band denoted as p . The dotted green lines represents the limits of the summation band with bandwidth of $2a$. Finally, the dashed red line indicates the maximum value of p and is denoted as p_{max} . Thus, the threshold value for the image, denoted by p_T , is chosen such that:

$$H(p_T) = \max \left\{ \sum_{x=(p-a)}^{(p+a)} H(x) \right\} \quad p = a, a + 1, a + 2, \dots, p_{max} \quad (4.3.1)$$

CHAPTER 4. PUPIL TRACKING

where H denotes the histogram array of the original image.

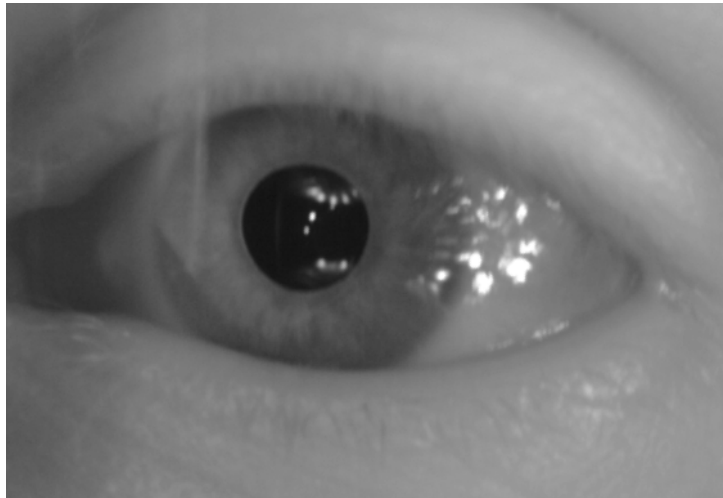


Figure 4.16: Real World VOG Image

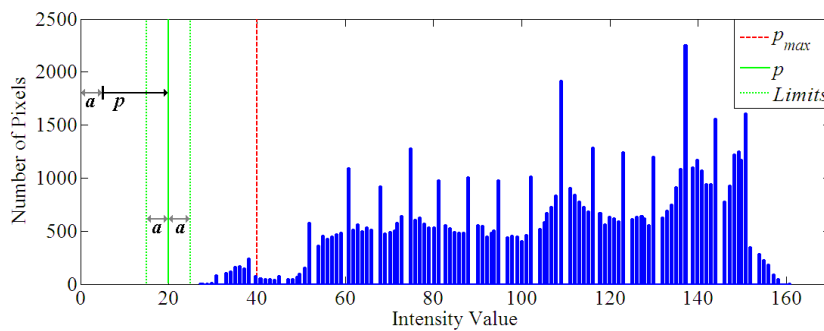


Figure 4.17: Histogram Processing of VOG Image

The reason for including a lower bound threshold is to reduce artefacts from malfunctioning sensors on the image sensor array, as well as the elimination of poorly lit regions in folds of the skin. As example, notice the standalone points away from the pupil in Figure 4.19.a. As demonstration of the algorithm, the real world image shown in Figure 4.16 will be analysed. It represents a single frame from a VOG video, and this specific frame has been chosen to demonstrate the accuracy of the algorithm, despite the fact that glints are located on the boundary of the pupil. Figure 4.18 shows the result of the threshold operation discussed above.

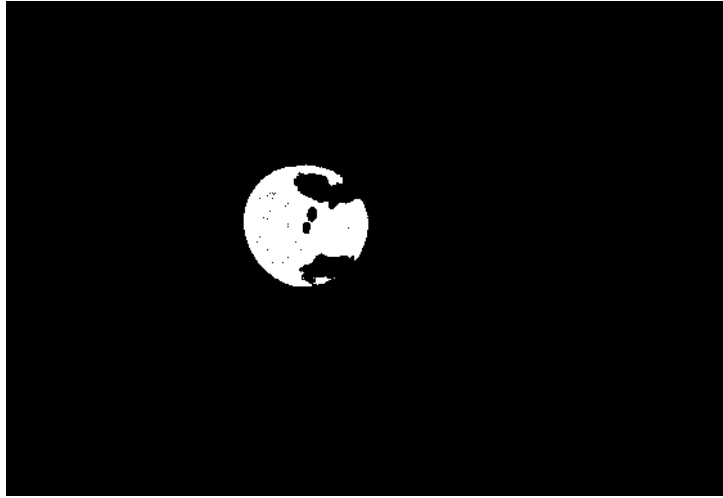


Figure 4.18: Threshold Image

After the pupil has been successfully extracted, another subroutine removes all standalone points from the edge image, the reasoning being the same as for the lower bound threshold operation as discussed above. This is simply implemented by visiting each pixel and determining which filled elements do not have any filled neighbouring elements. Artefact removal is an important part of the algorithm, since any remaining points that are not part of the pupil may cause an error in the subsequent pupil tracking subroutines. Figure 4.19 shows the threshold image after standalone points have been removed. Since the image in Figure 4.18 does not have any stand alone points, the subroutine is demonstrated on a different thresholded image that was also retrieved from a real world VOG recording.



Figure 4.19: Removal of Standalone Points: (a.) Thresholded Image Before Stand Alone Point Removal and, (b.) Image After Stand Alone Point Removal

The step that follows after the removal of standalone points, is the part of the algorithm that provides the robustness advantage over the originally proposed algorithm. Recall from Section 4.1.6 that closing is a morphological operation that can be used to fill small holes and gaps in an object body. Since the glints in an image is significantly smaller than the pupil itself, they simply cause small holes in the pupil body, and small gaps at the pupil boundary. These holes and gaps can be filled by choosing an

appropriate structuring element and performing a closing operation. Thus, the effect of the glint can be removed without relying on another adaptive threshold operation to isolate the glints. To ensure that the SE is large enough to fill all the gaps and yet small enough to minimize the computational load, the size of the SE is based on an estimated pupil size. The pupil centre is approximated first by finding the average of all the points in the threshold image. Next, the maximum and minimum x and y coordinate values are extracted from the thresholded image and used to estimate a pupil diameter. The SE is then chosen to have dimensions equal to one quarter of the estimated pupil diameter. The closing operation is then performed and the resulting image is assumed to accurately define the pupil. Figure 4.20 shows the result of closing of the threshold image. Notice that all holes inside the pupil body have been closed. The gaps caused by the glints located on the pupil boundary have been narrowed although they still remain.

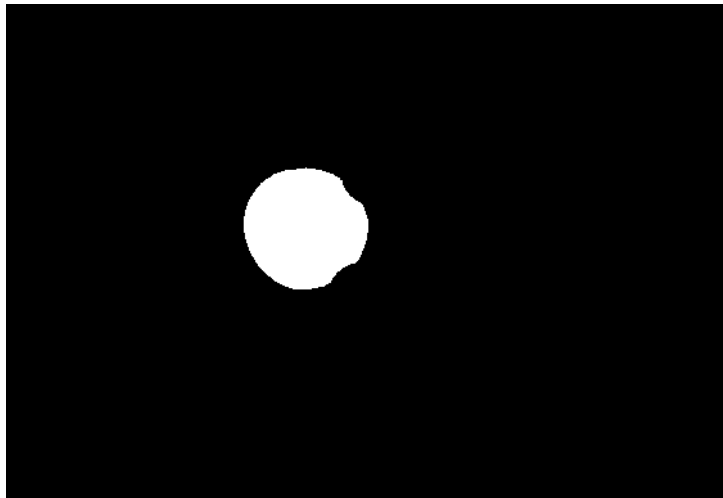


Figure 4.20: Result of Closing Operation

An edge detection algorithm is the final image processing operation before the ellipse fitting process. The edge of the reconstructed pupil is extracted by means of a Sobel edge detector which implements frequency domain filtering for computational efficiency purposes. An ellipse is then fit to the edge points with a least squares minimization method, since it is computationally cheaper than the alternative methods. The only advantage provided by the other ellipse fitting methods such as the modified Hough transform and the RANSAC algorithm, is a robust response to outlying points that are not part of the pupil boundary. Since care is taken to minimize the occurrence of such points, the least squares minimization is expected to perform just as well. Figure 4.21 shows the pupil boundary as detected by a Sobel edge detection operation from the closed image and represents the final estimation for the pupil edge.

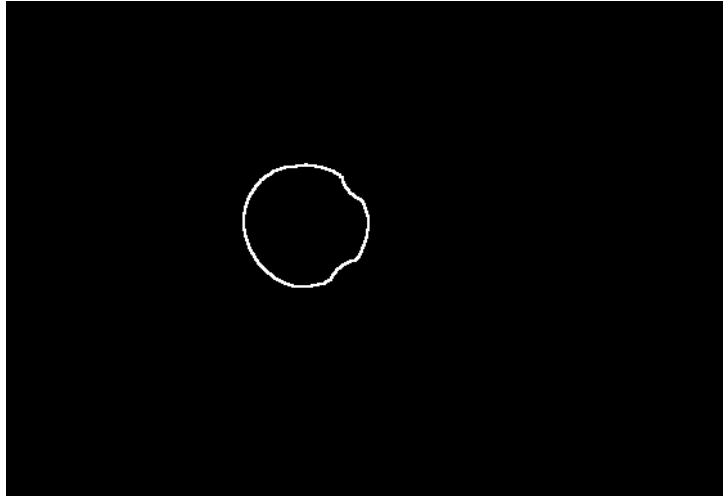


Figure 4.21: Estimated Pupil Boundary

The points that define the edge of the pupil, as shown in Figure 4.21, are extracted from the image and recorded in an array from which ellipse fitting is implemented. An ellipse can be defined in its quadratic form as shown in equation 4.3.2. This equation gives the generalized form of an ellipse in the global coordinate system, and thus includes rotation around the ellipse origin. A more detailed discussion of this generalized form is given in Section D.2.

$$a'x^2 + 2b'xy + c'y^2 + 2d'x + 2f'y + g' = 0 \quad (4.3.2)$$

Reorganizing equation 4.3.2 by dividing with a' and subtracting x^2 from each side of the equality, prepares it for implementation in a least squares minimization as shown in equation 4.3.3 (Zill and Cullen, 2000).

$$\mathbf{Y} = \mathbf{A}\mathbf{X} \quad (4.3.3)$$

The general least squares minimization problem, as given in equation 4.3.3, can be restated for the implementation of ellipse fitting as shown in equation 4.3.4.

$$\begin{bmatrix} -x_1^2 \\ -x_2^2 \\ \vdots \\ -x_n^2 \end{bmatrix} = \begin{bmatrix} 2x_1y_1 & y_1^2 & 2x_1 & 2y_1 & 1 \\ 2x_2y_2 & y_2^2 & 2x_2 & 2y_2 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 2x_ny_n & y_n^2 & 2x_n & 2y_n & 1 \end{bmatrix} \begin{bmatrix} b \\ c \\ d \\ f \\ g \end{bmatrix} \quad (4.3.4)$$

where n is the number of points extracted from the edge image and $a = 1$, $b = \frac{b'}{a'}$, $c = \frac{c'}{a'}$, etc.

The solution to the problem is found with equation 4.3.5 (Zill and Cullen, 2000). The parameters, b through g are extracted from matrix \mathbf{X} .

$$\mathbf{X} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Y} \quad (4.3.5)$$

CHAPTER 4. PUPIL TRACKING

With the extracted parameters the ellipse origin coordinate (x_0, y_0) , orientation (θ) , semimajor-axis (p) and semiminor-axis (q) values can be calculated from equations 4.3.6 through 4.3.10 available from *Wolfram Mathworld*¹.

$$x_0 = \frac{cd - bf}{b^2 - ac} \quad (4.3.6)$$

$$y_0 = \frac{af - bd}{b^2 - ac} \quad (4.3.7)$$

$$\theta = \begin{cases} 0 & \text{for } b = 0 \text{ and } a < c \\ \frac{1}{2}\pi & \text{for } b = 0 \text{ and } a > c \\ \frac{1}{2} \cot^{-1}\left(\frac{a-c}{2b}\right) & \text{for } b \neq 0 \text{ and } a < c \\ \frac{\pi}{2} + \frac{1}{2} \cot^{-1}\left(\frac{a-c}{2b}\right) & \text{for } b \neq 0 \text{ and } a > c \end{cases}$$

$$p = \sqrt{\frac{2(af^2 + cd^2 + gb^2 - 2bdf - acg)}{(b^2 - ac)[\sqrt{(a-c)^2 + 4b^2} - (a+c)]}} \quad (4.3.9)$$

$$q = \sqrt{\frac{2(af^2 + cd^2 + gb^2 - 2bdf - acg)}{(b^2 - ac)[-\sqrt{(a-c)^2 + 4b^2} - (a+c)]}} \quad (4.3.10)$$

With these parameters known, the ellipse can be drawn over the original frame to serve as a visual feedback. The centre coordinate is also stored in a data file for signal analysis purposes. This is the final step of the tracking algorithm. The result of an ellipse fit via least squares minimization to the points in the edge image shown in Figure 4.21 superimposed on the VOG frame of Figure 4.16, is shown in Figure 4.22.

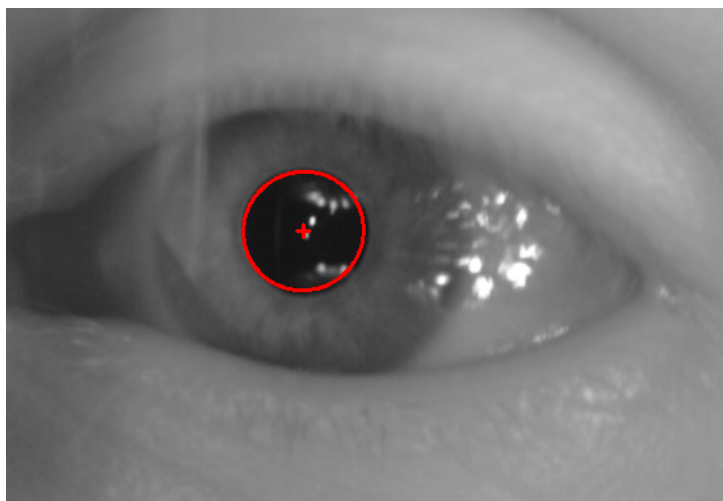


Figure 4.22: Estimated Pupil Boundary from Ellipse Fit

¹<http://mathworld.wolfram.com/Ellipse.html>

CHAPTER 4. PUPIL TRACKING

Figure 4.22 shows the results of the tracking algorithm introduced in this section. From the visual inspection from VOG video recordings, it was concluded that the algorithm performs very well, as pupil detection proved to be rather accurate with an exception to few frames that includes outlying artefacts that affects ellipse fitting. Pupil edge estimation remained accurate for lateral gaze situations where the pupil is deformed considerably, as shown in Figure 4.23. The algorithm's pupil boundary estimation also remained remarkably accurate despite instances when glints are located on the pupil boundary as previously shown in the demonstration figures for the tracking algorithm. The result of the tracking algorithm for an image when no glints are located on the boundary is shown in Figure 4.24 which demonstrates the accuracy of the ellipse fit procedure. It can thus be concluded that the developed algorithm proved to produce satisfactory results with regards to accuracy and robustness. Furthermore, the computational efficiency of the final algorithm averaged on about 2.2 frames per second.

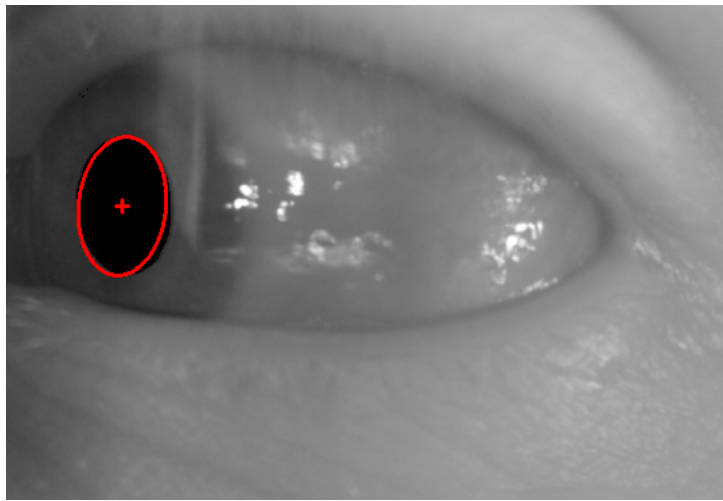


Figure 4.23: Lateral Gaze Tracking

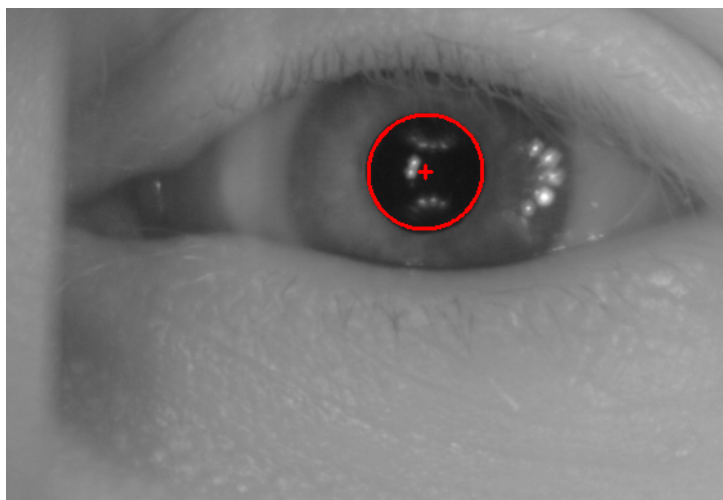


Figure 4.24: Estimated Pupil Boundary from Ellipse Fit for no Glints on Boundary

Chapter 5

Signal Processing

The algorithms discussed in Chapter 4 are purely concerned with pupil tracking. With the tracking algorithm that was presented, the pupil centre can be determined for each frame in a VOG recording and then stored in a data file. The stored pupil centre locations can then be extracted from the data file at a later time for signal analysis purposes. Signal analysis is the final step for the nystagmus identification process. This chapter will discuss the algorithms implemented to identify nystagmus beats within the signal data.

5.1 Parametric Data Analysis

Computerized analysis of ocular response data, such as nystagmus, have been in research for almost as long as computers were available. Research on computerized data analysis of eye movement recordings go back as far as the late 1960's (Houston and Watson, 1994). Despite the considerable duration of this research period, relatively little progress has been made. Until recently, all eye movement analysis algorithms implemented parametric analyses, since it provides a rather intuitive approach. However, Houston and Watson (1994) reported that parametric approaches usually require user interaction with the exception of a few unsupported claims of fully automated algorithms. Naturally, the ideal algorithm for data analysis in a clinical environment should be able to identify nystagmus completely automated such that the interaction time and knowledge requirements for the user are kept to a minimum.

Parametric approaches make use of the measurable parameters of the pupil position data such as the velocities and amplitudes of the recorded signal. This provides the opportunity to separate the quick phases and slow phases and perform subsequent constraint checks such that possible nystagmus beats can be identified. Although it is difficult to clearly define different subcategories for parametric methods, Houston and Watson (1994) identified two basic parametric methods that are implemented in existing systems. These two approaches are the directional method and the velocity method respectively. A brief discussion of the parametric analysis methods implemented for nystagmus beat identification will be presented in the following paragraphs.

Directional parametric analysis makes use of the predefined beat direction as criteria for slow phase identification. Additional criteria such as amplitude, duration and quick

phase direction and velocity can also be implemented in order to identify nystagmus beats.

Velocity-based parametric analysis of eye movement data essentially depend on velocity thresholds. Most algorithms differentiate the position signal in order to find the velocity of the eye over time. The differentiated signal is then used to identify quick and slow phases based on velocity.

Another approach is the combination of the methods discussed above. However, all these methods have drawbacks. As mentioned before, parametric methods usually require user interaction. For example, the directional method requires a beat direction as input. For velocity based algorithms it has also been argued that velocity thresholds are prone to errors since the quick phase velocity of one nystagmus beat can be lower than the slow phase velocity for another nystagmus beat located in close proximity to the first beat (Rey and Galiana, 1991). Recall from Figure 3.1 that the maximum velocity of a saccade is frequently dependent on the magnitude thereof. Thus, achieving an appropriate velocity threshold value for a robust analysis is problematic.

5.2 Wavelet Approach

In a recent paper, Jansen *et al.* (2010) reported successful implementation of nystagmus beat identification via wavelet transformations. They claimed that, to their knowledge, all previous nystagmus beat identification algorithms implemented parametric-based methods. However, Abel *et al.* (2008) implemented wavelet analysis for nystagmus identification in infants, although they have concluded that the limitations of wavelets outweighed their advantages and that parametric methods are generally more efficient. The results leading to this conclusion may be due to the fact that Abel *et al.* (2008) implemented Haar wavelets, which is not ideal for nystagmus identification. This will be discussed in more depth in Section 5.3.3.

The wavelet analysis implemented by Jansen *et al.* (2010) entails the identification of quick phases of possible nystagmus beats via wavelet decomposition with a Daubechies 2 (db2) mother wavelet. The sharp peaks of the nystagmus beats are then extracted with an appropriate threshold. Signal reconstruction is implemented by a quadratic spline interpolation, followed by a sequence of data checks to exclude irrelevant data. The algorithm evaluates the adherence of the signal to three conditions. Firstly, the direction of movement of the slow phase should be opposite to that of the surrounding quick phases. Secondly, the period of the slow phase should be longer than that of both surrounding quick phases. Finally, the velocity of both quick phases should exceed that of the slow phase.

The fully automatic wavelet method implemented by Jansen *et al.* (2010) was tested for various types of physiological nystagmus, including optokinetic nystagmus and acceleration induced ocular vestibular nystagmus waveforms. The study showed a significant improvement in specificity, although sensitivity results were less successful compared to parametric methods in some cases. Sensitivity and specificity as referred to here, are the statistical measurements of performance for the the analysis algorithm and are defined in equations 5.2.1 and 5.2.2 respectively.

$$\text{sensitivity} = \frac{\text{number of true positives}}{\text{number of true positives} + \text{number of false negatives}} \quad (5.2.1)$$

$$\text{specificity} = \frac{\text{number of true negatives}}{\text{number of true negatives} + \text{number of false positives}} \quad (5.2.2)$$

5.3 Proposed Signal Analysis Algorithm

In this section, the proposed algorithm for this thesis that was implemented to identify nystagmus beats within a signal will be discussed. The process consists of several distinct steps that will be discussed independently. Recall from Section 2.1.2 that there are several types of nystagmus and that nystagmus can present itself in one of two waveforms. Pendular nystagmus is predominantly associated with congenital sensory defects and usually have a slower oscillation. Since pendular nystagmus is rather simple to identify visually and provides little further information, this thesis is concerned with the identification of jerk type waveforms. Jerk type waveforms are associated with diagnostic tests for optokinetic and ocular-vestibular responses, since physiological nystagmus presents itself in jerk type waveforms. These types of tests where physiological nystagmus is induced to determine whether the patient has a normal response are typical clinical tests. Clinical tests that measure the physiological nystagmus responses in a patient can be used to diagnose various neurological or pathological conditions, as was discussed in Section 2.1.2. The caloric response can also provide information about balance issues. Jerk type waveforms thus present a wide field of diagnostic possibilities and will be the focus in this thesis.

5.3.1 Blink Removal

Preprocessing of the eye motion signal include the identification and removal or damping of anomalies in the signal. Sudden, large changes of the pupil position, due to the subject blinking, is an example of such an anomaly. Whenever the subject blinks and the pupil is covered by the eyelid, the pupil is either estimated to be at a shaded region of the skin or the position is recorded at the origin of the image. Since the eye is well lit, the last happens most frequently. Whatever the case may be, signal changes due to blinks need to be identified.

Since the blinks are represented by data far from the signal median, a threshold based on standard deviation of the signal is implemented. The standard deviation is calculated from the statistical relationship between the mean absolute deviation (MAD) and standard deviation as shown in equation 5.3.1, where x is the signal. Identification of the blinks is implemented by a $2\hat{\sigma}$ threshold. Rather than removing the data completely and causing discontinuities in the signal or reconstruct false eye motions, all outlying coordinates are relocated to the $2\hat{\sigma}$ threshold. A constant magnitude signal component is thus created on the threshold boundary. This does not remove the blink from the data set, but ensures that the blink data do not appear to represent a nystagmus beat. The threshold also minimize the effect of the blink on the analysis of surrounding components by reducing the deviation of the blink data from the mean. An example of the blink thresholds can be clearly observed in Figure 5.13.

$$\hat{\sigma} = \frac{MAD(x)}{0.6745} \quad (5.3.1)$$

5.3.2 Wavelet Theory

In order to compare signal content to a set of predefined general constraints and identify pieces of the signal that may be of interest, the possible locations of nystagmus beats have to be determined first. These locations can be found by identifying the sudden change in direction and velocity which is inherent to the quick phase portion of the nystagmus waveform. Wavelet analysis provides a reliable method for identification of these local minima and maxima. Therefore, since the signal processing algorithm for nystagmus beat identification in this thesis is based on wavelet transformations, a brief review on wavelet theory is presented here. The theory discussed in this section have been referenced from Gonzales and Woods (2008). Wavelet expansions are the representations of a signal with scaled and translated copies of a reference waveform period referred to as a mother wavelet. The major advantage of a wavelet transform over the more conventional Fourier transform is the efficiency in the detection of discontinuities and sharp peaks. A wavelet series expansion, as shown in equation 5.3.2, expresses a function in terms of a scaling function, $\varphi(x)$ and a wavelet function, $\psi(x)$ known as the mother wavelet.

$$f(x) = \sum_k a_{j_0}(k) \varphi_{j_0,k}(x) + \sum_{j=j_0}^{\infty} \sum_k d_j(k) \psi_{j,k}(x) \quad (5.3.2)$$

For the equation above, j specifies the scale and k specifies the translation of the wavelet. The terms $a_{j_0}(k)$ and $d_j(k)$ denote the scaling coefficients and wavelet coefficients respectively. These terms are also frequently referred to as the approximation coefficients and detail coefficients respectively and their calculation, as shown in equations 5.3.3 and 5.3.4, is defined as the forward wavelet transform.

$$a_{j_0}(k) = \int f(x) \varphi_{j_0,k}(x) dx \quad (5.3.3)$$

$$d_j(k) = \int f(x) \psi_{j,k}(x) dx \quad (5.3.4)$$

As with Fourier transformations, wavelet transforms can also be implemented digitally by means of the Discrete Wavelet Transform (DWT). The discrete wavelet series and the associated expansion coefficients for the DWT are given in equations 5.3.5 through 5.3.7.

$$f(n) = \frac{1}{\sqrt{M}} \sum_k W_{\varphi}(j_0, k) \varphi_{j_0,k}(n) + \frac{1}{\sqrt{M}} \sum_{j=j_0}^{\infty} \sum_k W_{\psi}(j, k) \psi_{j,k}(n) \quad (5.3.5)$$

where,

$$W_{\varphi}(j_0, k) = \frac{1}{\sqrt{M}} \sum_n f(n) \varphi_{j_0,k}(n) \quad (5.3.6)$$

$$W_\psi(j, k) = \frac{1}{\sqrt{M}} \sum_n f(n) \psi_{j,k}(n) \quad \text{for } j \geq j_0 \quad (5.3.7)$$

The expansion coefficients of equations 5.3.6 and 5.3.7 are defined for a discrete wavelet series expansion with M equally spaced samples. Thus, these two equations define the forward discrete wavelet transform. Fast implementation of the DWT is facilitated by the relationship between DFT coefficients of neighbouring scales. This Fast Wavelet Transform (FWT) is implemented with a sub-band coding filter bank, generally known as a Mallat tree. Figure 5.1 shows an example of a Mallat filter bank for a three level wavelet expansion.

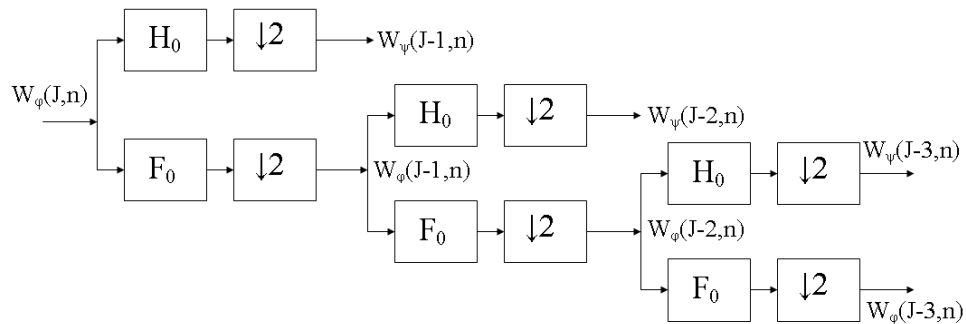


Figure 5.1: Mallat Tree with Three Level Wavelet Expansion

The symbol J in Figure 5.1 indicates the highest scale of the wavelet expansion. The number of samples M that represents f is usually chosen such that $M = 2^J$. F_0 and H_0 , respectively represents a low-pass filter associated with the scaling vector, h_φ and a high-pass filter associated with the wavelet vector, h_ψ . Equations 5.3.8 and 5.3.9 show the definition of the scaling function and wavelet function in terms of the scaling vector and wavelet vector. The $\downarrow 2$ symbol block in Figure 5.1 indicates a down-sample operation to half the number of samples of the previous scale. This is inherent to the FWT.

$$W_\varphi(j, k) = \sum_m h_\varphi(m - 2k) W_\varphi(j + 1, m) \quad (5.3.8)$$

$$W_\psi(j, k) = \sum_m h_\psi(m - 2k) W_\psi(j + 1, m) \quad (5.3.9)$$

A lower scale's approximation coefficients and detail coefficients are determined by two independent filter operations on the approximation coefficients of one scale above the said scale. The digital low-pass and high-pass filter operations are implemented by convolution with $h_\varphi(-n)$ and $h_\psi(-n)$ respectively. Equations 5.3.10 and 5.3.11 give the mathematical definition of the filter operations.

$$W_\varphi(j, k) = h_\varphi(-n) \star W_\varphi(j + 1, n) \quad |_{n=2k, k \geq 0} \quad (5.3.10)$$

$$W_\psi(j, k) = h_\psi(-n) \star W_\psi(j + 1, n) \quad |_{n=2k, k \geq 0} \quad (5.3.11)$$

Note from Figure 5.1 that at the highest scale of the wavelet decomposition, the approximation coefficients are assumed to be the sampled function itself, i.e. $W_\varphi(J, n) = f(n)$. Thus, higher scales contain more detailed information on the signal and the detail coefficients represents high frequency content. As the scale decreases, the content of the expansion becomes more coarse and lower frequencies are more prominent.

5.3.3 Quick Phase Detection

Recall from Section 5.2 that Abel *et al.* (2008) used Haar wavelets and Jansen *et al.* (2010) made use of Daubechies 2 wavelets to identify nystagmus beats. After a careful inspection of the performance of various standard wavelets with regards to nystagmus beat identification, it was concluded that the Daubechies 2 (db2) mother wavelet provides the most efficient analysis. Figure 5.2 shows the db2 mother wavelet and the accompanying scaling function. The important features of the db2 wavelet that make it an appropriate choice for quick phase detection are the sharp local extrema which are separated along the sample axis. This corresponds to the sharp local extrema at each end of the quick phases that are slanted along the sample axis. This is why the approach by Abel *et al.* (2008) with a Haar wavelet, which is simply an unit square wave function, is not ideal for nystagmus feature detection. Figure 5.3 shows the Haar mother wavelet and scaling function. Notice that the extrema are neither sharp nor spaced apart as is the case for the db2 wavelet and therefore the Haar wavelet does not correspond to a nystagmus beat. Therefore, the db2 wavelet was used for quick phase identification in this thesis.

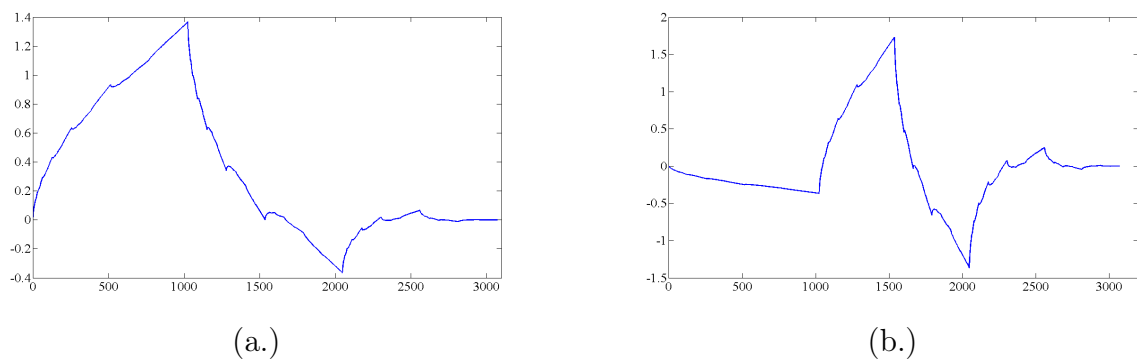


Figure 5.2: Daubechies 2 Wavelet: (a.) Scaling Function, φ and (b.) Wavelet Function, ψ

CHAPTER 5. SIGNAL PROCESSING

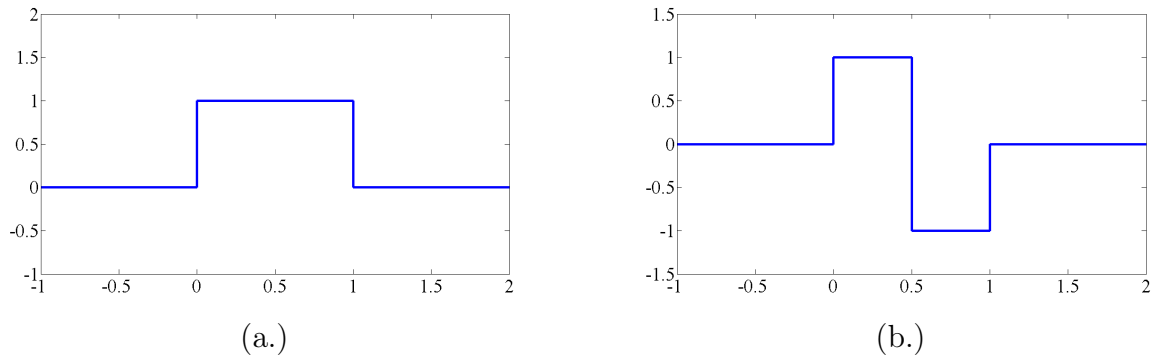


Figure 5.3: Haar Wavelet: (a.) Scaling Function, φ and (b.) Wavelet Function, ψ

Consider the artificially created nystagmus signal with two beats in Figure 5.4. The wavelet decomposition for the function is shown in Figure 5.5. Note that d_1 and d_5 are the detail coefficients of the finest and coarsest scales respectively. This might seem in contrast with the theory discussed in Section 5.3.2, but this numbering scheme is the convention in wavelet analysis software. For detail coefficients denoted as d_n , the subscript n denotes the decomposition level, which is an indication of how many times the signal has been filtered. Recall from Figure 5.1 that the first level of the filter is at the finest scale and thus the highest scale. Similarly, the last level is at the lowest scale.

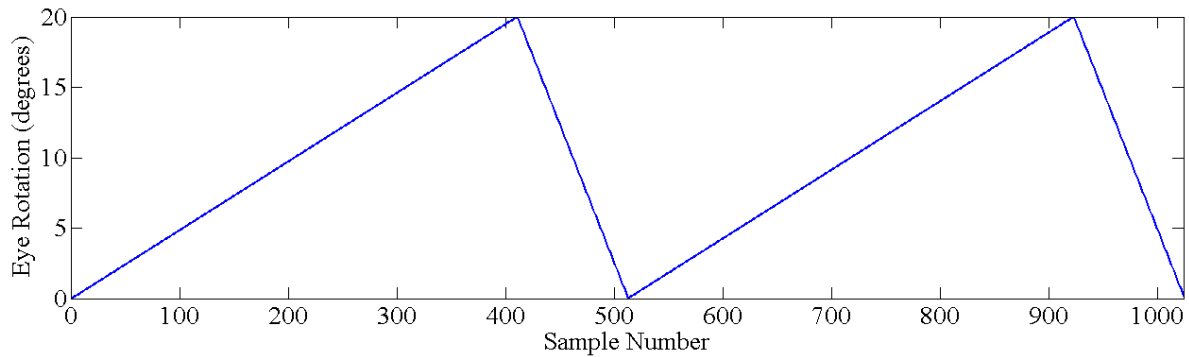


Figure 5.4: Demonstration Nystagmus Signal

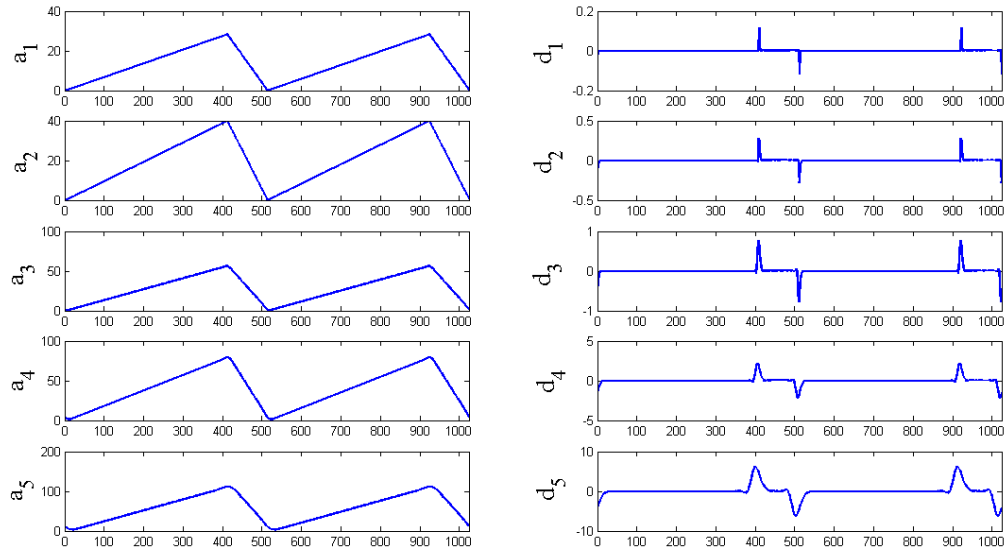


Figure 5.5: Five Level Wavelet Decomposition Coefficients

To ensure successful detection of the beats, it is preferred not to lose any resolution at the coarse, lower scales. Therefore, the down-sampling operation at each level of decomposition is not ideal. Fortunately, the Stationary Wavelet Transform (SWT) provides a wavelet decomposition that omits the down-sampling step by padding the filters with zeros (Nason and Silverman, 1995). Recall from Section 5.3.2 that, for the DWT, the number of samples, M , of the signal should be selected to be of the form $M = 2^J$, where J is the highest scale. Therefore, before the signal is fed to the analysis algorithm, the signal is down-sampled such that the number of samples adhere to the form given above. To ensure that as much samples as possible are still achieved, the maximum possible scale for the signal is determined by incrementing J from 1 until M is larger than the number of samples of the original discrete function. Then, J is decremented to find the maximum scale. The original signal is then down-sampled to 2^J samples with a FFT interpolation. This is the first step for the signal analysis after an initial smoothing filter that minimizes the pixelation effect.

After the correct number of samples and maximum scale have been determined, as was discussed in the previous paragraph, the SWT is performed on the signal. The next objective is to extract the quick phases by identifying the locations of the sharp peaked local minima and maxima in the signal. Figure 5.6 shows the response of the fine scale detail coefficients to the local sharp peaked minima and maxima of the nystagmus beats. This property presents the opportunity to extract the local extrema from the lower level detail coefficients.

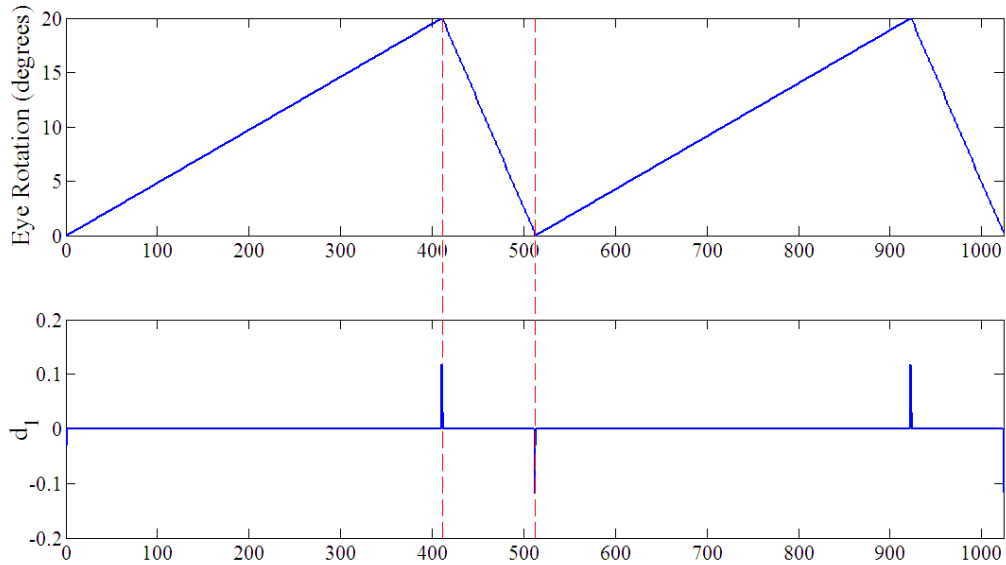


Figure 5.6: Detail Coefficients at Local Extrema

The locations of the local extrema of the signal in Figure 5.6 can be determined by simply finding the non-zero first level detail coefficients. However, for non-ideal signals, noise has to be accounted for. Since the lower levels contain more detailed information, noise components are captured by the detailed coefficients. Fortunately, this also presents the opportunity to estimate the noise which is needed to find the appropriate threshold value. Donoho (1995) proposed the noise reduction threshold for wavelet analysis given in equation 5.3.12. Here, N is the number of samples for the signal and σ is the estimated noise level. The scaling constant, γ_1 can be estimated mathematically (Donoho, 1995), but was found iteratively for robust application in this thesis.

$$\lambda = \gamma_1 \sigma \sqrt{\frac{2 \ln(N)}{N}} \quad (5.3.12)$$

Previous authors implemented the universal threshold shown in equation 5.3.12 in biomedical signal processing that makes use of wavelets, for example, nystagmus quick phase identification (Jansen *et al.*, 2010) and detection of abnormalities in heart sounds (Wang *et al.*, 2010). In order to implement this universal threshold, the noise level has to be estimated first. This can be done by using the statistical relationship between standard deviation and the Mean Absolute Deviation (MAD) of the first level detail coefficients as shown in equation 5.3.13. This statement is based on the rational assumption that the finer scale detail coefficients that contain the highest frequency components of the signal are predominantly influenced by the noise in the signal.

$$\hat{\sigma} = \frac{MAD(d_1)}{0.6745} \quad (5.3.13)$$

With the noise estimated as shown in equation 5.3.13, the signal specific threshold can be determined and implemented to extract the possible extrema of the signal from the

wavelet decomposition. Since the first level detail coefficients are assumed to be noise for threshold determination purposes, the threshold is performed on the second level detail coefficients. Consider the signal in Figure 5.7.a which is similar to the signal in Figure 5.4, but with added Gaussian white noise. The noise that was used in this example was created around a zero mean with a variance of 0.0025 and then added to the ideal signal shown in Figure 5.5. The variance was chosen such that the ratio of average noise to average extrema response in the first level detail coefficients was representative of that of a real VOG signal. Figures 5.7.b and 5.7.c shows the detail coefficients for the first two wavelet decomposition levels. Figure 5.8 shows the same information as shown in Figure 5.7, but for a real world VOG signal.

The threshold closest to the mean, indicated with red dashed lines, is the proposed threshold by Donoho (1995) with unity scaling factor. The second set of dashed lines, furthest away from the mean, is the threshold proposed by Jansen *et al.* (2010) defined as $\lambda = \sqrt{2 \ln N}$. Thus, a scaling constant of \sqrt{N} is used. Implementation of this threshold on real world VOG signals has proved to be too strict. Therefore, a modified version of the threshold is proposed here. A scaling constant of $\sqrt{N/2}$ is proposed for this thesis. The green dashed lines in Figure 5.7.c indicate the chosen threshold value.

The threshold value as illustrated in Figure 5.7.c, is chosen such that only relevant data points are extracted from the detail coefficients. However, it should also be conservative enough such that all possible locations of local extrema are found, which results in several points being extracted around each local extrema. Since only one of these coordinates represent the true extrema location, a simple three-point evaluation routine is implemented that discards all non-extrema samples from the original set. The end result of the extrema location algorithm is shown in Figure 5.9.

Notice from Figure 5.8.c that the threshold value proposed by Jansen *et al.* (2010) will not succeed in extracting the local extrema for certain nystagmus beats since the threshold value is higher than some of the responses in the second level detail coefficients. For the threshold value proposed in this thesis however, all the extrema responses are included by the the threshold while noise is entirely ommited.

CHAPTER 5. SIGNAL PROCESSING

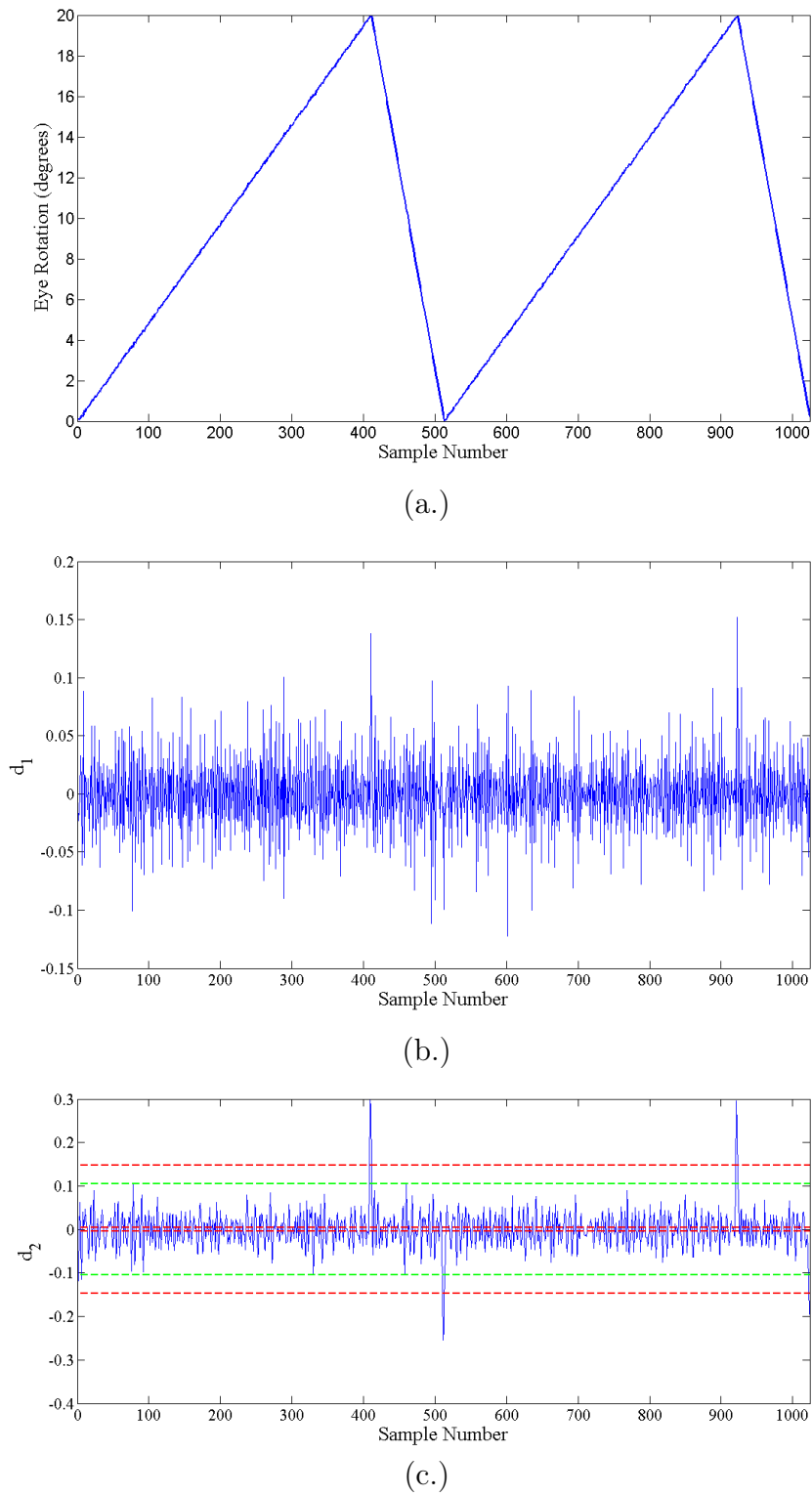
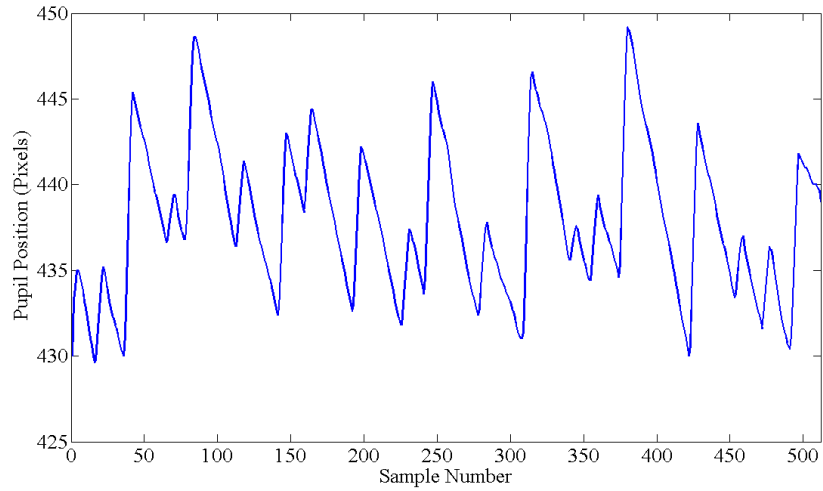
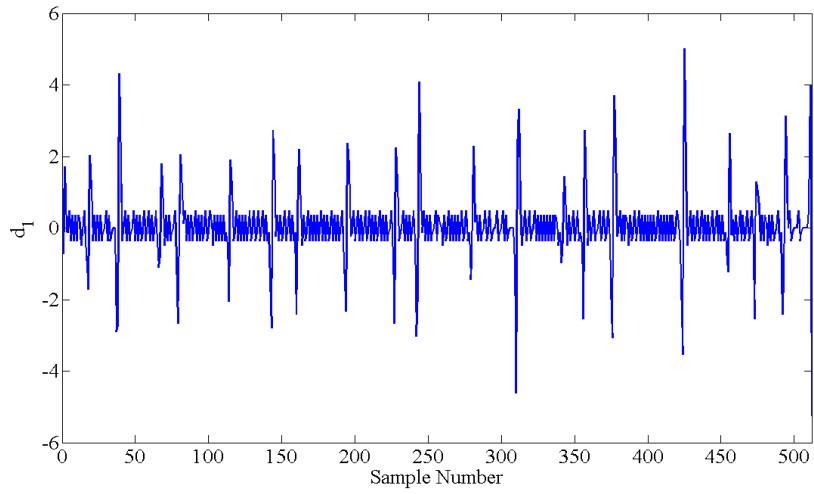


Figure 5.7: Noise Reduction by Threshold: (a.) Noisy Signal, (b.) First Level Detail Coefficients, and (c.) Second Level Detail Coefficients with Threshold Values

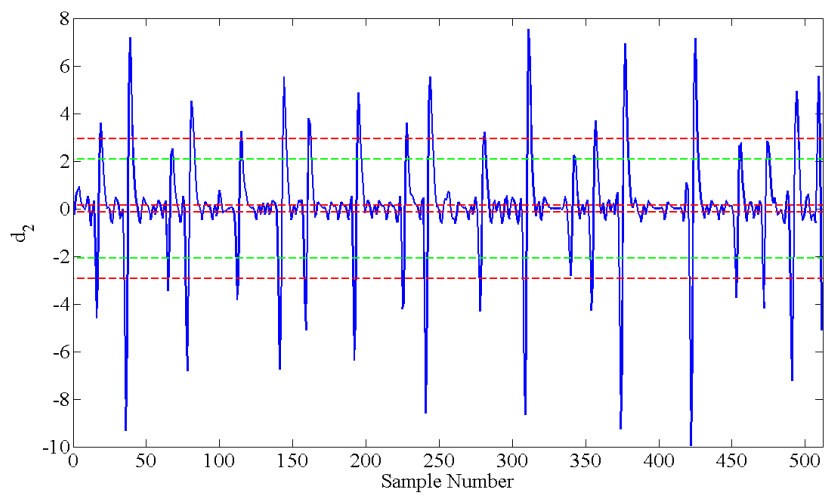
CHAPTER 5. SIGNAL PROCESSING



(a.)



(b.)



(c.)

Figure 5.8: Noise Reduction on Real VOG Signal: (a.) VOG Signal, (b.) First Level Detail Coefficients, and (c.) Second Level Detail Coefficients with Threshold Values

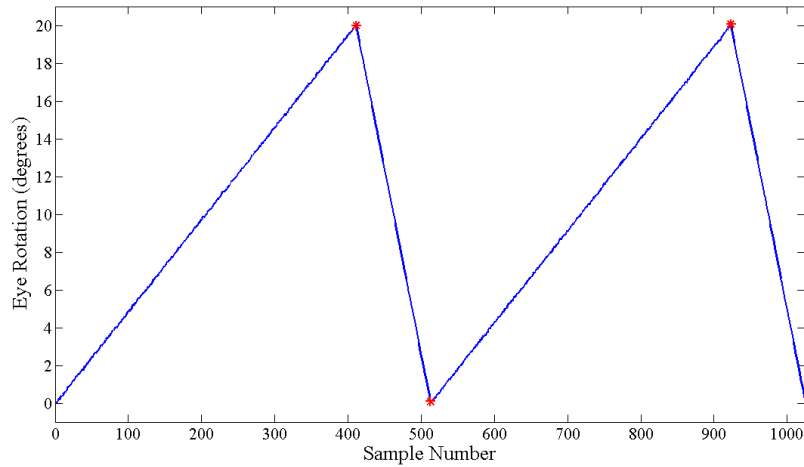


Figure 5.9: Extrema from Detail Coefficients

This concludes the quick phase identification through wavelet analysis and the subsequent extraction of the local minima and maxima associated with the quick phases in the signal. Discussions in the rest of this section will relate to nystagmus beat identification by parametric comparisons to predefined constraints.

5.3.4 Signal Reconstruction

With the extrema extracted from the signal, reconstruction can commence. Since both phases of the jerk for physiological nystagmus usually have constant velocities, the signal is reconstructed with straight lines between adjacent coordinates of the acquired local extrema. This is achieved through a simple linear interpolation routine. After the signal has been reconstructed, accuracy is estimated by finding the error relative to the signal span. Figure 5.10 shows the reconstruction of signal between the identified local extrema from a real world signal's original samples. The encircled sample points are the samples that were identified as the local extrema of the signal. The solid lines and the crosses give the reconstructed signal and the original samples, respectively.

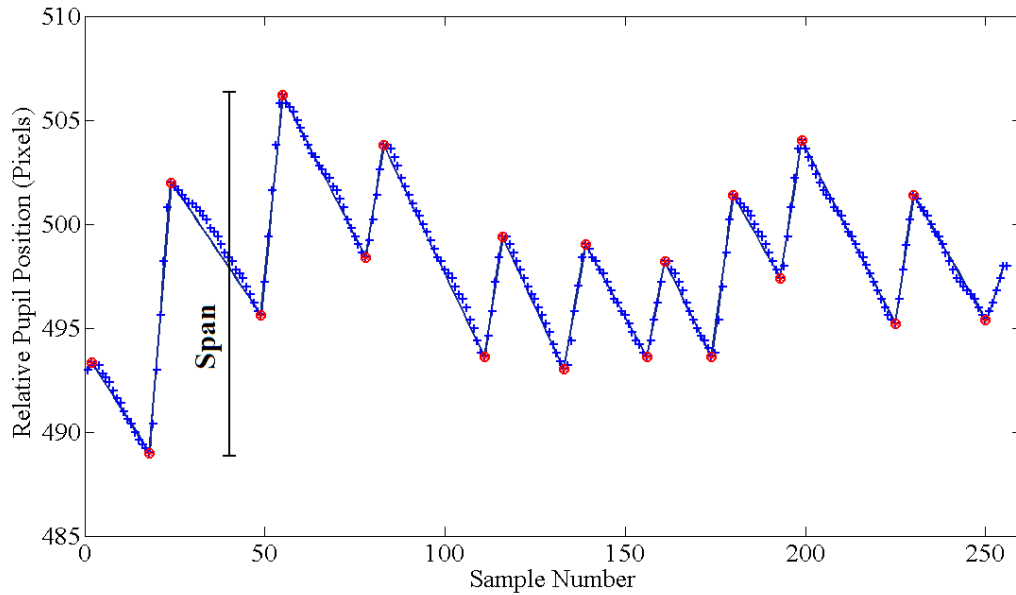


Figure 5.10: Reconstructed Signal

The error percentage is defined as one hundred times the sum of the magnitude differences, which are the vertical distances between the original data points and the reconstructed signal shown in Figure 5.10, divided by the signal span which is defined as the difference between the maximum and minimum values in the signal as indicated in Figure 5.10.

Before any parametric constraints are considered, the algorithm determines whether the reconstructed signal has a reasonable accuracy compared to the original signal content. For the real world data analysis tests discussed in Section 5.4.2, the lower reconstruction accuracy limit was set to 90%. If the reconstructed signal is within acceptable accuracy limits, it is assumed that the signal can be sufficiently reconstructed using straight lines and thus that the algorithm can proceed with parametric analysis.

5.3.5 Parametric Constraints

In order to make a positive identification of a nystagmus beat from the acquired extrema of a signal, the signal is evaluated for three apparent phases at a time and tested against some predefined geometrical constraints. These constraints are chosen to represent the best possible definition of a nystagmus beat. The three phases that are being analysed are found by cycling through four extrema at a time and incrementing one extrema position for every new evaluation. For each evaluation, the centre phase is assumed to be a slow phase. Before any constraint adherence evaluations are performed, a few region specific parameters are calculated from the extrema. These parameters include the magnitude, velocity and period of each apparent phase. Figure 5.11 shows the definition of the magnitudes (m) and periods (T) of the phases between the numbered local extrema that are used to construct the three phases. The velocity of each phase is defined as $v = \frac{m}{T}$.

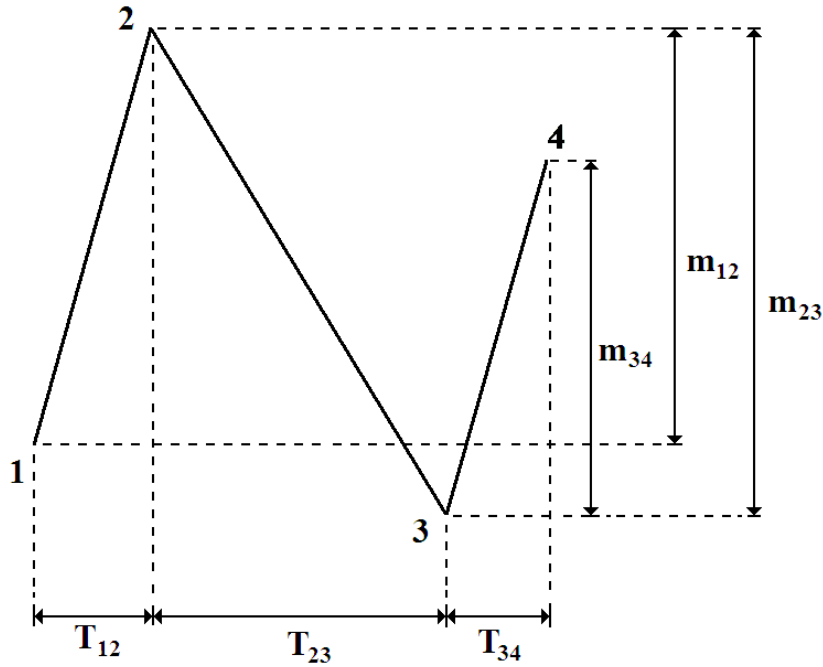


Figure 5.11: Parametric Definitions

The first constraint check that is made, is whether the direction of movement of the slow phase is in the opposite direction than each quick phase on either side of the slow phase. This is implemented by multiplying the slow phase velocity with each quick phase velocity and confirming that both results are negative. If either one of the products results in a positive value, the combination of extrema is excluded as a possible beat.

The second constraint states that the slow phase should have a period that is longer than that of either the surrounding quick phases. This constraint stems from the assumption that the magnitude of all three phases are more or less the same. This is not necessarily true, but it provides a better immunity against false negatives. Adherence to the constraint is tested by simply comparing the pre-calculated phase periods.

Perhaps the most important constraint is that the slow phase should have a velocity that is lower than that of both the surrounding quick phases. This is also a straight forward implementation, since the velocities were previously calculated.

One last constraint states that the phases should not differ too greatly. This prevents random constant velocity components to be included into adjacent nystagmus beats. When the random component is analysed as the slow phase, the period constraint may not exclude the random component. The check for constraint adherents is implemented by calculating the ratios between the first two phases and last two phases, respectively, and comparing them to predefined limits.

The four constraints discussed above in combination with the signal reconstruction accuracy discussed in Section 5.3.4, provide the filtering method that separates the possible nystagmus beats from the rest of the signal. If a given signal region between the four extrema that defines the reconstructed signal satisfies all the above constraints, the signal elements are stored in an array so that it can be superimposed on the original

signal after signal analysis is complete. This is done in order to provide a visual feedback for the analysis.

5.3.6 Windowed Analysis

The parametric and accuracy constraints coupled with other preprocessing subroutines, such as adaptive thresholding for blink removal, is an adequate method for signals with content that present little magnitude variance over the span of the signal. However, since the data is prone to contain random data components that are not nystagmus related, this is not always the case. Magnitude variation decreases the effectiveness of quick phase removal from the detail coefficients and consequently influences the sensitivity of the analysis for the worst. In order to minimize the effect of magnitude variation on the sensitivity, a windowed analysis scheme is applied. This simply entails that the signal is analysed in smaller subdivisions. The window increment is significantly smaller than the window width such that each specific region can be compared with its surrounding data. Care should be taken when defining the window width, since too few samples for the wavelet analysis can also have a significant negative effect.

5.3.7 Conclusion on Proposed Algorithm

Sections 5.3.3 through 5.3.6 discussed the proposed signal analysis algorithm for nystagmus beat identification. Although these sections focussed on the theory related to the implemented subroutines, it should be noted that the constraints and parameter values were based on a continuous development process that was built on real world signal analysis. This continuous development approach and the use of genuine nystagmus signal were especially important for the modifications that provided the robust analysis capability.

5.4 Test Method and Results

Previous sections discussed the various methods and subroutines used to identify possible nystagmus beats from pupil position data. This section will review the method that provides such data, as well as the results obtained from the analysed signals. All these signals are composed from coordinates along a single dimension, since the vertical and horizontal movements of the eye can be separated and analysed separately. This is true for all practical applications.

5.4.1 Data Collection Method

In order to test the capability of the analysis algorithm, a nystagmus signal is needed and thus real nystagmus eye motions are required. Recall from Section 2.1.2.4 that physiological nystagmus can be evoked as a normal response. This presents the opportunity to record nystagmus under a controlled test condition by some manner of physiological stimulus. There are several types of physiological nystagmus that can be

achieved, each by a unique stimulus. The first is optokinetic nystagmus (OKN) that can be induced by an optokinetic stimulation such as an optokinetic drum. This is a common ophthalmology tool used to test for normal motor cortex behaviour. The second type of physiological nystagmus that can be achieved through stimulus, is ocular-vestibular nystagmus via a caloric test which entails controlled flow of hot and cold water or air into the ear. Another type of physiological nystagmus that can be evoked, is motion induced ocular-vestibular nystagmus. This can be achieved by letting the subject sit in a rotating chair that can provide a velocity step rotation.

Motion induced ocular-vestibular nystagmus would necessitate the development of a controllable rotating chair which is outside of the scope for this thesis. It would also present some difficulty with regards to the VOG device that needs to be connected to a computer for recordings. The caloric test is too invasive for data collection purposes and is therefore also not considered. Optokinetic stimulation provides the most feasible and least invasive method to evoke nystagmus. Inspired by the optokinetic drum used in ophthalmology, a stimulation application has been developed that moves a pattern of vertical black and white stripes over the monitor of a computer. This mimics the effect of vertical stripes on a rotating cylinder, which is basically the principle of a optokinetic drum.

The application described above provokes a pure nystagmus eye motion. In order to introduce random motion components into the signal, another stimulation application was developed that alternates between optokinetic stimulation and random motion stimulation. The random motion was simply achieved by moving a small circle around the monitor at random. The subjects were instructed to follow the circle with their eyes and to keep a steady gaze when the optokinetic stimulation appeared. Thus, the subject does not fixate on a single stripe and follow it along the movement direction. The alternating pattern with the high contrast stripes causes a natural nystagmus eye motion. Each type of stimulus displayed for five seconds at a time.

Subjects were positioned at approximately 0.5 m from the monitor. The stimulation application alternated between random- and optokinetic stimulation at intervals of five seconds. Recordings were limited to 30 seconds and less. All signals were recorded at an image resolution of 640×480 with almost full field of view focus on the eye. The sampling rate was set to 75 Hz, which is the maximum achievable sampling rate at the aforementioned resolution.

5.4.2 Test Results

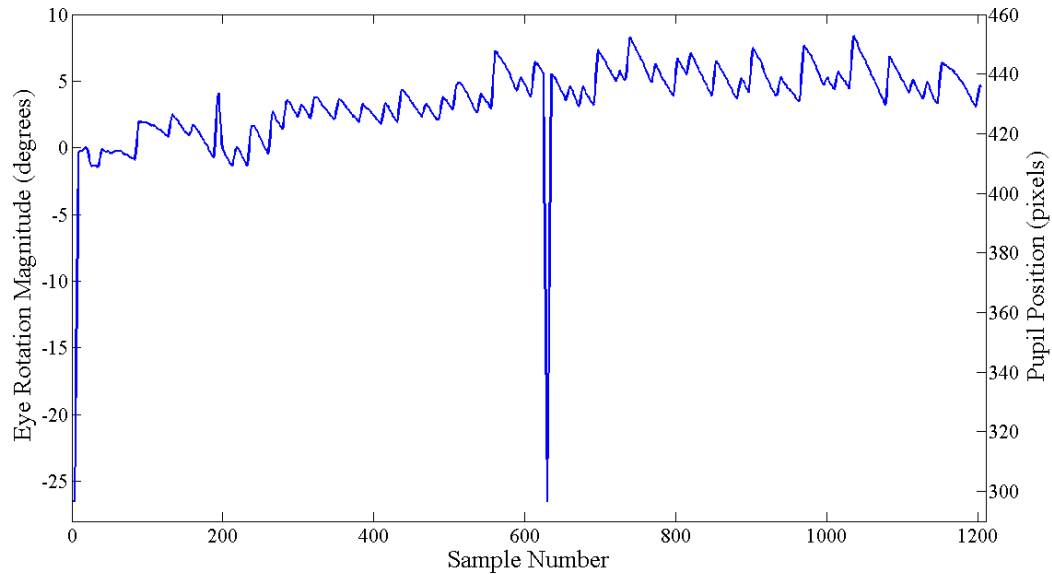
Results of the signal analyses algorithm, as discussed in Section 5.3, are presented here. The signals that were analysed, were retrieved with the method described in the previous section. Data was collected from six individuals. Two types of signals are reviewed, namely signals that were in response to pure optokinetic stimulation and signals that were the result of mixed component stimulation. All signals were analysed with the same parameters such that the robustness of the automated analysis could be proven. The reconstruction accuracy constraint was set to 90% and the window width and increment were selected to be 5 seconds and 1 second, respectively. Over 500 beats were analysed for the pure nystagmus signals and approximately 350 seconds of randomly generated data mixed with nystagmus were analysed.

Although the general trend followed in the literature is analysis of pure nystagmus signals, random signals are also reviewed here to reinforce the specificity results of the proposed algorithm. It may be argued that the algorithm implemented by Rey and Galiana (1991) does not analyse pure nystagmus signals, however they implement a model based analysis with head position as input. Thus, this analysis only provides results for strictly controlled test environments and cannot diagnose nystagmus from signals with unknown content. Jansen *et al.* (2010) presented specificity results, but there is no mention of any non-nystagmus data present in their test data. On the contrary the description of their test procedures suggest that all signals had pure nystagmus content. This lead the author to question the legitimacy of the specificity values presented by Jansen *et al.* (2010). It should be noted, however, that one of the tests conducted by Jansen *et al.* (2010) is a torsional swing test which causes a pure nystagmus signal, but with a sinusoidal shaped median. Jansen *et al.* (2010) analysed data for four different types of physiological nystagmus. Approximately 500 beats per type of nystagmus were analysed and the performance of the algorithm varied as shown in Table 5.1.

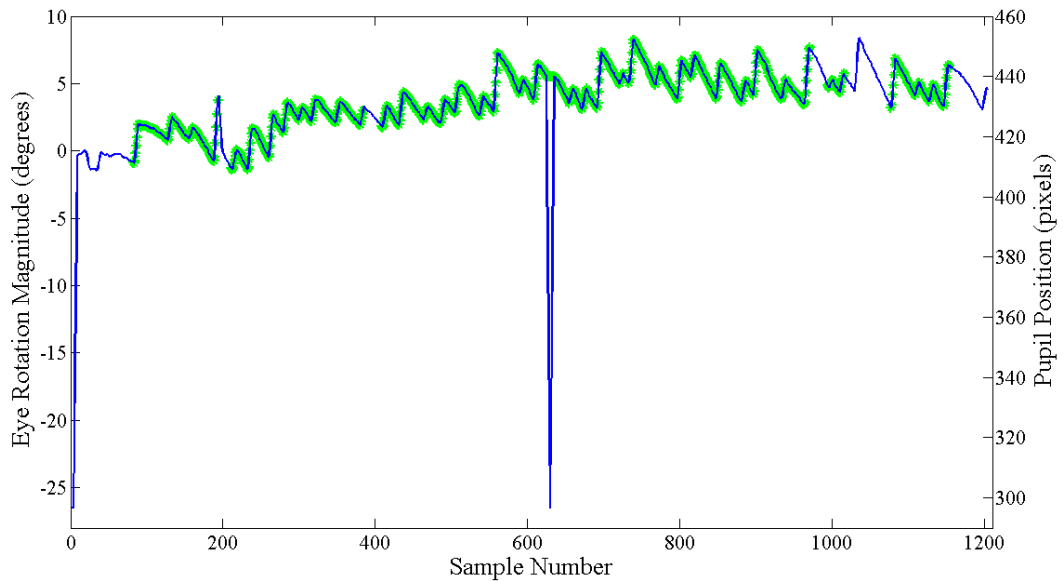
Table 5.1: Statistical Results achieved by Jansen *et al.* (2010)

Nystagmus Type	Sensitivity	Specificity
OKN	98.91%	96.74%
Torsion Swing	92.77%	94.19%
Velocity Step	93.50%	93.44%
Caloric Test	95.74%	95.65%

Figure 5.12.a shows an OKN signal as response to a pure optokinetic stimulation. Notice that almost all signal components belong to a nystagmus beat. Exceptions include the transient content at the start of the stimulation and the sudden and considerable dip in the data which is caused by the subject blinking. Figure 5.12.b shows the result of the data analysis algorithm. The positively identified nystagmus beats are highlighted with green.



(a.)



(b.)

Figure 5.12: Nystagmus Diagnosis: (a.) Original Signal, and (b.) Processed Signal

Apart from the non-nystagmus components introduced into the signal by the transient period and the blink, Figure 5.12 only demonstrates the identification of nystagmus beats in an environment where all eye motions are part of a nystagmus beat. Two common statistical values to measure the performance of a diagnostic algorithm, is sensitivity and specificity. Sensitivity and specificity were defined in Section 5.2, but are stated again in this section for convenience in equations 5.4.1 and 5.4.2.

$$\text{sensitivity} = \frac{\text{number of true positives}}{\text{number of true positives} + \text{number of false negatives}} \quad (5.4.1)$$

$$\text{specificity} = \frac{\text{number of true negatives}}{\text{number of true negatives} + \text{number of false positives}} \quad (5.4.2)$$

For the analysis of pure nystagmus signals, as shown in Figure 5.12, the parameters of sensitivity and specificity as was calculated for this thesis are defined below. Please refer to Appendix E for a graphical illustration of these definitions.

Number of true positives: The number of real phases, quick and slow, that were identified by the algorithm.

Number of false positives: The sum of all line segments that do not belong to a nystagmus beat, but were identified by the algorithm.

Number of true negatives: The sum of all line segments that do not belong to a nystagmus beat and were also not identified by the algorithm.

Number of false negatives: All real quick phases and slow phases that were not identified by the algorithm.

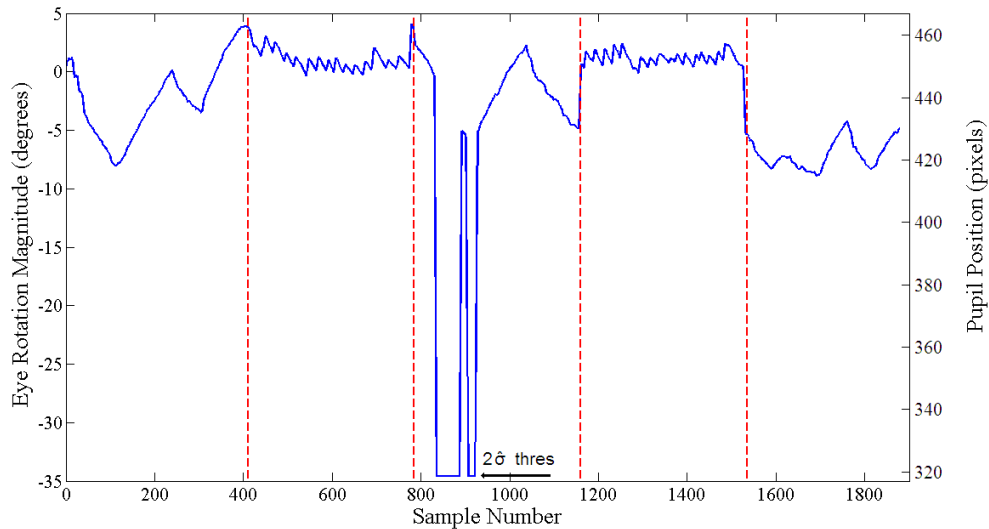
Using the above definitions for sensitivity and specificity, the performance of the analysis algorithm for pure nystagmus signals were calculated. The recorded VOG signals are given in in Appendix E. The algorithm achieved a sensitivity of 91.84% and a specificity of 96.47%. Since the same amount of beats were analysed as by Jansen *et al.* (2010), the data should be fairly comparable. Comparing the results found in this thesis with the OKN results shown in Table 5.1, it is clear that the sensitivity found by the proposed algorithm is not as high as the sensitivity claimed by Jansen *et al.* (2010). However, compared to the rest of the results in Table 5.1, the sensitivity of the proposed algorithm performs well, and the specificity even outperforms the tabulated results in some cases. Furthermore, the strict constraints implemented in the proposed algorithm also cause more true nystagmus beats to be excluded as possible beats by the algorithm. This is a necessary trade-off in order to avoid inclusion of non-nystagmus data, such that the algorithm can perform robustly for all signals and keep the specificity in an acceptable range.

Table 5.2 gives the tabulated parameters that were found by signal analysis of various real world pure nystagmus signals of six individuals. More than 500 nystagmus beats were analysed.

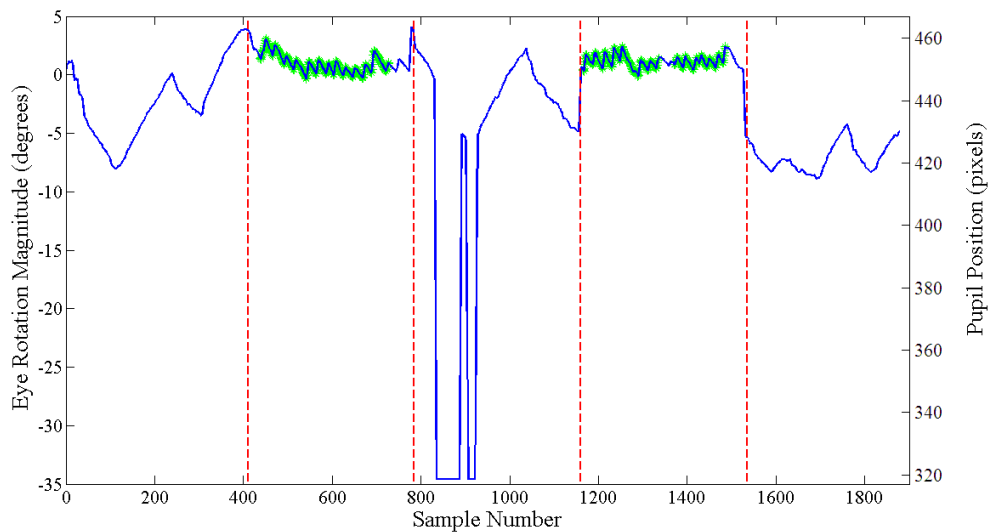
Table 5.2: Collected Data for Pure Nystagmus Signals

Signal #	# of True Positives	# of False Positives	# of True Negatives	# of False Negatives
1	74	1	5	5
2	72	1	11	10
3	73	1	12	9
4	47	1	5	5
5	75	0	5	4
6	67	0	5	2
7	61	0	6	3
8	61	0	8	8
9	49	0	7	9
10	64	0	4	1
11	133	0	1	6
12	120	0	6	22
13	28	0	19	2
14	76	0	6	6

In order to provide a more convincing demonstration of the resilience of the proposed algorithm to random components, the responses to the second type of mixed stimulation, as described in Section 5.4.1, are presented here. Figure 5.13.a shows a signal recorded from a mixed stimulus and Figure 5.13.b shows the identified nystagmus beats highlighted in green. The vertical dashed lines show the approximate transition times from one stimulus to the next.



(a.)



(b.)

Figure 5.13: Nystagmus Diagnosis in Random Signal: (a.) Original Signal, and (b.) Processed Signal

Sensitivity for these data sets was calculated in the same manner as for the pure nystagmus signals. Specificity was calculated with new definitions for the specificity parameters. These parameters were redefined to provide a more concrete basis of statistic analysis and avoid objectivity bias with regards to the definition of line segments. The number of false positives and true negatives were subsequently region based. Thus, if any data points were identified in a random component or blink region, the region was counted as a false positive. If the entire region was correctly excluded from the possible nystagmus beats by the algorithm, it was counted as a single true negative. The results with the newly defined parameters were calculated and a sensitivity of 87.79% was achieved. The specificity was found to be 91.07%, however, it should be noted that the true specificity is much higher, since only a few points were incorrectly identified as part of a nystagmus beat. If the overly conservative definition of

CHAPTER 5. SIGNAL PROCESSING

the specificity parameters are ignored, it would be reasonable to assume a specificity comparable with that found for pure nystagmus signals. Once again Appendix E may be consulted for a graphical definition of the parameters and a demonstration on the conservativeness of the specificity calculation. The recorded VOG signals with mixed content are also given in Appendix E.

Table 5.3 gives the tabulated parameters that were found by signal analysis of mixed signals from 6 individuals and approximately 350 seconds of mixed signals were analysed.

Table 5.3: Collected Data for Mixed Signals

Signal #	# of True Positives	# of False Positives	# of True Negatives	# of False Negatives
1	39	0	3	8
2	54	1	3	4
3	40	0	3	4
4	47	1	3	3
5	53	0	3	4
6	50	0	3	7
7	40	0	3	9
8	34	0	3	5
9	21	0	3	14
10	47	1	3	5
11	44	0	3	4
12	33	0	3	3
13	23	1	3	2
14	26	1	3	2

It can thus be concluded from these performance indexes that, although the sensitivity is lower than that achieved by Jansen *et al.* (2010), the constraints that reduce the sensitivity is a trade-off for a robust specificity. The performance values that were achieved are sufficient to provide the necessary proof of concept.

Chapter 6

Conclusion

From the thorough literature review in Chapter 2, an appropriate eye tracking method has been chosen. A binocular VOG goggle prototype was designed and manufactured. The binocular design facilitates recording of both eyes at the same time, while the goggle concept provides a controlled light environment. Video-oculography provides a highly accurate non-invasive solution to a potentially invasive procedure. It was also shown in Chapter 3 that the sampling rate for the selected cameras are more than sufficient. This statement is also supported by the non-aliased real world nystagmus signals recorded by the developed VOG device. Furthermore, the use of high sampling rate USB cameras, as implemented for the VOG device presented in this thesis, facilitate a cost-effective system compared to commercial systems that use expensive FireWire cameras, while still providing an adequate sampling rate.

Image processing algorithms that were developed to extract pupil location from the recorded VOG images, proved to be accurate and robust. Combined methods that provided a hybrid algorithm of colour transformation and feature tracking were implemented to extract the pupil position from each image. Histogram processing and morphological operations implemented in the tracking algorithm assist in minimizing the effect of glints, also known as corneal reflexes, located on the pupil boundary. These operations completely eliminate the effect of glints when they are located away from the pupil boundary. Ellipse fitting to the pupil boundary through least squares minimization provides accurate parameters for the ellipse that describe the pupil location and form. The algorithm provides a computationally efficient method to accurately track the pupil location from a VOG recording and provides robustness against far lateral gazes and artefact and glint locations in the image, which leaves little room for tracking improvement.

Signal processing that was discussed in the previous chapter, implemented a wavelet approach. Wavelets have only been implemented in nystagmus diagnosis in recent years. The wavelet approach provides a considerably more robust approach than conventional parametric methods. Exploitation of Daubechies 2 wavelet to identify possible quick phase locations from a signal is key for the success of the analysis algorithm. The wavelet decomposition provides a response that can be used to extract reliable information about quick phase location, whereas parametric approaches follow less reliable velocity threshold methods. However, it should be noted that the wavelet analysis is sensitive to some parameters, such as detail coefficient level threshold values and window size. Window size should be chosen such that enough samples are included

CHAPTER 6. CONCLUSION

for a meaningful wavelet transform and should also be narrow enough such that nystagmus data is not dwarfed by other motion components. Despite these sensitivities, it was shown that the algorithm can perform completely automated without any intervention and deliver reliable results. The sensitive parameters can be well-defined based on knowledge of the system and test conditions, which will provide stable performance and an automatic diagnosis capability.

Finally, the performance results of the analysis algorithm were very satisfactory and even though the sensitivity is not as high as some results found in the literature, it has been argued that the robust performance of the analysis algorithm achieved even when analysing mixed signal content, as well as the very competing specificity values, entirely justifies the decreased sensitivity. The algorithm proved successful in situations where signal content did not consist purely of nystagmus beats, something which have not been investigated anywhere else in the literature. With an average sensitivity around 90% and specificity values typically above 90%, the proposed nystagmus diagnosis algorithm prove to be reliable and most certainly provides a convincing proof of concept.

It can thus be concluded that all the project objectives were completely satisfied and results proved to be very pleasing. This provides proof of concept for a video based eye analysis system which can be used for automatic diagnosis of nystagmus and may also serve as development platform for other eye motion tests and research on ocular motor responses. The platform can also be used to expand the scope of the analysis software to include torsional tracking and pupil dilation reflexes in response to sharp light.

Appendices

Appendix A

Safety Calculations

A.1 LED Specifications

The LEDs that were chosen for illumination are the BIR-BM13E4G-2 IR LEDs manufactured by *Mobicon*. These LEDs are specified to have a wavelength of 940 nm, a half-angle (half of the viewing angle) of 50 degrees and an irradiance of 2.5 mW cm⁻². Most emission limits given by safety standards are specified in radiance, of which the units are W m⁻² sr⁻¹. Thus, in order to find the radiance of the LEDs from the specified irradiance, the solid angle of illumination needs to be calculated first.

A.1.1 Derivation of the Solid Angle

A solid angle is a three dimensional angle of which the units are steradians, and can be visualized as a pyramid shape with any number of sides of which the base will form a smooth parametric surface. With regards to lighting, the angular content of interest will be symmetrical along the viewing axis (centre axis) and the solid angle can subsequently be described as a half-angle revolved around the centre line that divides the viewing angle into two equal angles. It can thus best be visualized as a conical segment from a sphere. Figure A.1 shows a graphical representation of this concept. In order to derive the formula for a solid angle of such a section, it is necessary to consider the first principles of the solid angle. Figure A.2 shows a solid angle Ω subtended by the smooth parametric surface S at P. This means that the solid angle is the set of lines that originates at P and passes through S. Stewart (2003) gives the following equation for a solid angle,

$$\Omega(S) = \iint_S \frac{\mathbf{r} \cdot \mathbf{n}}{r^3} dS \quad (\text{A.1.1})$$

where \mathbf{r} is the radius vector from P to any point on S, $r = |\mathbf{r}|$ and \mathbf{n} is the unit normal vector directed away from P. The vectors \mathbf{r} and \mathbf{n} are represented in the xyz-cartesian space as:

$$\mathbf{r} = x \mathbf{i} + y \mathbf{j} + z \mathbf{k} \quad (\text{A.1.2})$$

APPENDIX A. SAFETY CALCULATIONS

$$\mathbf{n} = \mathbf{i} + \mathbf{j} + \mathbf{k} \tag{A.1.3}$$

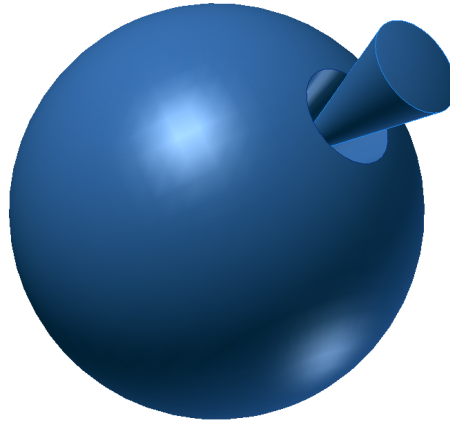


Figure A.1: Visual Representation of a Conical Section Solid Angle

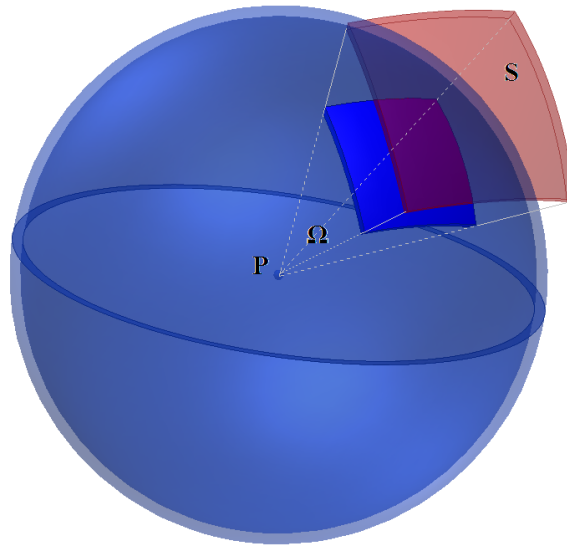


Figure A.2: Solid Angle Definition

Evaluation of the double integral in equation A.1.1 can be simplified by transforming the problem to spherical coordinates in the $\rho\phi\theta$ -cartesian space. Substituting spherical coordinates into the double integral in equation A.1.1 requires a specific transformation method. From calculus it is known that,

$$\iiint_D F(x, y, z) \, dx \, dy \, dz = \iiint_G H(\rho, \phi, \theta) J(\rho, \phi, \theta) \, d\rho \, d\phi \, d\theta \tag{A.1.4}$$

where, G is a cube in the cartesian $\rho\phi\theta$ -space and D is a spherical wedge in the xyz-space, both with infinitesimal dimensions, while J is the Jacobian for the transformation (Thomas *et al.*, 2005). The Jacobian is given by,

APPENDIX A. SAFETY CALCULATIONS

$$J(\rho, \phi, \theta) = \begin{vmatrix} \frac{\partial x}{\partial \rho} & \frac{\partial x}{\partial \phi} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial \rho} & \frac{\partial y}{\partial \phi} & \frac{\partial y}{\partial \theta} \\ \frac{\partial z}{\partial \rho} & \frac{\partial z}{\partial \phi} & \frac{\partial z}{\partial \theta} \end{vmatrix} \quad (\text{A.1.5})$$

where,

$$x = \rho \sin\phi \cos\theta \quad (\text{A.1.6})$$

$$y = \rho \sin\phi \sin\theta \quad (\text{A.1.7})$$

and

$$z = \rho \cos\phi \quad (\text{A.1.8})$$

are the conversion functions to obtain spherical coordinates. When converting the problem to $\rho\phi\theta$ -cartesian space using equations A.1.6 to A.1.8, Equations A.1.2 and A.1.3, respectively becomes:

$$\mathbf{r} = \rho \sin\phi \cos\theta \mathbf{i} + \rho \sin\phi \sin\theta \mathbf{j} + \rho \cos\phi \mathbf{k} \quad (\text{A.1.9})$$

and

$$\mathbf{n} = \sin\phi \cos\theta \mathbf{i} + \sin\phi \sin\theta \mathbf{j} + \cos\phi \mathbf{k} \quad (\text{A.1.10})$$

Using these equations for \mathbf{r} and \mathbf{n} in the $\rho\phi\theta$ -space simplifies the vector dot product in equation A.1.1 to:

$$\mathbf{r} \cdot \mathbf{n} = \rho \quad (\text{A.1.11})$$

Furthermore, determining the partial derivatives for equations A.1.6 to A.1.8 and substituting the results into equation A.1.5 and then evaluating the determinant, gives:

$$J(\rho, \phi, \theta) = \rho^2 \sin\phi \quad (\text{A.1.12})$$

Finally, transforming the problem in equation A.1.1 using equations A.1.4, A.1.11 and A.1.12 gives,

$$\Omega(\phi, \theta) = \iiint_{\mathbf{R}} \sin\phi \, d\phi \, d\theta \quad (\text{A.1.13})$$

where \mathbf{R} is the representation of \mathbf{S} in the $\rho\phi\theta$ -space and the integral over $d\rho$ is omitted, since Ω is per definition independent of ρ .

Since the solid angle of a conical section can be found by fully revolving the apex angle of the cone, the integral in equation A.1.13 can be evaluated as follows,

APPENDIX A. SAFETY CALCULATIONS

$$\Omega = \int_0^{2\pi} \int_0^\alpha \sin\phi \, d\phi \, d\theta = 2\pi(1 - \cos\alpha) \quad (\text{A.1.14})$$

where α is the apex angle of the conical section, and in this specific case, half the viewing angle of the illumination source.

A.1.2 LED Radiance

Substituting the specified half-angle for the LED into equation A.1.14 of Section A.1.1, the solid angle of the illumination source can be calculated. A half-angle of 50 degrees, as is the case for this particular LED, gives a solid angle of approximately 2.24443 steradians. With the solid angle of the LED known, the radiance of the LED can be calculated. The IEC (2007) safety standard on laser products specifies radiance in terms of irradiance and the solid angle of the source as follows,

$$L = \frac{dE}{d\Omega} \cdot \cos\theta \quad (\text{A.1.15})$$

where θ , in this case, is the viewing angle of the subject relative to the source of illumination. From equation A.1.15, the radiance of the LED can be determined. Assuming a worst case viewing angle ($\theta = 0^\circ$) and converting to SI units, the radiance of the LED can be calculated as shown in A.1.16.

$$L = \frac{2.5}{2.24443} \cdot \frac{1}{1000} \cdot \frac{100^2}{1} = 11.139 \left[\frac{\text{W}}{\text{m}^2 \text{ sr}} \right] \quad (\text{A.1.16})$$

The illumination source for each of the cameras of the VOG device developed for the purpose of this thesis will consist of twelve of the chosen LEDs. This will result in a radiance of $133.668 \text{ W m}^{-2} \text{ sr}^{-1}$ for the illumination source.

A.2 Safety Evaluation

With the radiance known, the safety of the illumination source can be evaluated. IEC (2007) gives the maximum permitted radiances over an exposure time of 100 seconds for Class 1 and Class 1M sources with regards to retinal hazards in table E.1 of the standard. Table A.1 shows an extract from the IEC table. Since all standard LEDs fall within the Class 1 category and recording time will typically be limited to one minute, this table may be used for the safety evaluation of the illumination source. The interpolated maximum radiance at 940 nm from Table A.1 is $146'504.2 \text{ W m}^{-2} \text{ sr}^{-1}$. Furthermore, Matthes (2000) concluded that radiances of $1800 \text{ W m}^{-2} \text{ sr}^{-1}$ over exposure times of 100 seconds are harmless. Recalling that the radiance of the illumination source is $133.7 \text{ W m}^{-2} \text{ sr}^{-1}$ from Section A.1.2, it can thus be concluded that the illumination source is well within safe limits according to both these established safety limits.

It should be noted that IEC (2007) specifies the solid angle with a different mathematical relationship to α . IEC (2007) specifies the solid angle as shown in equation A.2.1.

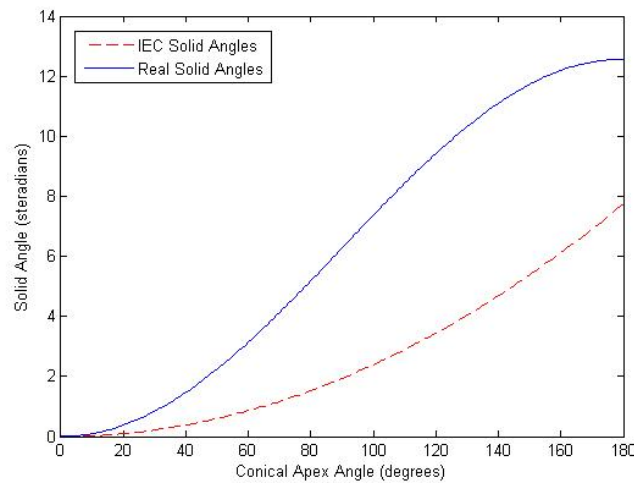
APPENDIX A. SAFETY CALCULATIONS

Table A.1: IEC Maximum Radiance of Class 1 Source Over 100s

Wavelength nm	Radiance $W \cdot m^{-2} \cdot sr^{-1}$
850	96 403
900	121 365
950	152'789
1000	192'350

$$\Omega = \frac{\pi \alpha^2}{4} \quad (\text{A.2.1})$$

The use of equation A.2.1 to calculate the solid angle results in an inaccurate solution. The comparison between the solid angles as calculated by IEC (2007) and those calculated by the mathematical definition of a solid angle is shown in Figure A.3. The IEC (2007) standard does not specify how equation A.2.1 is derived. Since this equation results in a more conservative approach, safety evaluation of the radiance calculated by this approach should also be considered. The solid angle for a LED with half-angle calculated with equation A.2.1 is 0.598 steradians, which results in a radiance of $41.798 \text{ W m}^{-2} \text{ sr}^{-1}$. Thus, an illumination source making use of twelve of these LEDs will have a radiance of $501.576 \text{ W m}^{-2} \text{ sr}^{-1}$. Comparing this value to the maximum radiance levels stated earlier in this section, it is clear that the designed illumination source is well within safe exposure limits and should not pose any health hazards despite the considerably conservative approach that was followed during calculations.

**Figure A.3:** Comparison of Solid Angles as Calculated by IEC and Solid Angle Definition

Appendix B

Frequency Domain Theory

Frequency domain filtering discussed in Section 4.1.4 is an important tool used in various image processing algorithms. This appendix discusses the theory regarding the Fourier Transform (FT) and the discrete Fourier transform (DFT), as well as the derivation of some important mathematical formulae.

B.1 Fourier Transform Theory

French mathematician Joseph Fourier introduced an invaluable tool for signal processing and image processing alike with his contribution of the Fourier transform. Fourier stated that any periodic function can be expressed as a sum of sines and cosines with different frequencies and coefficients. This sum is known today as a Fourier series. The mathematical definition of a Fourier series as stated by Gonzales and Woods (2008) is given in equation B.1.1.

$$f(t) = \sum_{n=-\infty}^{\infty} c_n e^{j \frac{2\pi n}{T} t} \quad (\text{B.1.1})$$

where $e^{j \frac{2\pi n}{T} t}$ is obtained from Euler's formula and c_n are the coefficients given by:

$$c_n = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(t) e^{-j \frac{2\pi n}{T} t} dt$$

for $n = 0, \pm 1, \pm 2, \pm 3 \dots$

Fourier's theory is not limited to periodic functions however, but can also be applied to non-periodic functions with finite sub-curve areas. The mathematical definition for the Fourier transform, also known as the continuous Fourier transform (CFT), is given in equation B.1.2 (Gonzales and Woods, 2008).

$$\mathcal{F}\{f(t)\} = F(\mu) = \int_{-\infty}^{\infty} f(t) e^{-j 2\pi \mu t} dt \quad (\text{B.1.2})$$

APPENDIX B. FREQUENCY DOMAIN THEORY

A remarkable characteristic of both the Fourier series and Fourier transform, is that any signal represented in the Fourier domain can be completely reconstructed in the time domain via the inverse Fourier transform (IFT) without any loss of information. The definition of the inverse Fourier transform is given in equation B.1.3 (Gonzales and Woods, 2008). The period of the periodic signal is given by T .

$$f(t) = \int_{-\infty}^{\infty} F(\mu) e^{j 2\pi\mu t} d\mu \quad (\text{B.1.3})$$

The above characteristic grants the ability to transform a signal or image back and forth between the spacial domain and the Fourier domain, also known as the frequency domain, such that a signal or image can be processed in the frequency domain and then fully reconstructed to the spacial domain. This is arguably one of the most important properties of the Fourier transform, and certainly very important for the application of the Fourier transform in image processing algorithms as discussed in Section 4.1.

B.2 Discrete Fourier Transform

The theory presented in Section B.1 discusses the continuous Fourier transform. For the application on digital signal and image processing, a computationally efficient version of the Fourier transform needs to be implemented. A numerical algorithm of the FT for the transformation of digital signals is given by the discrete Fourier transform (DFT). The derivation of the DFT, as given by Gonzales and Woods (2008), is given below. The first step is to start with the definition of the Fourier transform:

$$\tilde{F}(\mu) = \int_{-\infty}^{\infty} \tilde{f}(t) e^{-j 2\pi\mu t} dt \quad (\text{B.2.1})$$

where \tilde{f} indicates a sampled function and \tilde{f} is defined as:

$$\tilde{f}(t) = f(t) s_{\Delta T}(t) = \sum_{n=-\infty}^{\infty} f(t) \delta(t - n\Delta T) \quad (\text{B.2.2})$$

where δ denotes an impulse and ΔT is the time between consecutive impulses.

The DFT of a sampled function can thus be rewritten from equations B.2.1 and B.2.2 as:

$$\tilde{F}(\mu) = \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} f(t) \delta(t - n\Delta T) e^{-j 2\pi\mu t} dt \quad (\text{B.2.3})$$

The equality in equation B.2.3 can be simplified sampling theory to the DFT definition given in equation B.2.4.

APPENDIX B. FREQUENCY DOMAIN THEORY

$$\tilde{F}(\mu) = \sum_{n=-\infty}^{\infty} \tilde{f}_n e^{-j2\pi\mu n\Delta T} \quad (\text{B.2.4})$$

In image processing literature the convention is to work with u and v as the frequency domain variables and x and y as the spacial domain variables. For this case the equality in equation B.2.4 can be rewritten as:

$$F(u) = \sum_{x=0}^{M-1} f(x) e^{-j2\pi ux/M} \quad u = 0, 1, 2, \dots, M-1 \quad (\text{B.2.5})$$

where M denotes the number of samples.

The inverse discrete Fourier transform (IDFT) is defined in equation B.2.6.

$$f(x) = \frac{1}{M} \sum_{u=0}^{M-1} F(u) e^{j2\pi ux/M} \quad x = 0, 1, 2, \dots, M-1 \quad (\text{B.2.6})$$

Image processing algorithms poses the special need of a 2-D Fourier transform that can calculate the Fourier transform of a matrix. The definitions for a 2-D DFT and 2-D IDFT, as stated by Gonzales and Woods (2008), is given in equations B.2.7 and B.2.8 respectively.

$$F(u, v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) e^{-j2\pi(ux/M+vy/N)} \quad (\text{B.2.7})$$

where $f(x, y)$ is a digital image with dimensions of $M \times N$, and $u = 0, 1, 2, \dots, M-1$ and $v = 0, 1, 2, \dots, N-1$ are the discrete variables.

$$f(x, y) = \frac{1}{MN} \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u, v) e^{j2\pi(ux/M+vy/N)} \quad (\text{B.2.8})$$

where x and y are defined with the same values as u and v , respectively, in B.2.7. Recall that x and y are the spacial variables represented by u and v in the frequency domain. Note that the 2-D FT of an image can be found by simply finding the sequential 1-D FT's along the rows and then along the columns of the image.

B.3 Fast Fourier Transform

Application of the Fourier transform only really became practical in the 1960's with the discovery of the fast Fourier transform (FFT). The FFT provides a feasible alternative to the, computationally expensive, traditional FT. The FFT provides a computational efficiency that allows computations to be $\frac{M}{\log_2 M}$ times quicker than that of the traditional FT (Gonzales and Woods, 2008). The derivation of the fast Fourier transform following the original successive-doubling method algorithm is given below. For this algorithm the

APPENDIX B. FREQUENCY DOMAIN THEORY

number of samples, M , is assumed to be of the form 2^n , where n is positive integer. The explanation of the method was slightly modified from the method presented by Gonzales and Woods (2008). The definition for the DFT in equation B.2.5 can be rewritten as shown in equation B.3.1

$$F(u) = \sum_{x=0}^{M-1} f(x) W_M^{ux} \quad (\text{B.3.1})$$

with $u = 0, 1, 2, \dots, M-1$ and where

$$W_M = e^{-j2\pi/M} \quad (\text{B.3.2})$$

The signal being evaluated can be split into two sub-signals, one containing the even numbered samples and the other containing the uneven numbered samples as shown in equations B.3.3 and B.3.4, where $x = 0, 1, 2, \dots, \frac{M}{2} - 1$ in both cases.

$$f_{\text{even}}(x) = f(2x) \quad (\text{B.3.3})$$

$$f_{\text{odd}}(x) = f(2x + 1) \quad (\text{B.3.4})$$

By splitting the input signal as shown in equations B.3.3 and B.3.4, equation B.3.1 can be rewritten as shown below.

$$F(u) = \sum_{x=0}^{\frac{M}{2}-1} f(2x) W_M^{2xu} + \sum_{x=0}^{\frac{M}{2}-1} f(2x+1) W_M^{(2x+1)u} \quad (\text{B.3.5})$$

From the definition of W_M , it is easy to show that $W_M^2 = W_{\frac{M}{2}}$. Denoting the Fourier transforms of $f_{\text{even}}(x)$ and $f_{\text{odd}}(x)$ as $F_{\text{even}}(u)$ and $F_{\text{odd}}(u)$, respectively, it then follows from equations B.3.1 and B.3.5 that:

$$F(u) = F_{\text{even}}(u) + W_M^u F_{\text{odd}}(u) \quad (\text{B.3.6})$$

Furthermore, from the fact that F_{even} and F_{odd} are periodic, each with a period of $N_{\frac{M}{2}}$, it can be stated that:

$$F_{\text{even}}(u) = F_{\text{even}}\left(u + \frac{N}{2}\right) \quad (\text{B.3.7})$$

and

$$F_{\text{odd}}(u) = F_{\text{odd}}\left(u + \frac{N}{2}\right) \quad (\text{B.3.8})$$

Also, from symmetry and periodicity follows the respective statements in equations B.3.9 and B.3.10.

APPENDIX B. FREQUENCY DOMAIN THEORY

$$W_M^{u+\frac{M}{2}} = -W_M^u \quad (\text{B.3.9})$$

$$W_M^{u+M} = W_M^u \quad (\text{B.3.10})$$

By utilising equations B.3.7 through B.3.10, equation B.3.6 can be rewritten as shown in equation B.3.11.

$$F\left(u+\frac{M}{2}\right) = F_{\text{even}}(u) - W_M^u F_{\text{odd}}(u) \quad (\text{B.3.11})$$

It can be shown that the implementation of the procedure discussed above present a computational advantage of $\frac{M}{\log_2 M}$, as stated earlier in this section. The implementation of the DFT and the FFT for the efficient numerical implementation of the Fourier transform provides a feasible and efficient alternative to spacial filtering via brute-force computations. Subsequently, the computationally demanding algorithms associated with image processing can be facilitated and implemented in realistic applications. The following section will discuss how frequency domain filtering can replace spacial domain filtering by declaring convolution in terms of the frequency domain.

B.4 Convolution in the Frequency Domain

As stated repeatedly before, the implementation of filtering in the spacial domain is computationally speaking, extremely expensive. Therefore, filters are applied in the frequency domain, and hence the inclusion of this chapter on Fourier transforms to review the relevant, related theory. Convolution was defined in Section 4.1.3.1. Convolution of two 1-D functions with each other can also be defined as in equation B.4.1, where τ is a dummy variable and t is the spacial domain variable.

$$f(t) \star h(t) = \int_{-\infty}^{\infty} f(\tau) h(t - \tau) d\tau \quad (\text{B.4.1})$$

Since the interest lies in frequency domain filtering, the Fourier transform of the equality in equation B.4.1 needs to be found. The Fourier transform of this equation is shown below, where μ is the frequency domain variable.

$$\mathcal{F}\{f(t) \star h(t)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\tau) h(t - \tau) e^{-j2\pi\mu t} dt \quad (\text{B.4.2})$$

Using the properties of e and from Fubini's theorem (Stewart, 2003), the above can be reorganized as shown in equation B.4.3.

$$\mathcal{F}\{f(t) \star h(t)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\tau) h(t - \tau) e^{-j2\pi\mu t} dt d\tau \quad (\text{B.4.3})$$

APPENDIX B. FREQUENCY DOMAIN THEORY

Furthermore, since $f(\tau)$ is independent of t and substituting $z = t - \tau$ such that $dt = dz$, the inner integral of equation B.4.3 becomes:

$$\int_{-\infty}^{\infty} f(\tau)h(t - \tau)e^{-j2\pi\mu t} dt = f(\tau) \int_{-\infty}^{\infty} h(z)e^{-j2\pi\mu\tau} e^{-j2\pi\mu z} dz \quad (\text{B.4.4})$$

From equations B.4.3 and B.4.4 it then follows:

$$\mathcal{F}\{f(t)\star h(t)\} = \int_{-\infty}^{\infty} f(\tau) \int_{-\infty}^{\infty} h(z)e^{-j2\pi\mu\tau} e^{-j2\pi\mu z} dz d\tau \quad (\text{B.4.5})$$

Simplifying equation B.4.5 then gives:

$$\mathcal{F}\{f(t)\star h(t)\} = H(\mu) \int_{-\infty}^{\infty} f(\tau)e^{-j2\pi\mu\tau} d\tau = H(\mu)F(\mu) \quad (\text{B.4.6})$$

From the definition derived above it can thus be said that convolution in the spacial domain gives the same results as multiplication in the frequency domain. Although this derivation was based on the convolution of 1-D functions, it can be expanded for multidimensional functions. Using this property while implementing the FFT, greatly simplify and accelerate filtering computations.

Appendix C

Manufactured Device and Selected Components

This chapter is included to show the manufactured VOG device along with some of the chosen components.

Figure C.1 shows the VOG assembly, labeled number one in the exploded model of Figure 3.8 in Section 3.5, from the face contact side. Figure C.2 shows the VOG assembly from the front with the visual shield, labeled number eight in Figure 3.8, added to the device.

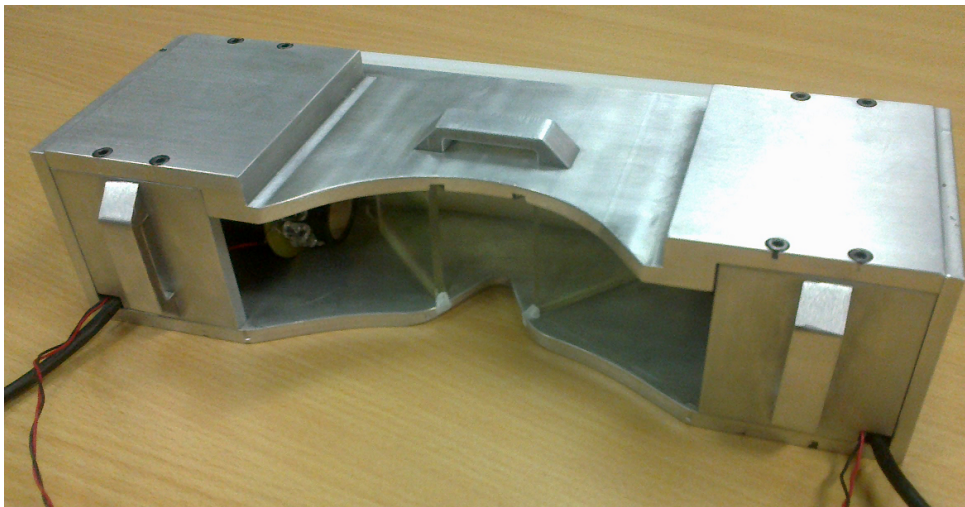


Figure C.1: VOG Device Face Contact Side

APPENDIX C. MANUFACTURED DEVICE AND SELECTED COMPONENTS

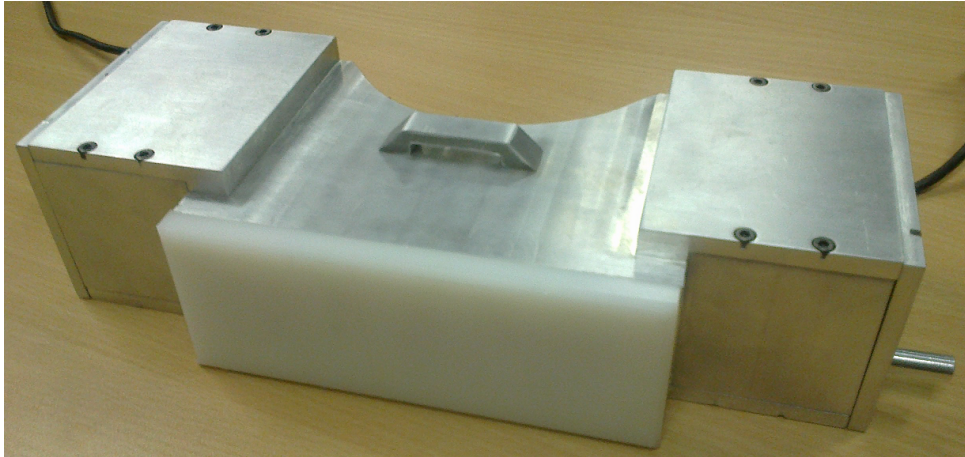


Figure C.2: VOG Device Front Side and Visual Shield

Figure C.3 and Figure C.4 show the selected IR bandpass filter, labeled number two in Figure 3.8, and dichroic filter, labeled number seven in Figure 3.8, respectively.

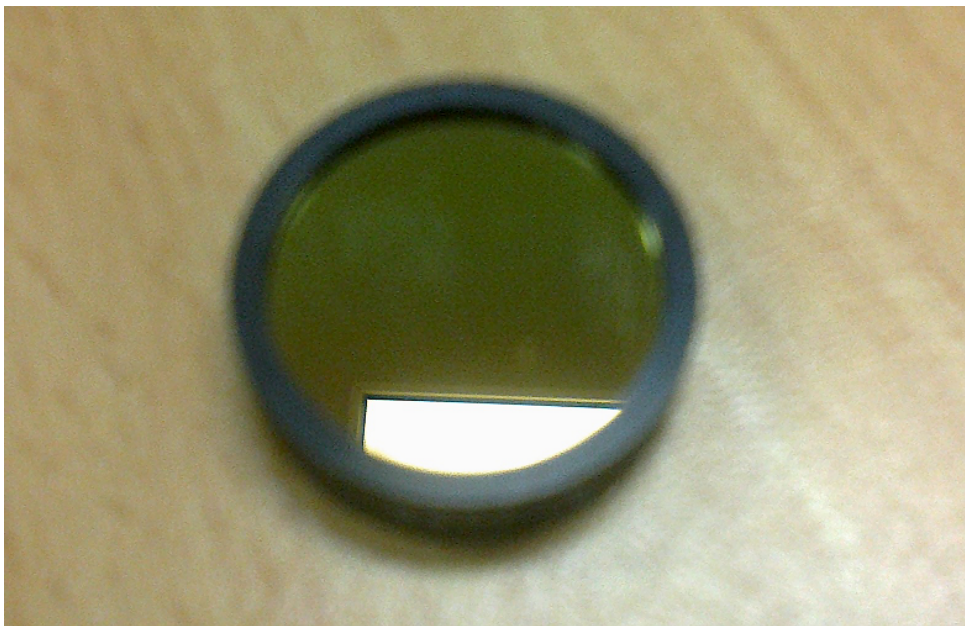


Figure C.3: IR Bandpass Filter

APPENDIX C. MANUFACTURED DEVICE AND SELECTED COMPONENTS



Figure C.4: Dichroic Filter

Figure C.5 shows the lens mounting socket that was specially designed to mount the custom selected lens to the standard PS3 Eye circuit board. This part was labeled number four in Figure 3.8.



(a.)



(b.)

Figure C.5: Custom Mounting Socket: (a.) Front View, and (b.) Rear View

Figure C.6 shows an IR light source PCB, labeled number five in Figure 3.8.

APPENDIX C. MANUFACTURED DEVICE AND SELECTED COMPONENTS



(a.)



(b.)

Figure C.6: LED PCB: (a.) Front View, and (b.) Rear View

Figure C.7 shows the USB power supply circuit board for the LED light sources.



Figure C.7: USB LED Power Circuit

Appendix D

Feature Tracking and Related Theory

Chapter 4 discuss two main types of techniques with which pupil tracking can be implemented, namely histogram processing, or colour transformation, and feature tracking. This appendix is included to provide some additional theory and methods that were not presented in Chapter 4, but will shed more light on the extent of the scope for some methods.

D.1 Bresenham Algorithm

Recall from Section 4.2.2.1 that the Hough transform draws a circle with a predefined radius around each edge pixel and accumulates the points of the circle. The circle that is drawn around each of the edge pixels is generally referred to as a digital circle. The motivation behind the use of a digital circle is that it provides a circle with unity line width and no gaps. Here, unity line width is defined as a line width of one pixel. Note that, since an image is a matrix, the exact coordinates of a perfect circle cannot be indicated, and therefore the coordinate positions are rounded to integers such that the appropriate image elements can be accessed. This rounding scheme creates a gapped circle when calculating circle coordinates along one dimension as shown in Figure D.1.

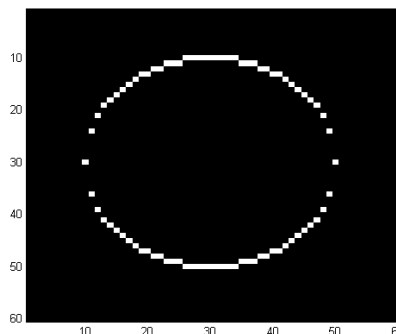


Figure D.1: Circle Rounding Errors

The Bresenham algorithm provides an efficient method to construct a digital circle. The first step is to calculate to octants of the circle, one along each spacial dimension of the

APPENDIX D. FEATURE TRACKING AND RELATED THEORY

image. These two octants are then appended to form one quarter of the circle. The quarter is mirrored around one axis to find a half-circle, and the half-circle is mirrored around the second axis to find the full circle. The final step is to translate all the elements of the circle with the desired centre coordinate values. Equations D.1.1 through D.1.6 gives the mathematical procedure for the Bresenham algorithm.

$$y_1(i) = x_2(i) = i \quad \text{for } i = 0, 1, \dots, r \quad (\text{D.1.1})$$

$$x_1(i) = \text{round}(\sqrt{r^2 - (y_1(i))^2}) \quad \text{for } i = 0, 1, \dots, r \quad (\text{D.1.2})$$

$$y_2(i) = \text{round}(\sqrt{r^2 - (x_2(i))^2}) \quad \text{for } i = 0, 1, \dots, r \quad (\text{D.1.3})$$

$$\begin{aligned} Y_1 &= [y_1 \quad y_2] \\ X_1 &= [x_1 \quad x_2] \end{aligned} \quad (\text{D.1.4})$$

$$\begin{aligned} Y_2 &= Y_4 = -Y_1 \\ Y_3 &= Y_1 \\ X_2 &= X_1 \\ X_3 &= X_4 = -X_1 \end{aligned} \quad (\text{D.1.5})$$

$$\begin{aligned} \mathbf{Y} &= [Y_1 \quad Y_2 \quad Y_3 \quad Y_4] \\ \mathbf{X} &= [X_1 \quad X_2 \quad X_3 \quad X_4] \end{aligned} \quad (\text{D.1.6})$$

Figure D.1 shows a digital circle created by the Bresenham algorithm.

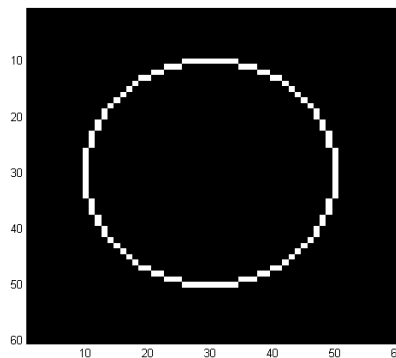


Figure D.2: Digital Circle

D.2 Ellipse Fitting Through Least Squares Minimization

Ellipse fitting through least squares minimization as implemented in this thesis and discussed in Section 4.3, requires the ellipse to be defined in a global coordinate system.

The first objective is to define the ellipse with a quadratic equation. In this manner, the ellipse parameters can be calculated from the given coordinates as extracted from an image. With the ellipse expressed in this form, the least squares minimization problem can be defined.

D.2.1 Ellipse Definition

An ellipse is defined mathematically as shown in equation D.2.1 (Stewart, 2003). Figure D.2.1 shows an ellipse in the case where $(x_0, y_0) = (0, 0)$. This form of the ellipse does not account for the situation where the ellipse is rotated around its origin. Therefore, the definition in equation D.2.1 only defines the ellipse in a local coordinate system. In order to generalize the ellipse for a global coordinate system, coordinate transformation can be implemented. The geometrical relation between a local- and global coordinate system is shown in Figure D.4.

$$\frac{(x - x_0)^2}{p^2} + \frac{(y - y_0)^2}{q^2} = 1 \quad (\text{D.2.1})$$

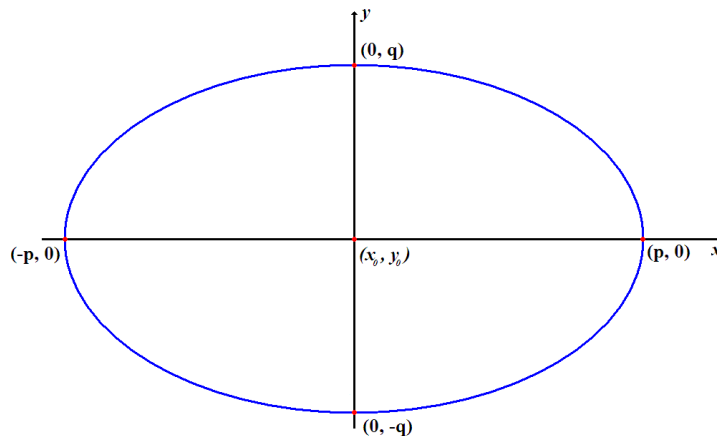


Figure D.3: Ellipse

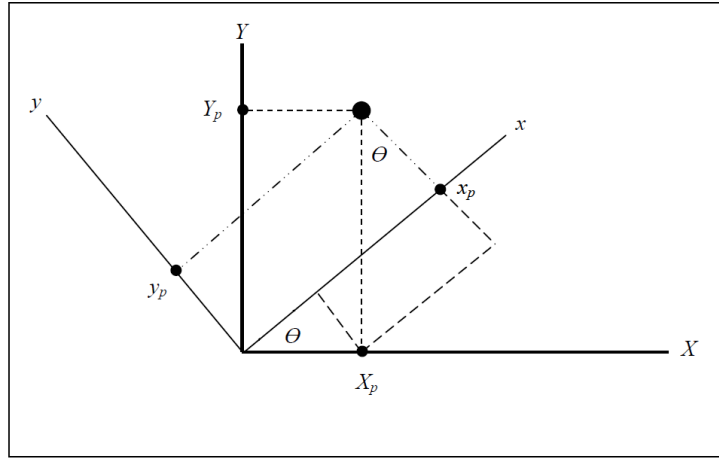


Figure D.4: Coordinate Transformation (Coetzee, 2006)

The coordinate transformation for the coordinate system shown in Figure D.4 is given in equation D.2.2 (Coetzee, 2006).

$$\begin{bmatrix} x_p \\ y_p \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} X_p \\ Y_p \end{bmatrix} \quad (\text{D.2.2})$$

From the above, an ellipse rotated around its origin by θ degrees in a global xy -coordinate system is given in equation D.2.3. Note that the global coordinates of the ellipse are available from the pupil edge detection. These coordinates are the (X_p, Y_p) coordinates. The interest lies with defining the ellipse in the local coordinate system where the ellipse is not rotated, such that the coordinates can be used in the mathematical definition of an ellipse as shown in equation D.2.1. The origin, (X_0, Y_0) is included in the transformation.

$$\frac{((X_p - X_0) \cos \theta + (Y_p - Y_0) \sin \theta)^2}{p^2} + \frac{(-(X_p - X_0) \sin \theta + (Y_p - Y_0) \cos \theta)^2}{q^2} = 1 \quad (\text{D.2.3})$$

The generalized form, as shown in equation 4.3.2, can be found by expanding the equation above. The ellipse parameters can then be expressed in terms of the coefficients by solving simultaneous equations. The coefficients are found from a least square minimization as discussed in Section 4.3 and can then be plugged into the parameter equations shown in equations 4.3.6 through 4.3.10.

Appendix E

Test Data

This appendix is included to provide clarity about the definition of the parameters for sensitivity and specificity. The collected data sets that were used to calculate the sensitivity and specificity are also given in this appendix.

E.1 Parameter Definitions

This section restates the definitions of the parameters that were used to evaluate the sensitivity and specificity values of the algorithm and also provide graphical illustrations of the definitions. The definitions of the parameters, as given in Section 5.4.2, are stated below.

Number of true positives: The number of real phases, quick and slow, that were identified by the algorithm.

Number of false positives: The sum of all line segments that do not belong to a nystagmus beat, but were identified by the algorithm.

Number of true negatives: The sum of all line segments that do not belong to a nystagmus beat and were also not identified by the algorithm.

Number of false negatives: All real quick phases and slow phases that were not identified by the algorithm.

Figure E.1 shows a hypothetical analysis of a pure nystagmus signal with a transition phase at the start. The numbered regions are used to indicate where the positives and negatives, that were found by the analysis, are. The highlighted regions were identified as nystagmus beats by the algorithm. Note once again that the positives and negatives are counted per line segment in all cases for pure nystagmus signals.

APPENDIX E. TEST DATA

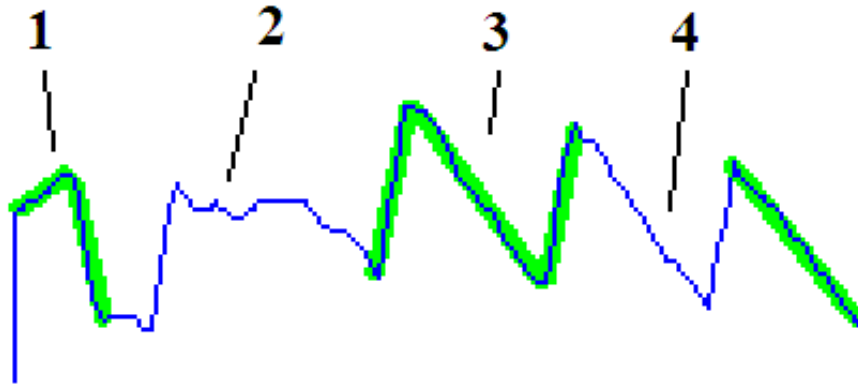


Figure E.1: Parameter Definitions for Pure Nystagmus

The first numbered region contains two highlighted line segments that do not belong to a nystagmus beat. Thus, this region contributes two false positives. The second region contains approximately three non-highlighted line segments that also do not belong to any nystagmus beats and thus this region contributes three true negatives. The third numbered region contains three highlighted line segments that are all part of a nystagmus beat and thus three true positives are contributed by this region. The last numbered region contains two non-highlighted and one highlighted line segments, all of which are part of a nystagmus beat. Therefore, the last region contributes two false negatives and one true positive.

For signal analysis of signals that contain both nystagmus- and random components, the number of false positives and true negatives were determined on a region basis. The reason for the redefinition of these parameters for mixed signal analysis, is to avoid objectivity bias regarding line segment definition. However, the per region approach does present a very conservative approach since even the smallest number of false positives result in the entire region of random motion components to be counted as false positives. Consider the signal in Figure E.2. The vertical dashed lines are the separators between the pure nystagmus content region and the random motion components regions on either side.

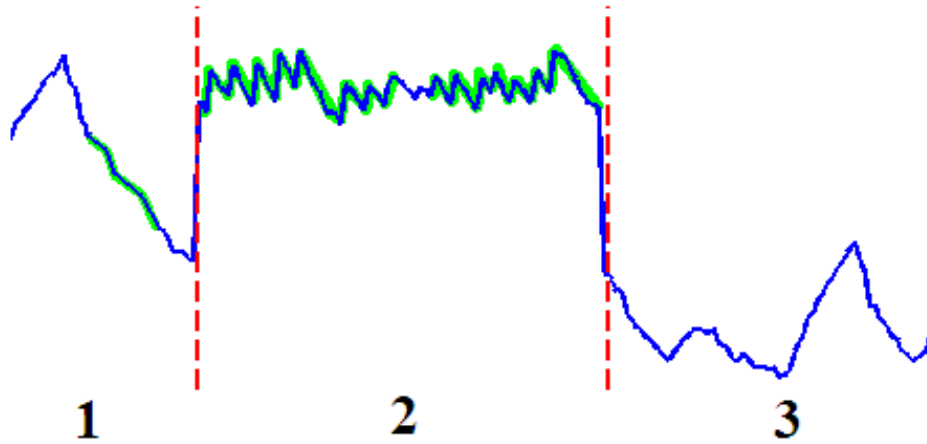


Figure E.2: Parameter Definitions for Mixed Signals

The first numbered region contains no nystagmus beats, but is counted as a false positive since there are a number of highlighted samples. The third region also contains no nystagmus beats and is counted as a true negative, since there are no highlighted samples. The second region contains only nystagmus components and the parameters are determined as discussed for pure nystagmus signals. Therefore, there are 29 true positives and 3 false negatives.

To demonstrate the conservativeness of the region based approach, consider the following analysis. It may be argued that there are approximately 2 line segments in the first numbered region and 5 line segments in the third numbered region of Figure E.2. Using a line based approach, the analysis would have identified 6 true negatives and 1 false positive. The specificity of the analysis, as defined by equation 5.4.2, is given by,

$$\text{Specificity} = \frac{6}{6 + 1} = 85.71\%$$

while for a region based analysis the specificity, calculated in a similar manner, is 50%.

E.2 Collected Data

This section includes presents the VOG data recordings for the signals analyzed and for which specific parameters are recorded in Tables 5.2 and 5.3 in Section 5.4.2. Figures E.3 through E.16 shows the pure nystagmus signals that were analysed and Figures E.17 through E.30 shows the mixed content signals that were analysed.

APPENDIX E. TEST DATA

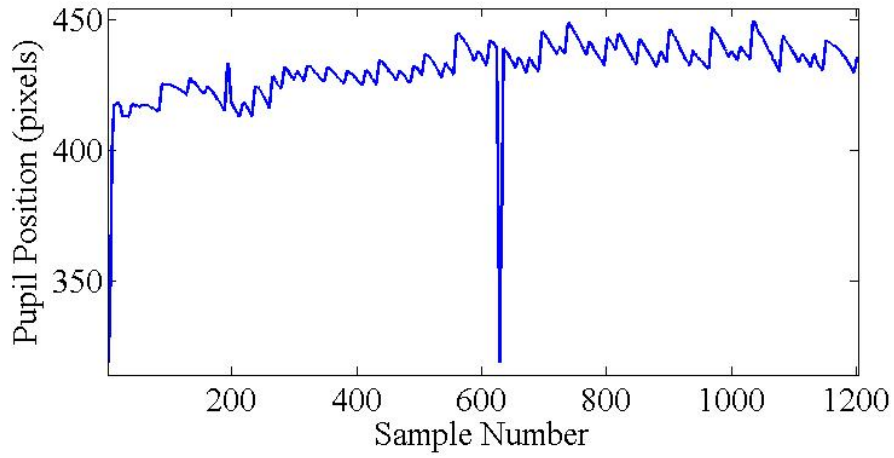


Figure E.3: Signal 1

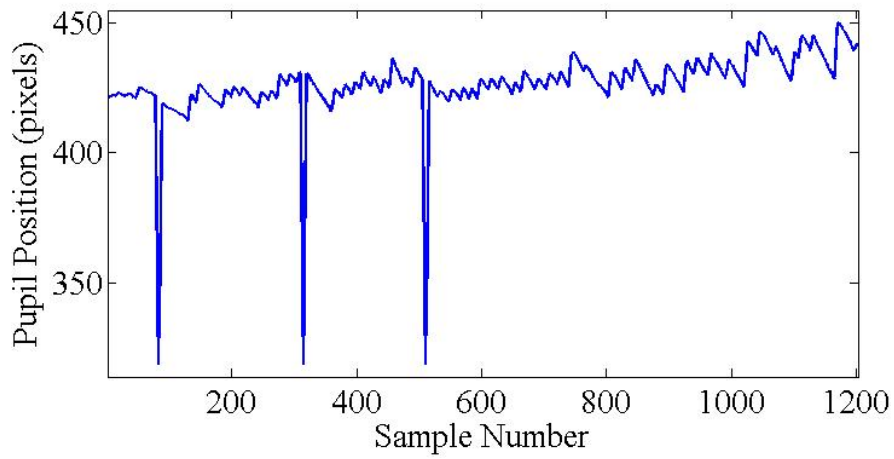


Figure E.4: Signal 2

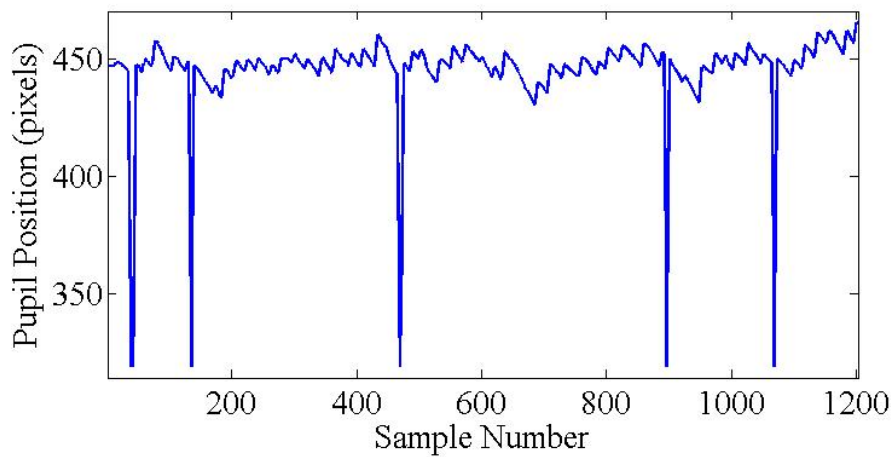


Figure E.5: Signal 3

APPENDIX E. TEST DATA

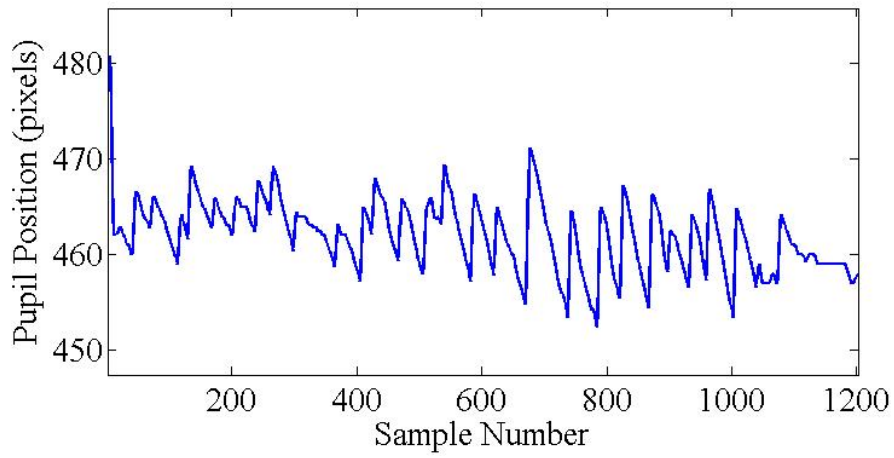


Figure E.6: Signal 4

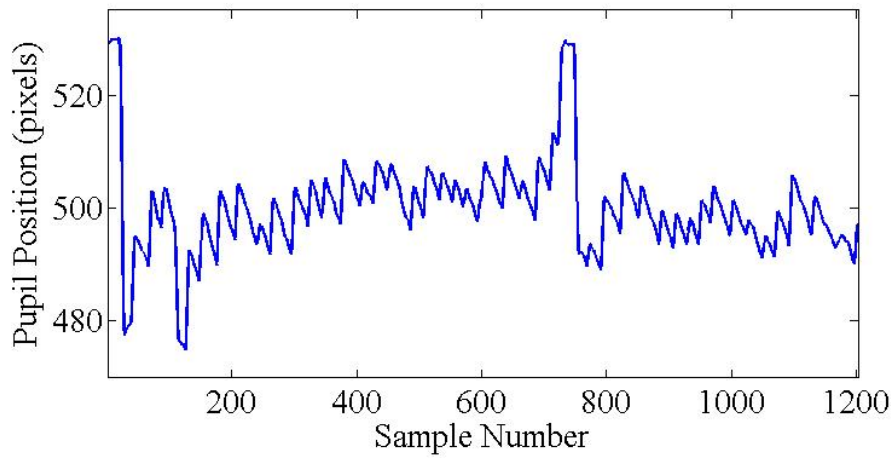


Figure E.7: Signal 5

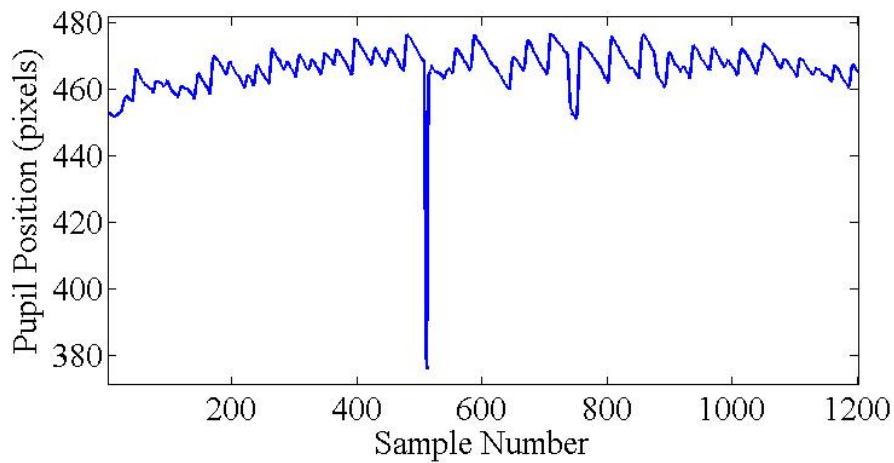


Figure E.8: Signal 6

APPENDIX E. TEST DATA

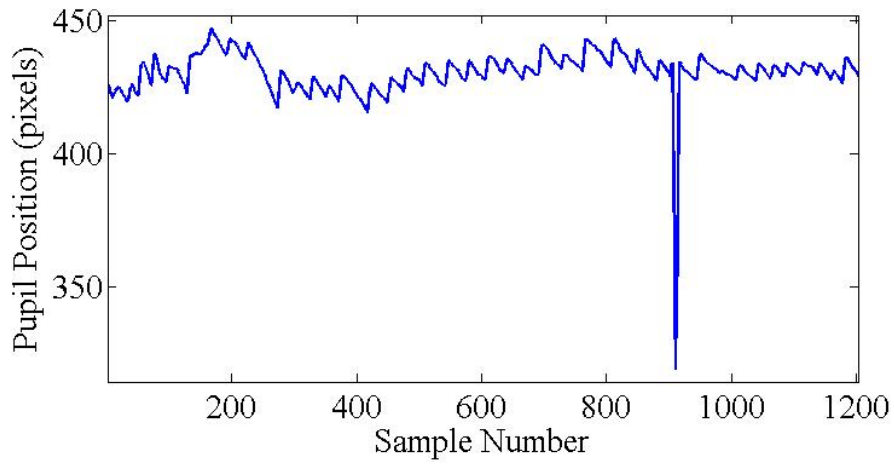


Figure E.9: Signal 7

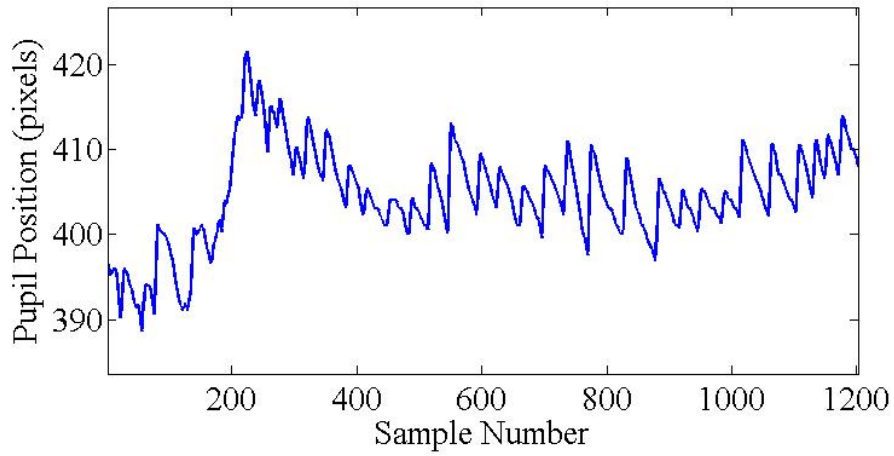


Figure E.10: Signal 8

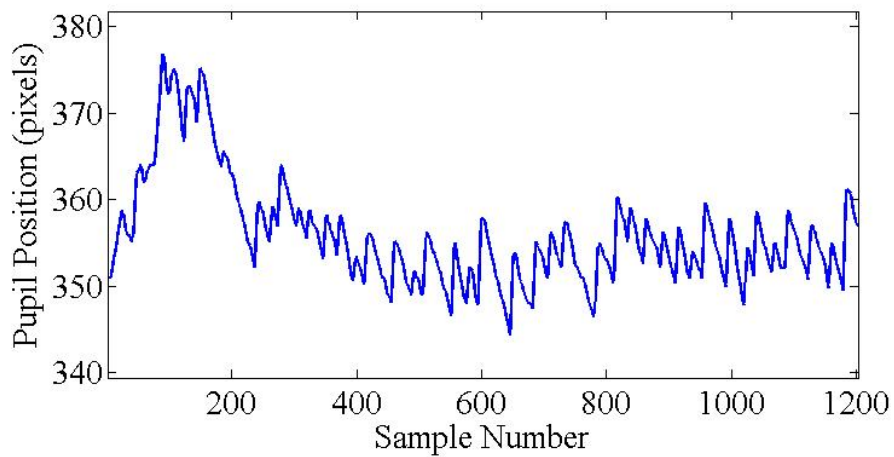


Figure E.11: Signal 9

APPENDIX E. TEST DATA

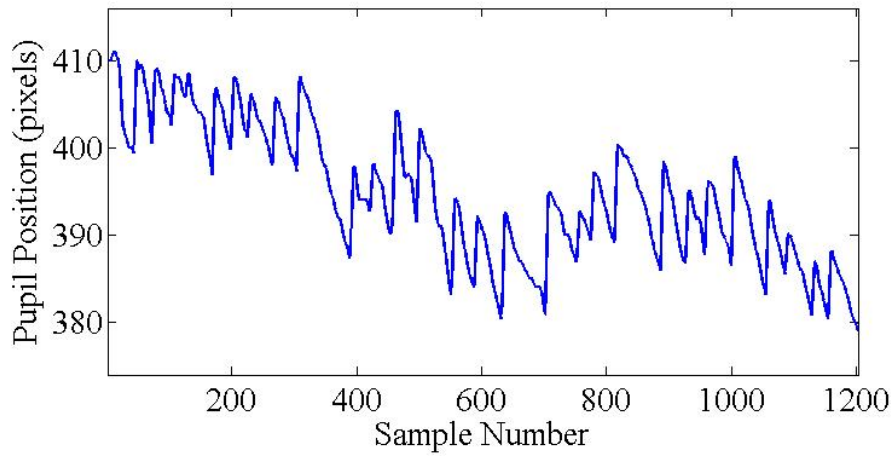


Figure E.12: Signal 10

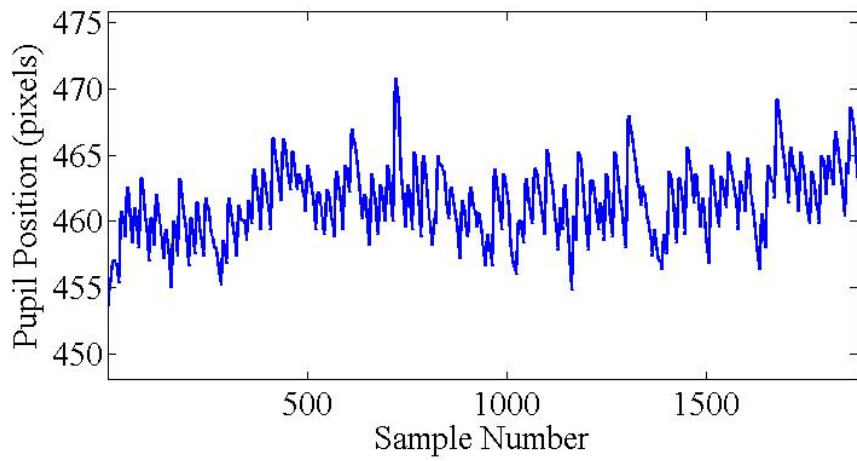


Figure E.13: Signal 11

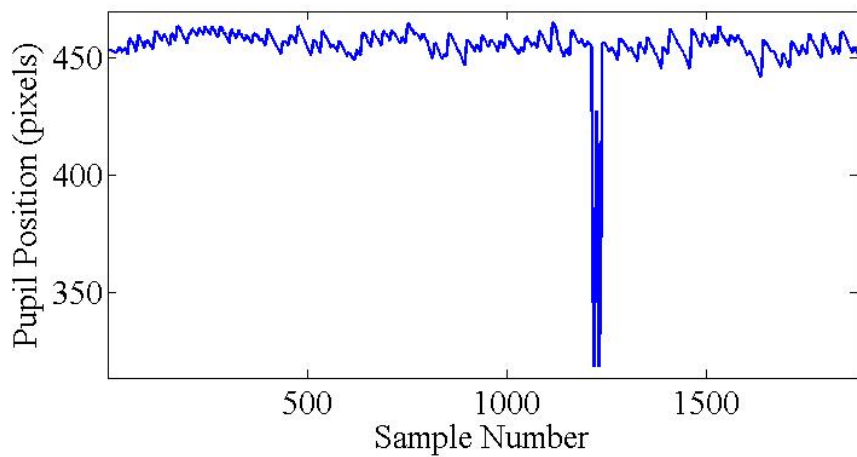


Figure E.14: Signal 12

APPENDIX E. TEST DATA

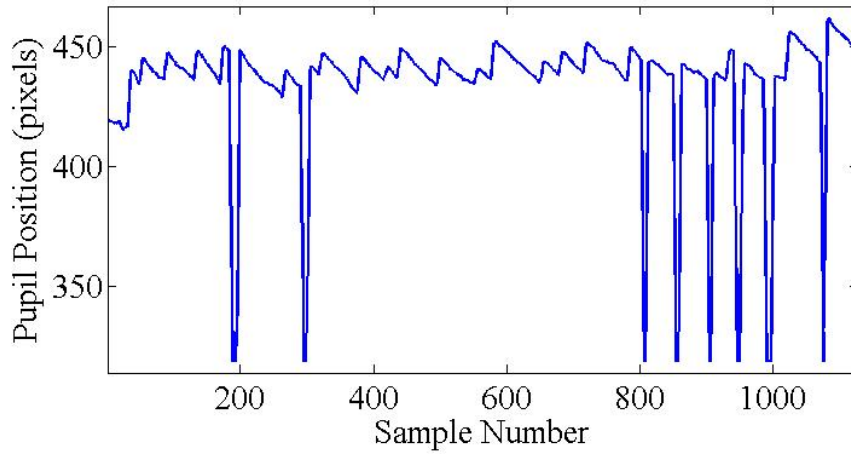


Figure E.15: Signal 13

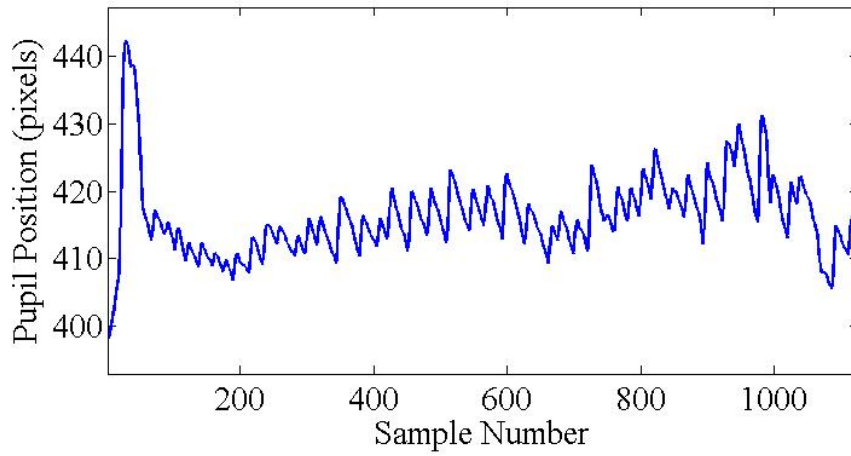


Figure E.16: Signal 14

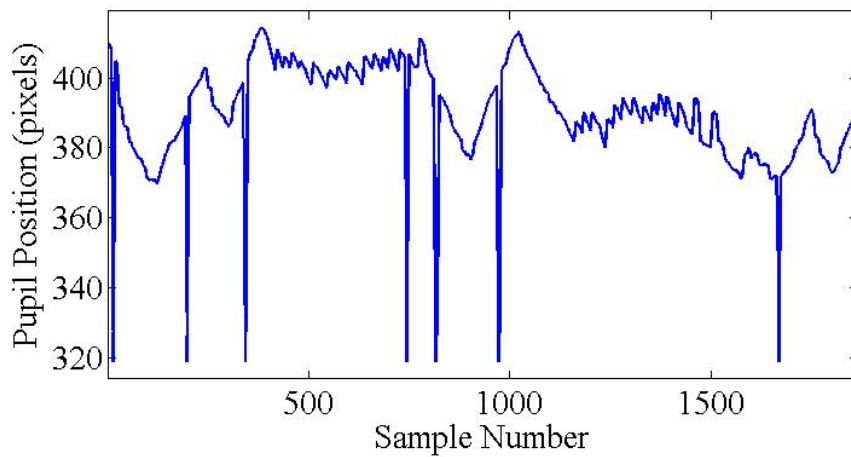


Figure E.17: Mixed Signal 1

APPENDIX E. TEST DATA

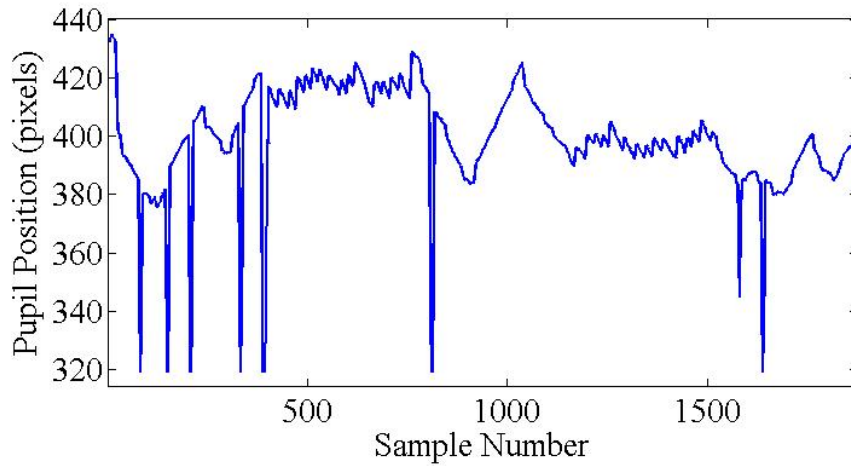


Figure E.18: Mixed Signal 2

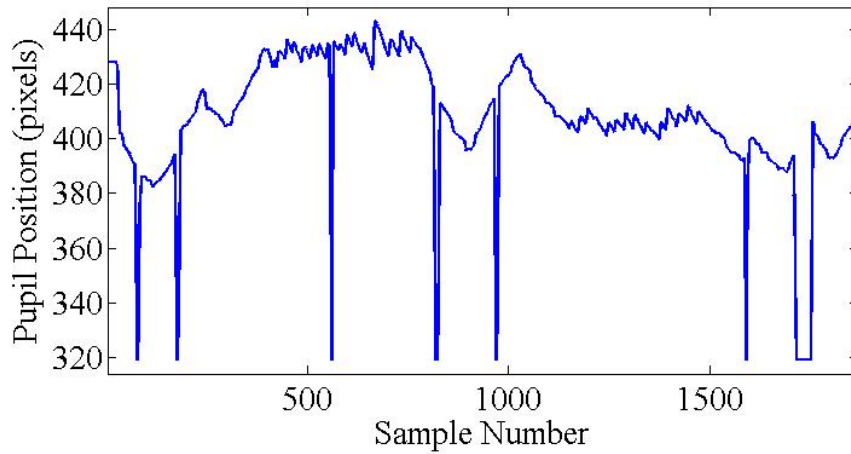


Figure E.19: Mixed Signal 3

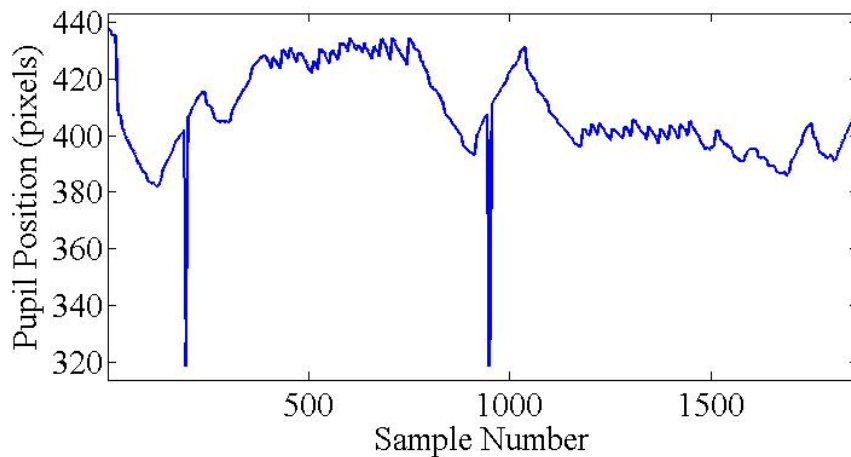


Figure E.20: Mixed Signal 4

APPENDIX E. TEST DATA

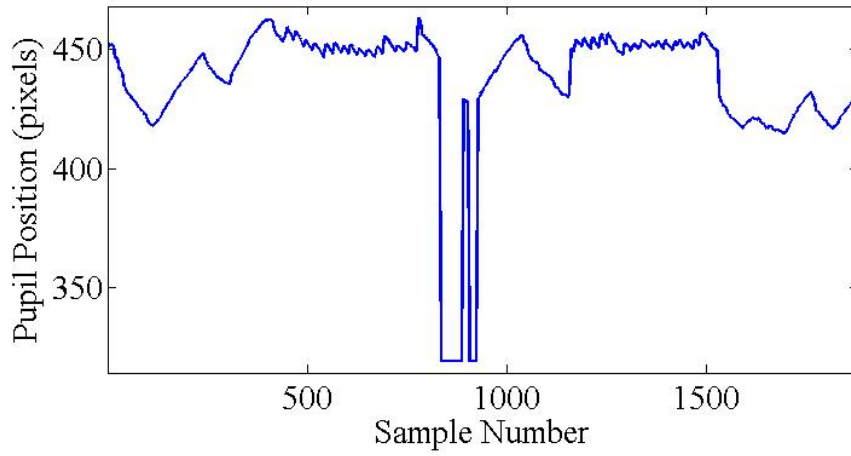


Figure E.21: Mixed Signal 5

Figure E.22: Mixed Signal 6

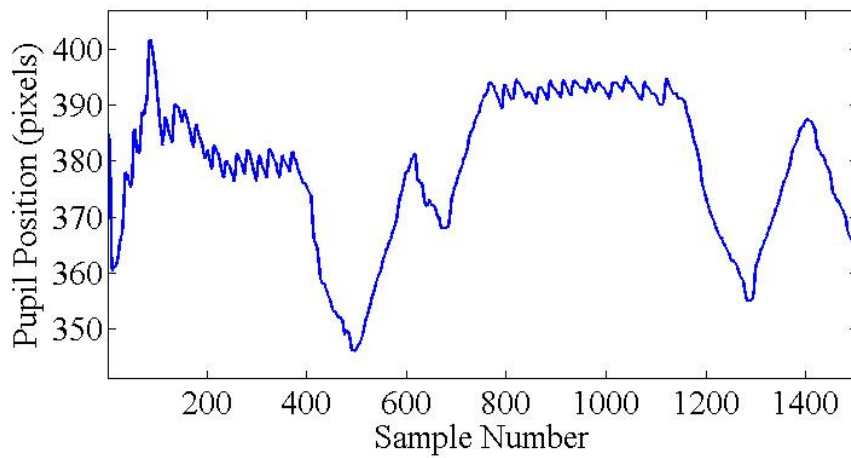


Figure E.23: Mixed Signal 7

APPENDIX E. TEST DATA

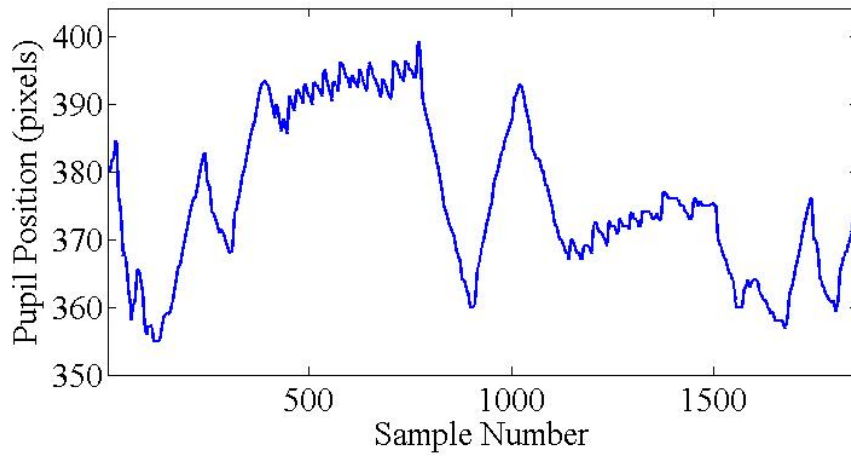


Figure E.24: Mixed Signal 8

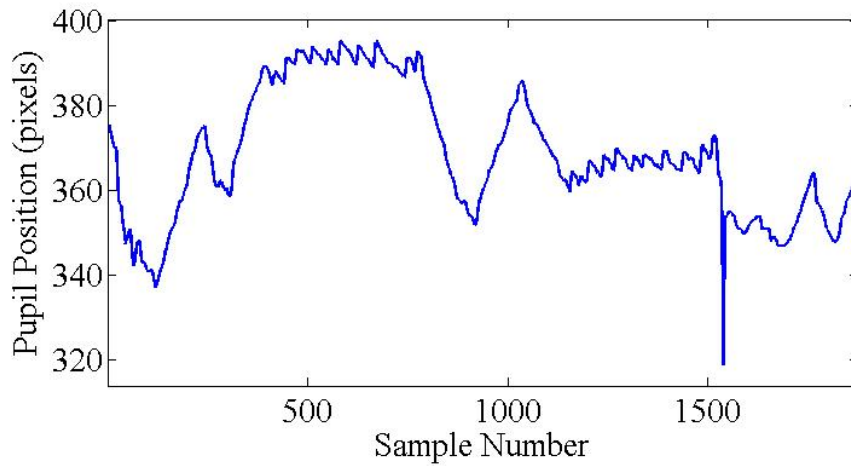


Figure E.25: Mixed Signal 9

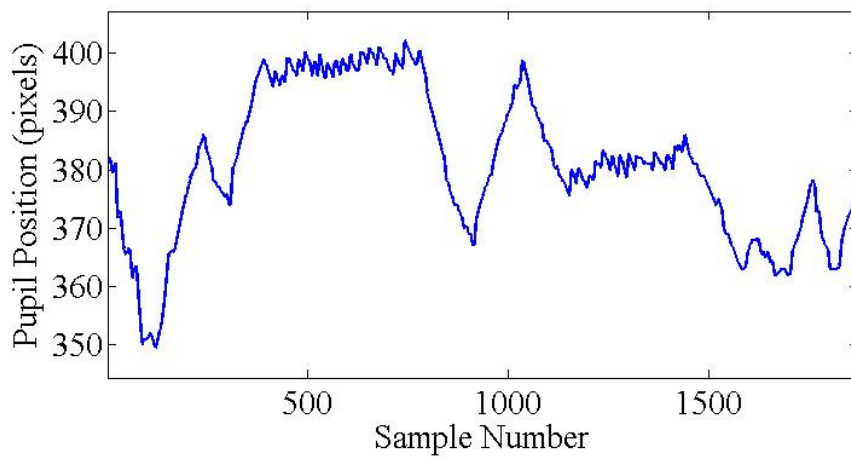


Figure E.26: Mixed Signal 10

APPENDIX E. TEST DATA

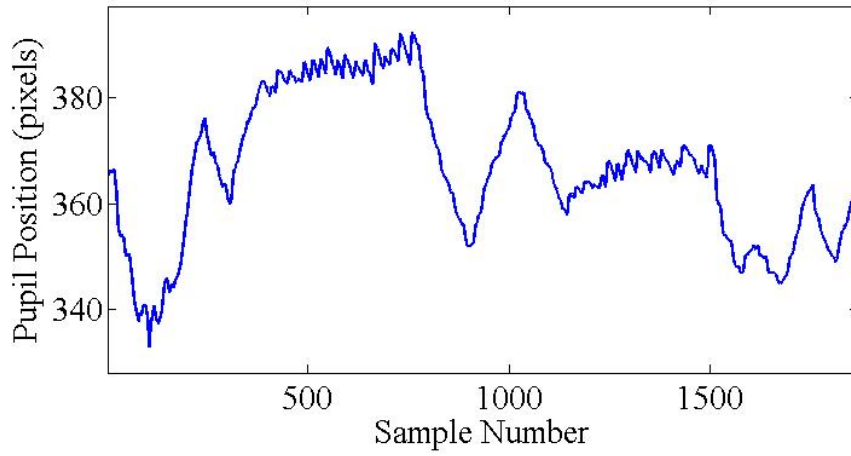


Figure E.27: Mixed Signal 11

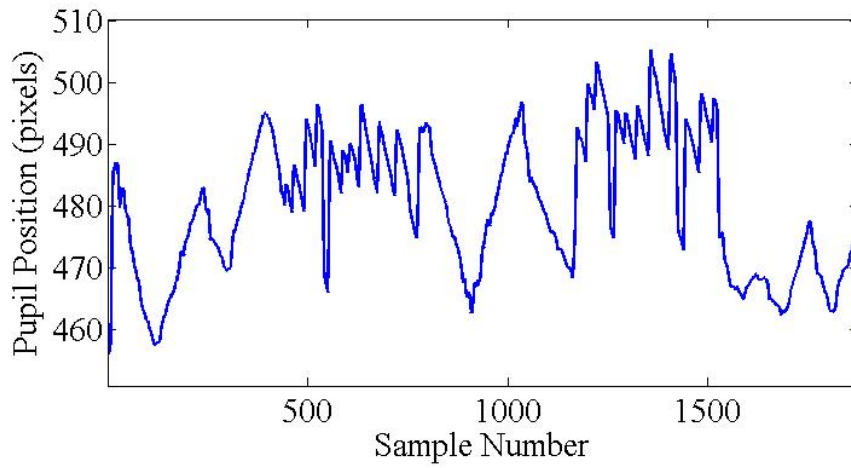


Figure E.28: Mixed Signal 12

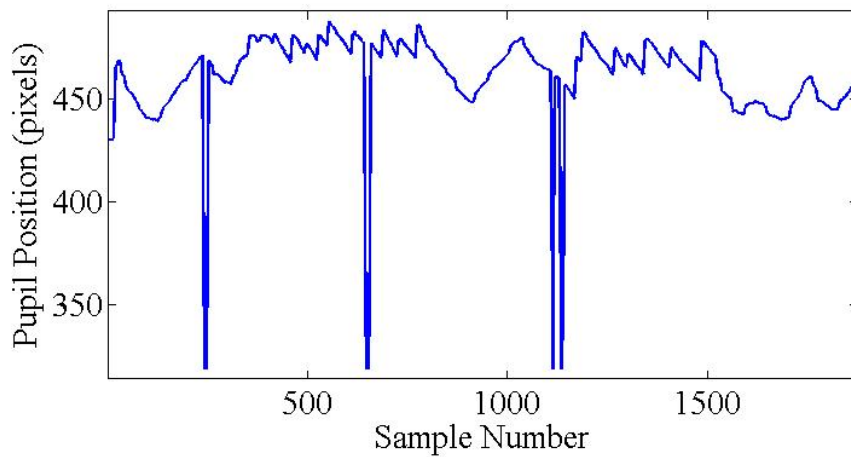


Figure E.29: Mixed Signal 13

APPENDIX E. TEST DATA

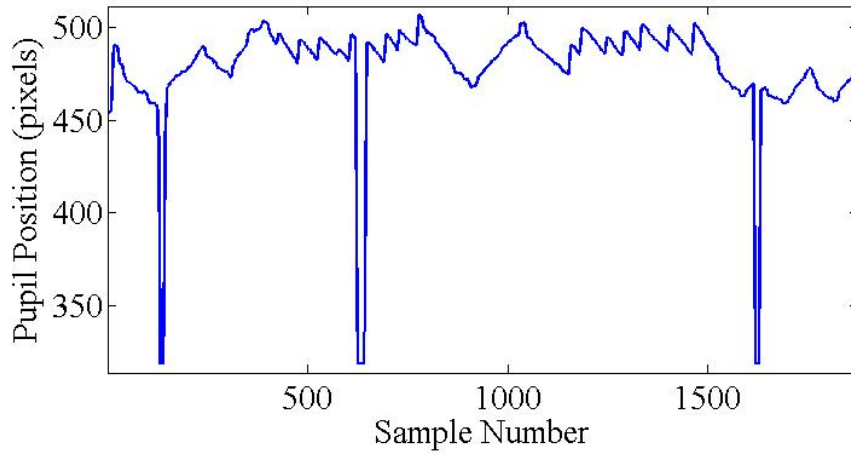


Figure E.30: Mixed Signal 14

Appendix F

Software Development

This chapter presents more information about the developed software concerning user interface and code structure. Since wavelet analysis could not be incorporated into a C++ application within the thesis time and scope constraints and consequently the image capture, image processing and signal processing were developed into different applications.

F.1 Image Capture and User Interface

Since PS3Eye cameras were used to capture the images, the capture application needed to be modified accordingly. The first requirement that derived from the camera choice is the camera driver. Since the camera was not originally designed for PC use, a special driver pack is needed. *Code Laboratories* provides these drivers that supports the with a *MS Windows* OS. In order to communicate with the camera, Code Laboratories have also developed the necessary C++ libraries. Communication with the cameras can also be implemented with *OpenCV* but only one camera is supported at any given instance. Thus, in order to enable stereo recordings, libraries supplied *Code Laboratories* were implemented. A GUI application was developed to provide a simply, user friendly control tool for video capture.

F.1.1 Image Capture

Code Laboratories provide a camera class as part of the PS3Eye libraries. With this class, multiple camera instances can be created and controlled. The standard class was modified to adhere to the requirements for the user interface that was developed for the purpose of this thesis. At application startup all PS3Eye cameras are identified and a camera instance is created for each of the identified cameras. Each of the identified cameras are listed in a message box in the application. If no cameras are found, the user is informed in the same message box.

Through the camera class the application can activate the periodic capture of raw images from the image sensor. Based on application settings the raw image is either processed or written to a video file. This will be discussed in more detail in Section F.1.2.

F.1.2 User Interface

Figure F.1 shows the layout of the user interface. In this case, no cameras were found at application startup, hence the message in the message box. If the *Connect* button is clicked after no cameras were found, the application will search for cameras again. If cameras were already identified, the application starts a camera instance for each identified camera.

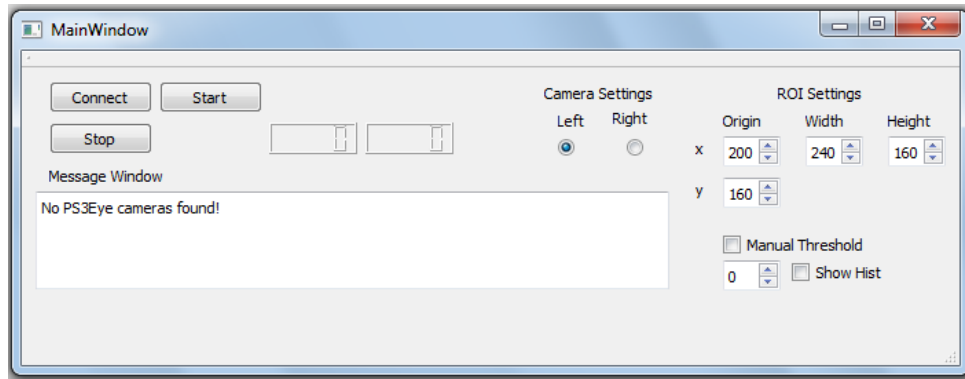


Figure F.1: Graphical User Interface with No PS3Eye Cameras Connected

Figure F.2 shows the message window of the user interface reporting two PS3Eye cameras connected to the PC. For each camera the message window reports the that the camera has been detected and displays the identification code (GUID) of the camera. Every camera manufactured has a unique GUID.

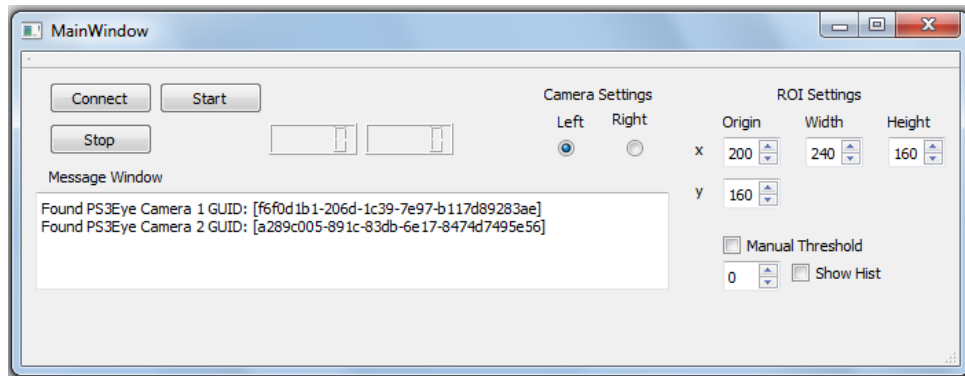


Figure F.2: Graphical User Interface with Two PS3Eye Cameras Detected but Unconnected

Based on an application setting (currently hard-coded) the *Start* and *Stop* buttons either control the histogram tracking operation or video capture function. If histogram tracking is activated, the raw image that is periodically captured by the camera class is converted to an *OpenCV* image and processed according to GUI settings. The result is displayed in a window in order to provide visual feedback for the user. The estimated pupil center coordinate is then recorded to a data file as raw data. If the application is set to record video files, a *cvVideoWriter* is created on the *Start* button click and

APPENDIX F. SOFTWARE DEVELOPMENT

released at the *Stop* button clicked. Between these events the raw image captured by the camera class is converted into an *OpenCV* image and written to the specified video file.

Figure F.3 shows the user interface after connection to the cameras have been established. The windows below the interface displays each frame as it is captured. In the case the histogram tracking is activated the estimated pupil position is also displayed with a white cross. The apparent difference in resolution between the images is the result of the different lenses that were used. Recall from Section 3.4 that the standard PS3Eye lens was replaced to improve the resolution of the captured images.

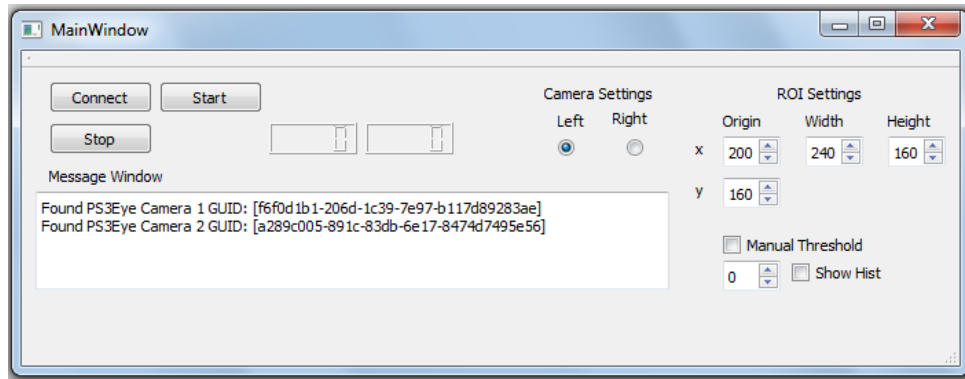


Figure F.3: Graphical User Interface with Two PS3Eye Cameras Detected and Connected

Note the four dials in the top right hand corner of the user interface marked *ROI Settings*. These Region of Interest (ROI) integer dials allow the user to set the ROI for the image for histogram processing purposes by specifying the origin coordinate and image width and height. The ROI can be set separately for each camera by using the radio buttons at the top center of the user interface labeled *Camera Settings*. These radio buttons also applies to the threshold settings that are discussed in the following paragraph. The reason for including the ROI settings is to lighten the computational load for high speed (125 Hz) stereo tracking purposes. Currently, the capture rate of the cameras are controlled by hard-coded settings in the application.

The controls at the bottom right corner of the user interface help the user to define a custom threshold value. If the tick box labeled *Manual Threshold* is checked, the application omits the automatic threshold detection routine described in Section 4.3 and illustrated in Figure 4.17. The threshold value is then read directly from the integer dial below the check box. The check box labeled *Show Hist* enables the user to open a new window that the shows the threshold image of the frame captured by the camera. The pupil the appears as a white blob with a black background.

F.2 Tracking Software

The signal analysis software was also developed in C++ and can therefore easily be included into an application containing the user interface and camera class files that was discussed in the previous section. Currently the processing application simply creates a video reader which is used to read the frames from the recorded VOG video. The

application can then be set to process a certain frame range or all of the frames from the video. For every frame that gets processed, the application grabs the grey image from the captured video and creates a colour frame with the same dimensions and attributes such as depth. These two frames are then fed to the frame process function which implements the proposed tracking algorithm discussed in detail in Section 4.3. The frame process function then estimates the pupil center and records it to a raw data file. Depending on hard-coded settings the colour frame is then modified to contain the estimated ellipse or pupil center or both.

F.3 Signal Analysis Software

The signal analysis software was implemented in a Matlab program since wavelet analysis could not be implemented in C++. The Matlab program reads the recorded pupil positions from the raw data files and converts the data into the vertical and horizontal motion arrays. These arrays are then processed in the manner that was discussed in detail in Section 5.3. The output of this program is the signal with the identified nystagmus beats highlighted as shown in Figures 5.12.b and 5.13.b. These figures were analyzed manually to gather the data shown in Tables 5.2 and 5.3.

List of References

- Abel, L., Parker, L., Daroff, R. and Dell'Osso, L. (1978). End-Point nystagmus. *Investigative Ophthalmology*, vol. 17, pp. 539–544.
- Abel, L., Wang, Z. and Dell'Osso, L. (2008). Wavelet Analysis in Infantile Nystagmus Syndrome: Limitations and Abilities. *Investigative Ophthalmology Visual Science*, vol. 49, no. 8, pp. 3413–3423.
- Allison, R., Eizenman, M. and Cheung, B. (1996). 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.
- Barea, R., Boquete, L., Mazo, M., López, E. and Bergasa, L. (2003). Electrooculographic Guidance of a Wheelchair Using Eye Movements Codification. *The International Journal of Robotics*, vol. 22, no. 7-8, pp. 641–652.
- Bashshur, R., Sanders, J. and Shannon, G. (1997). *Telemedicine: Theory and Practice*. 1st edn. Charles C Thomas Publisher.
- Böhme, M., Meyer, A., Martinetz, T. and Barth, E. (2006). Remote eye tracking: State of the art and directions for future development. In *Proceedings of the 2006 Conference on Communication by Gaze Interaction (COGAIN)*, pages 12-17, 2006.
- Clarke, A. (1998). Vestibulo-Oculomotor Research and Measurements Technology for the Space Station Era. *Brain Research Reviews*, vol. 28, no. 1, pp. 173–184.
- Clarke, A. (2004). Chronos eye tracking device - user manual. [Online.] Available: <http://www.eidactics.com/Downloads/Refs-Methods/Chronos EyeTracking Device UserManual 2006.pdf>. [2011, June 24].
- Coetzee, C. (2006). *Finite Element Methods*. Finite Element Method 414 Subject Notes, University of Stellenbosch (2006).
- Cogan, D. (1967). Congenital Nystagmus. *Canadian Journal of Ophthalmology*, vol. 2, no. 1, pp. 4–10.
- Cogan, D.G. In Haik, G. (1962). *Strabismus. Symposium of the New Orleans Academy of Ophthalmology*. Mosby-Year Book.
- Collewijn, H. (1977). Eye- and Head Movements in Freely Moving Rabbits. *Journal of Physiology*, vol. 266, no. 2, pp. 471–498.
- Collewijn, H., van der Mark, F. and Jansen, T. (1975). Precise Recording of Human Eye Movements. *Vision Research*, vol. 15, no. 3, pp. 447–450.

APPENDIX F. SOFTWARE DEVELOPMENT

- Collewijn, H., Van der Steen, J., Ferman, L. and Jansen, T. (1985). Human ocular counterroll: assessment of static and dynamic properties from electromagnetic scleral coil recordings. *Experimental Brain Research*, vol. 59, no. 1, pp. 185–196.
- Daroff, R., Dell’Osso, L.F. In Thompson, H., Daroff, R., Frisén, L., Glaser, J. and Sanders, M. (1979). *Topics in Neuro-Ophthalmology*. Williams and Wilkins.
- Daunys, G., Böhme, M., Droege, D., Villanueva, A., Delbrück, T., Hansen, D., Stepankova, O., Ramanauskas, N. and Kumpys, L. (2007). *D5.3 Eye Tracking Hardware Issues*. [Online]. Available: <http://www.cogain.org/results/reports/COGAIN-D5.3.pdf>. [2011, July 4].
- Dell’Osso, L. (1973). Fixation Characteristics in Hereditary Congenital Nystagmus. *American Journal of Optometry and American Academy of Optometry*, vol. 50, no. 2, pp. 85–90.
- Dell’Osso, L. (1989). Nystagmus, Saccadic Intrusions/Oscillations, and Oscillopsia. *Curr Neuro Ophthalmol*, vol. 1, no. 2, pp. 147–182.
- Dell’Osso, L. and Daroff, R. (1975). Congenital Nystagmus Waveforms and Foveation Strategy. *Documenta Ophthalmologica*, vol. 39, no. 1, pp. 155–182.
- Dell’Osso, L., Flynn, J. and Daroff, R. (1974). Hereditary Congenital Nystagmus: An Intrafamilial Study. *Archives of Ophthalmology*, vol. 92, no. 5, pp. 366–374.
- Dell’Osso, L., Hertle, R. and Daroff, R. (2007). Sensory and Motor Nystagmus: Erroneous and Misleading Terminology Based on Misinterpretation of David Cogan’s Observations. *Archives of Ophthalmology*, vol. 125, no. 11, pp. 1559–1561.
- Dongheng, L., Winfield, D. and Parkhurst, D. (2005). Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR05 Workshops.
- Donoho, D. (1995). De-Noising by Soft-Thresholding. *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627.
- Duda, R. and Hart, P. (1972). Use of the Hough Transformation to Detect Lines and Curves in Pictures. *Communications of the ACM*, vol. 15, no. 1, pp. 11–15.
- Enderle, J., Blanchard, S. and Bronzino, J. (2005). *Introduction to Biomedical Engineering*. 2nd edn. Elsevier.
- Fischler, M. and Bolles, R. (1982). Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, vol. 24, no. 6, pp. 381–395.
- Fox, S. (2009). *Fundamentals of Human Physiology*. 1st edn. McGraw-Hill.
- Gans, R. (2001). Video-oculography: A new diagnostic technology for vestibular patients. *The Hearing Journal*, vol. 54, no. 5, pp. 40–42.
- Goneid, A., El-Gindi, S. and Sewisy, A. (1997). A Method for the Hough Transform Detection of Circles and Ellipses Using a 1-Dimensional Array. *Systems, Man, and Cybernetics, 1997. ‘Computational Cybernetics and Simulation’.*, 1997 IEEE International Conference, vol. 4, no. 10, pp. 3154–3157.
- Gonzales, R. and Woods, R. (2008). *Digital Image Processing*. 3rd edn. Pearson Education.

APPENDIX F. SOFTWARE DEVELOPMENT

- Good, G. and Augsburger, A. (1986). Use of horizontal gaze nystagmus as part of roadside sobriety testing. *American Journal of Optometry and Physiological Optics*, vol. 63, no. 6, pp. 467–471.
- Gresty, M., Bronstein, A., Page, N. and Rudge, P. (1991). Congenital-Type Nystagmus Emerging in Later Life. *Neurology*, vol. 41, no. 5, pp. 653–656.
- Gresty, M., Page, N. and Barratt, H. (1984). The differential diagnosis of congenital nystagmus. *Journal of Neurology, Neurosurgery, and Psychiatry*, vol. 47, no. 9, pp. 936–942.
- Guestrin, E. and Eizenman, M. (2006). General Theory of Remote Gaze Estimation Using the Pupil Center and Corneal Reflections. *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 6, pp. 1124–1133.
- Halíř, R. and Fusser, J. (1998). Numerically stable direct least squares fitting of ellipses. *WSCG*.
- Ham, W., Mueller, H. and Sliney, D. (1976). Retinal sensitivity to damage by short-wavelength light. *Nature*, vol. 260, pp. 153–155.
- Hansen, D., Hansen, J., Nielsen, M., Johanson, A. and Stegmann, M. (2002). Eye Typing using Markov and Active Appearance Models. *Applications of Computer Vision 2002. IEEE Workshop on Applications on Computer Vision*, pp. 132–136.
- Houben, M., Goumans, J. and van den Steen, J. (2006). Recording Three-Dimensional Eye Movements: Scleral Search Coils versus Video Oculography. *Investigative Ophthalmology Visual Science*, vol. 47, no. 1, pp. 179–187.
- Houston, H. and Watson, D. (1994). A Review of Computerized Electronystagmography Technology. *British Journal of Audiology*, vol. 28, no. 1, pp. 41–46.
- Hutchinson, T., White, K.P., J., Martin, W., Reichert, K. and Frey, L. (1989). Human-Computer Interaction Using Eye-Gaze Input. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 19, no. 6, pp. 1527–1534.
- IEC (2007). *Safety of Laser Products - Part 1: Equipment Classification and Requirements*. IEC 60825-1 International Standard.
- Imai, T., Sekine, K., Hattori, K., Takeda, N., Koizuka, I., Nakamae, K., Miura, K., Fujioka, H. and Kubo, T. (2005). Comparing the accuracy of video-oculography and the scleral search coil system in human eye movement analysis. *Auris Nasus Larynx*, vol. 32, no. 1, pp. 3–9.
- Imai, T., Takeda, N., Morita, M., Koizuka, I., Kubo, T., Miura, K., Nakamae, K. and Fujioka, H. (1999). Rotation Vector Analysis of Eye Movement in Three Dimensions with Infrared CCD Camera. *Acta Oto-Laryngologica*, vol. 119, no. 1, pp. 24–28.
- Jansen, S., Kingma, H. and Peeters, R. (2010). A Wavelet Approach for Unsupervised Nystagmus Analysis on ENG and VOG Recordings. *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pp. 6333–6336.
- Kenyon, R. (1985). A Soft Contact Lens Search Coil for Measuring Eye Movements. *Vision Research*, vol. 25, no. 11, pp. 1629–1633.
- Koekemoer, D., M. (2011). Medical doctor. personal interview. 16 March, Stellenbosch.
- Kumar, N., Kohlbecher, S. and Schneider, E. (2009). A novel approach to video-based pupil tracking. *IEEE Conference: Systems, Man and Cybernetics, 2009*.

APPENDIX F. SOFTWARE DEVELOPMENT

- Liu, C., Chaovalitwongse, W., Pardalos, P., Seref, O. Xanthopoulos, P., Sackellares, J. and Skidmore, F. (2007). Quantitative analysis on electrooculography (EOG) for neurodegenerative disease. *AIP Conference Proceedings*, vol. 953, no. 1, pp. 246–253.
- Matsumoto, Y. and Zelinsky, A. (2000). An algorithm for real-time stereo vision implementation of head pose and gaze direction measurement. *Proc. 4th IEEE Int. Conf. Automatic Face Gesture Recognition*, pp. 499–504.
- Matthes, R. (2000). ICNIRP Statement on Light-Emitting Diodes (LEDs) and Laser Diodes: Implications for Hazard Assessment. *Health Physics*, vol. 78, no. 6, pp. 744–752.
- Merchant, J., Morrisette, R. and Porterfield, J. (1974). Remote measurement of eye direction allowing subject motion over one cubic foot of space. *IEEE Transactions on Biomedical Engineering*, vol. BME-21, no. 4, pp. 309–317.
- Milekic, S. (2004). *The More You Look The More You Get: Intention-based Interface Using Gaze-tracking*. [Online]. Available: <http://www.archimuse.com/mw2003/papers/milekic/milekic.html>. [2011, July 4].
- Moore, S., Curthoys, I. and McCoy, S. (1991). VTM - an image-processing system for measuring ocular torsion. *Computer Methods and Programs in Biomedicine*, vol. 35, no. 3, pp. 219–230.
- Moore, S., Halswanter, T., Curthoys, I. and Smith, S. (1996). A Geometric Basis for Measurement of Three-Dimensional Eye Position Using Image Processing. *Vision Research*, vol. 36, no. 3, pp. 445–459.
- Mulvey, F., Villanueva, A., Sliney, D., Lange, R., Cotmore, S. and Donegan, M. (2008). *D5.3 Eye Tracking Hardware Issues*. [Online]. Available: <http://www.cogain.org/w/images/c/c8/COGAIN-D5.4.pdf>. [2011, July 5].
- Nacci, P., Turner, C., Waldron, R. and Broyels, E. (2002). Implementing Telemedicine in Correctional Facilities. *U.S. Department of Justice Documentation*.
- Nason, G. and Silverman, B. (1995). *The Stationary Wavelet Transform and some Statistical Applications*. Lecture Notes in Statistics, Department of Mathematics, University of Bristol.
- Newman, R., Matsumoto, Y., Rougeaux, S. and Zelinsky, A. (2000). Real-time stereo tracking for head pose and gaze estimation. *Proc. 4th IEEE Int. Conf. Automatic Face Gesture Recognition*, pp. 122–128.
- Orssaud, C. (2003). *Congenital Nystagmus*. Orphanet Encyclopedia. [Online]. Available: <http://www.orpha.net/data/patho/GB/uk-nystagmus.pdf>.
- Reulen, J., Marcus, J., Koops, D., de Vries, F., Tiesinga, G., Boshuizen, K. and Bos, J. (1988). Precise Recording of Eye Movement: the IRIS Technique Part 1. *Medical and Biological Engineering and Computing*, vol. 26, no. 1, pp. 20–26.
- Rey, C. and Galiana, H. (1991). Parametric Classification of Segments in Ocular Nystagmus. *IEEE Transactions on Biomedical Engineering*, vol. 38, no. 2, pp. 142–148.
- Robinson, D. (1963). A method of measuring eye movement using a scleral search coil in a magnetic field. *Institute of Electrical and Electronics Engineers, Transactions on Bio-Medical Electronics*, vol. BME-10, no. 1, pp. 137–145.
- RSA-MIL (1994). *Ergonomic Design: Anthropometry and Environment*. RSA-MIL-STD-127 : Volume 1.

APPENDIX F. SOFTWARE DEVELOPMENT

- Shelhamer, M. (1998). *Contact lens and wire coil on subject's eye*. [Online]. Available: <http://mvl.mit.edu/Research/Neurovestibular/Pages/project1.html>.
- Shinomiya, K., Itsuki, N., Kubo, M. and Shiota, H. (2008). Analyses of the characteristics of potential and cross-talk at each electrode in electro-oculogram. *The Journal of Medical Investigation*, vol. 55, no. 1-2, pp. 120–126.
- Sliney, D. (2007). *Lasers and Optical Radiation Hazards in Laboratory-understanding the biological effects*. [Online]. Available: <http://www.boulder.nist.gov/div815/lmsc/presentations/Laser-Optical Hazards RD Labs-for DSLINEY use.pdf>. [2011, July 5].
- Stewart, J. (2003). *Calculus*. 5th edn. Thomson Brooks/Cole.
- Stowe, G. (2007). *Interacoustics VNG system VO425*. [Online]. Available: http://www.gordonstowe.com/products/vng_systems.htm. [2011, July 4].
- Sugioka, A., Ebisawa, Y. and Ohtani, M. (1996). Noncontact video-based eye-gaze detection methods allowing large head displacements. *Proc. 18th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 2, pp. 526–528.
- Thomas, G., Weir, M., Hass, J. and Giordano, F. (2005). *Thomas's Calculus*. 11th edn. Pearson Education.
- Trikha, M., Bhandari, A. and Gandhi, T. (2007). Automatic electrooculogram classification for microcontroller based interface design. *2007 IEEE Systems and Information Engineering Design Symposium*.
- Von Noorden, G. and Campos, E. (2002). *Binocular Vision and Ocular Motility: Theory and Management of Strabismus*. 6th edn. Mosby.
- Wang, Y., Wang, H., Liu, L., Fang, Y. and Zhang, J. (2010). A New Diagnosis Method for Heart Valve Disease by Using Antibody Memory Clone Clustering Algorithm Based on Supervised Garth-Geva Algorithm. *Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference*, pp. 150–154.
- Young, L. and Sheena, D. (1975). Methods Designs: Survey of eye movement recording methods. *Behaviour Research Methods Instrumentation*, vol. 7, no. 5, pp. 397–429.
- Yu, L. and Eizenman, M. (2004). A new methodology for determining point-of-gaze in head-mounted eye tracking systems. *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 10, pp. 1765–1773.
- Zifkin, B. and Trenité, D. (2000). Reflex epilepsy and reflex seizures of the visual system: a clinical review. *Epileptic Disorders*, vol. 2, no. 3, pp. 129–136.
- Zill, G. and Cullen, R. (2000). *Advanced Engineering Mathematics*. 2nd edn. Jones and Bartlett.