Doctoral Thesis

Eye movement analysis for context inference and cognitive-awareness
wearable sensing and activity recognition using
electrooculography

Author[s]:
Bulling, Andreas

Publication Date:
2010

Permanent Link:
https://doi.org/10.3929/ethz-a-006208589

Rights / License:
In Copyright - Non-Commercial Use Permitted

This page was generated automatically upon download from the ETH Zurich Research Collection. For more information please consult the Terms of use.
Eye Movement Analysis for Context Inference and Cognitive-Awareness: Wearable Sensing and Activity Recognition Using Electrooculography

A dissertation submitted to
ETH ZURICH
for the degree of
Doctor of Sciences

presented by
ANDREAS BULLING
Dipl.-Inform., Technical University of Karlsruhe, Germany
date of birth 14 February 1981
citizen of Germany

Prof. Gerhard Tröster, ETH Zurich, examiner
Prof. Peter Robinson, University of Cambridge, co-examiner

2010
To my father
# Table of Contents

Abstract xi

Zusammenfassung xiii

1. Introduction 1
   1.1. Context-Awareness and Activity Recognition 2
   1.2. Eye Movements as a Modality for Context-Awareness 2
   1.3. Eye Movement Analysis 4
   1.4. Eye Tracking and Applications in Human-Computer Interaction 4
   1.5. Aims of the Work 5
   1.6. Thesis Outline 7
   1.7. Additional Publications 10

Bibliography 16

2. Thesis Summary 17
   2.1. Summary of Contributions 18
      2.1.1. Development of a Wearable Eye Movement Sensor 18
      2.1.2. HCI in Stationary Settings Using Relative Eye Movements Detected From EOG 19
      2.1.3. Development of an Eye-Based Activity Recognition Architecture 21
      2.1.4. Recognition of Daily Life Activities From Eye Movements 23
      2.1.5. Towards Cognitive-Awareness Using Eye Movement Analysis 25
   2.2. Conclusion 27
   2.3. Limitations and Relevance 29
   2.4. Outlook 30

3. Wearable Eye Movement Sensing 31
   3.1. Introduction 32
      3.1.1. Contributions 33
      3.1.2. Paper Organisation 33
   3.2. Eye Movements: Physiology, Sensors and Applications 34
      3.2.1. Physiology of Eye Motion 34
      3.2.2. Eye Movements in Daily Life 34
      3.2.3. Video-Based Eye Tracking 35
      3.2.4. EOG-Based Eye Tracking 36
3.2.5. Summary ........................................... 37
3.3. Design and Implementation of a Wearable Context-Aware EOG Sensor ............................................. 38
  3.3.1. Requirements ........................................ 38
  3.3.2. Challenges .......................................... 39
  3.3.3. Hardware Design ...................................... 39
  3.3.4. Firmware ............................................ 42
3.4. EOG Signal Processing for Context Sensing .......... 44
  3.4.1. Adaptive Filtering ..................................... 44
  3.4.2. Blink Detection ....................................... 47
  3.4.3. Saccade Detection ..................................... 49
  3.4.4. Blink Removal ......................................... 49
  3.4.5. Eye Gesture Recognition ............................. 50
3.5. Case Studies ........................................... 51
  3.5.1. Activity Recognition ................................... 51
  3.5.2. Human-Computer Interaction ........................ 53
3.6. Discussion ................................................ 53
  3.6.1. On Wearable Eye Tracking Using EOG ............ 53
  3.6.2. On the Case Studies .................................. 55
3.7. Conclusion and Outlook .................................. 55
Bibliography .................................................. 61

4. Relative Eye Movements for Human-Computer Interaction 63
4.1. Introduction ............................................. 64
4.2. Related Work ............................................ 65
  4.2.1. Activity Recognition ................................. 65
  4.2.2. Eye-based Interaction ................................ 65
  4.2.3. EOG-based Interfaces ................................. 66
  4.2.4. Eye Tracking Devices ................................. 66
4.3. Wearable Electrooculography ............................ 66
  4.3.1. Eye Movement Characteristics ....................... 66
  4.3.2. Design and System Architecture ..................... 67
4.4. EOG Signal Processing .................................. 69
  4.4.1. Blink Detection ...................................... 70
  4.4.2. Saccade Detection .................................... 70
  4.4.3. Blink Removal ........................................ 70
  4.4.4. Eye Gesture Recognition ............................. 71
  4.4.5. Artefact Compensation ............................... 72
4.5. Experiment I - Eye Gestures for Stationary HCI .... 73
  4.5.1. Setup ................................................ 73
  4.5.2. Experimental Procedure .............................. 74
  4.5.3. Results ............................................... 76
4.6. Experiment II - Eye Movements for Mobile HCI ...... 78
  4.6.1. Setup ................................................ 78
  4.6.2. Experimental Procedure .............................. 79
5. **Eye Movements for Context-Aware Gaming**

5.1. **Introduction** .................................................. 88
5.2. **Related Work** .................................................. 89
  5.2.1. Eye-based Human-Computer Interaction ................. 89
  5.2.2. Eye Movements and Games ................................. 89
  5.2.3. EOG-based Interfaces ...................................... 89
5.3. **Wearable Electrooculography** ............................ 89
  5.3.1. Eye Movement Characteristics ............................ 89
  5.3.2. EOG Data Recording ...................................... 90
  5.3.3. EOG Signal Processing .................................... 90
5.4. **Application Scenarios** .................................... 92
  5.4.1. Stationary Games .......................................... 92
  5.4.2. Pervasive Games .......................................... 95
  5.4.3. Future Context-Aware Games ............................ 99
5.5. **Discussion and Conclusion** .............................. 100

Bibliography .......................................................... 104

6. **Recognition of Reading Activity in Transit**

6.1. **Introduction** .................................................. 106
  6.1.1. **Related Work** .......................................... 107
6.2. **Eye Movement Analysis** .................................... 109
  6.2.1. Wearable Electrooculography ............................ 109
  6.2.2. Eye Movement Characteristics ............................ 109
  6.2.3. Baseline Drift ............................................. 110
6.3. **Data Collection** ............................................. 111
  6.3.1. Experimental Procedure .................................. 111
6.4. **Continuous Recognition Methods** ....................... 115
  6.4.1. String Matching ........................................... 116
  6.4.2. Hidden Markov Models .................................... 117
6.5. **Results** ...................................................... 118
  6.5.1. Parameter Selection and Training ....................... 118
  6.5.2. Continuous Classification ................................ 119
6.6. **Discussion** .................................................. 122
  6.6.1. Conclusion ................................................ 125

Bibliography .......................................................... 129
7. Recognition of Office Activities

7.1. Introduction ............................................. 132
  7.1.1. Paper Scope and Contributions ...................... 133
  7.1.2. Paper Organisation .................................. 133

7.2. Related Work ............................................ 134
  7.2.1. Electrooculography Applications ...................... 134
  7.2.2. Eye Movement Analysis ................................. 134
  7.2.3. Activity Recognition ................................ 135

7.3. Background .............................................. 135
  7.3.1. Electrooculography .................................. 135
  7.3.2. Eye Movement Types ................................. 135

7.4. Methodology .............................................. 137
  7.4.1. Recognition Architecture .............................. 137
  7.4.2. EOG Signal Processing ................................. 138
  7.4.3. Detection of Basic Eye Movement Types ............... 139
  7.4.4. Analysis of Repetitive Eye Movement Patterns ....... 143
  7.4.5. Feature Extraction ................................... 145
  7.4.6. Feature Selection and Classification ............... 147

7.5. Experiment .............................................. 147
  7.5.1. Apparatus .......................................... 148
  7.5.2. Procedure ......................................... 149
  7.5.3. Parameter Selection and Evaluation .................. 149

7.6. Results .................................................. 149
  7.6.1. Classification Performance ......................... 149
  7.6.2. Eye Movement Features ............................... 150

7.7. Discussion .............................................. 153
  7.7.1. Robustness Across Participants ...................... 153
  7.7.2. Results for Each Activity ............................ 153
  7.7.3. Feature Groups ..................................... 154
  7.7.4. Features for Each Activity Class .................... 154
  7.7.5. Activity Segmentation Using Eye Movements ........ 155
  7.7.6. Limitations ....................................... 156
  7.7.7. Considerations for Future Work ..................... 156

7.8. Conclusion .............................................. 157

Bibliography .................................................. 164

8. What’s in the Eyes for Context-Awareness? 165

8.1. Introduction ............................................ 166
8.2. Tracking Eye Movements ................................. 167
8.3. Eye Movement Analysis .................................. 167
8.4. Context Recognition Using Eye Movement Analysis .... 168
8.5. Case Study I: Recognition of Reading .................. 168
  8.5.1. Lessons Learned .................................... 169
8.6. Case Study II: Recognition of Office Activities .... 169
  8.6.1. Lessons Learned .................................... 170
Abstract

Context-awareness has emerged as a key area of research in mobile and ubiquitous computing. Context-aware systems aim to sense the user’s context and use this information to proactively adapt their behaviour to the user’s personal needs. The context of a person is usually defined as a combination of different personal and environmental factors. Among these factors, physical activity is widely considered to be one of the most important contextual cues. Accordingly, machine recognition of human physical activity has attracted considerable research interests in recent years.

Advances in activity recognition were traditionally achieved by using modalities such as body posture and gestures, interactions between people, or by combining on-body and ambient sensors. Despite these advances, activity recognition systems that use such modalities are limited with respect to the types and complexity of activities they are able to detect. In addition, activity recognition using more subtle cues, such as user attention or intention, or recognition that takes into account cognitive aspects of a user’s activity, such as the experience or task engagement, remains largely unexplored.

The current work introduces eye movement analysis as a novel modality for context inference and cognitive-awareness. The movements our eyes perform as we carry out different activities reveal much about the activities themselves. In the same manner, location or a particular environment influence our eye movements. Finally, unconscious eye movements are strongly linked to a number of cognitive processes of visual perception. This link to cognition makes eye movements a distinct source of information on a person’s context beyond physical activity and location. Eventually, information derived from eye movements may allow us to extend the current notion of context with a cognitive dimension, leading to so-called cognitive-aware systems.

This thesis comprises six scientific publications that address five aims of the work: (1) to introduce eye movement analysis as a modality for context-awareness and activity recognition as well as a trailblazer for cognitive-aware systems; (2) to develop a wearable and self-contained eye tracker for long-term eye movement recordings based on electrooculography (EOG); (3) to investigate eye gestures continuously detected from EOG signals for real-time eye-based human-computer interaction; (4) to develop an architecture for eye-based activity recognition comprising eye movement features and algorithms for EOG signal processing and eye movement analysis; and (5) to recognise daily life activities from eye movement data in natural mobile and stationary settings using this architecture.

This work developed the so-called wearable EOG goggles, an embedded eye tracker that - in contrast to common systems using video - relies on EOG. Challenges associated with wearability, eye movement analysis, and signal artefacts caused by physical activity were addressed with a combination of a light-weight mechanical design, robust algorithms for eye movement detection, and adaptive EOG signal processing. The final system features real-
time EOG signal processing, on-board data storage, and wireless transmission. Its low-power implementation allows for up to 7.2 hours of autonomous eye movement recordings.

The robustness and performance of the wearable EOG goggles was evaluated on the example of eye-based HCI in a multi-participant study. The thesis developed algorithms to encode eye movements detected in EOG signals into an alphabet of characters. Sequences of several consecutive movements, so-called eye gestures, can then be linked to (application-dependent) input commands. For continuous eye gesture recognition, the work adopted a string matching approach to scan the encoded eye movement sequence for eye gestures. Results from the study confirmed comparable performance to a state-of-the-art video-based eye tracker.

The eye tracker was complemented with a software architecture comprising eye movement features and algorithms specifically geared towards eye-based activity recognition. The work developed and evaluated new algorithms for detecting three different eye movement characteristics from EOG signals: saccades, fixations, and blinks. In addition, the work proposed a new method for analysing repetitive eye movement patterns using a wordbook analysis. These characteristics formed the basis for developing 90 different eye movement features that describe eye movement dynamics.

The recognition architecture was evaluated on two multi-participant activity recognition problems. A first study investigated the problem of spotting and recognising reading activity of people in transit in a daily life scenario. Reading was recognised across eight participants with a top recognition rate of 80.2% using a string matching approach. In a second study, the thesis investigated the problem of recognising typical office activities from the eye movements of eight participants in a natural work environment. Using a support vector machine classifier an average precision of 76.1% and recall of 70.5% was obtained over all classes and participants.

Finally, as a first step towards cognitive-aware systems, the thesis investigated the utility of eye movements for assessing memory recall processes of people looking at familiar and unfamiliar faces. The study revealed significant differences in eye movement features across four exposures. This result is promising for wearable eye tracking to capture eye movement characteristics that reflect memory recall processes.
Zusammenfassung


In dieser Arbeit wurde ein mobiler Eyetracker auf Basis von EOG entwickelt. Die EOG-Brille kombiniert ein leichtgewichtiges Design mit robusten Algorithmen für die adaptive EOG-Signalverarbeitung und Augenbewegungserkennung und ermöglicht die Langzeitaufnahme, Echtzeitverarbeitung und Speicherung von Augenbewegungsdaten. Das System wurde in
einer Benutzerstudie am Beispiel der augenbasierten Mensch-Maschine Interaktion mit dem Computer evaluiert. Die EOG-Brille erzielte dabei eine zu aktuellen videobasierten Systemen vergleichbare Leistungsfähigkeit.


This chapter provides an introduction to context-awareness and activity recognition, eye movement analysis, and applications of wearable eye tracking in human-computer interaction. By reviewing the state-of-the-art in each of these fields, the chapter first identifies limitations to current approaches. It then outlines the particular characteristics and potentials of wearable eye movement analysis for context-awareness and activity recognition. Finally, the chapter presents the aims and an outline of the thesis together with a list of publications that resulted from this work.
1.1. Context-Awareness and Activity Recognition

Building a common ground for humans and machines to interact with each other is a challenging problem. Context-awareness is a promising solution to this problem and has attracted significant research interests in recent years [1, 2]. Context-aware systems aim to proactively assist and adapt their behaviour to users’ personal needs by sensing the users’ context [3]. The context of a person is typically defined as a combination of different personal and environmental factors [4]. Besides location, physical activity is one of the most important contextual cues [5]. The definition of an “activity” can be broad and entirely application dependent. The goal of activity recognition is to infer activity by observing a person’s actions in daily life situations. Activity recognition has also become an important topic for a broad range of real-life applications such as personal health and lifestyle monitoring, medical rehabilitation, work safety, or human-computer interaction (HCI) in entertainment.

Research in computer vision has traditionally been at the forefront of work on activity recognition (see [6–8] for surveys). The growing use of ambient and body-worn sensors has paved the way for other sensing modalities. In ubiquitous computing and HCI, activity recognition research mostly focused on using accelerometers or gyroscopes to recognise body movement [9], hand and arm gestures [10], or posture [11]. Important advances in these fields were achieved by incorporating additional modalities such as sound [12], interactions between people [13], or object use [14, 15], and by combining wearable with ambient sensors [16].

There are, however, limitations to current activity recognition systems. Common wearable sensors are limited to sensing physical activity; they cannot easily be used for detecting predominantly visual tasks such as reading. This is despite the fact that information on a person’s reading activities can be a useful indicator of daily situations [17]. Ambient sensors typically used in activity recognition, such as reed switches or light sensors, are limited in that they only detect basic activity events such as entering or leaving a room or switching an appliance [18]. These examples illustrate the need to investigate additional sensing modalities and to further develop the recognition methods currently employed for the machine recognition of physical activities.

1.2. Eye Movements as a Modality for Context-Awareness

The current work introduces eye movement analysis as a novel sensing modality for context-awareness and activity recognition. Gaze (the “what a person is looking at”) is well-known in ubiquitous computing and HCI [19]. The less common approach is to analyse the eye movement dynamics over time (the “how a person is looking at something”). The central hypothesis of this work is that analysing these dynamics allows us to infer certain aspects of a person’s context and their activity.

The movement patterns our eyes perform as we carry out different activities are influenced by the activities themselves. This includes information on
1.2. Eye Movements as a Modality for Context-Awareness

visual tasks, such as reading, but also on many predominantly physical activities, such as driving a car. Reading is a pervasive activity, e.g. on computer screens at work, advertisements and signs in public, and books read at home or while travelling. It is comprised of a sequence of characteristic small eye movements while scanning the line and larger movements to jump back to the beginning of the next line [20]. A context-aware reading interface could provide assistance to people with reading disabilities by automatically magnifying or explaining words or context in the text or flipping pages (see [19, 21] for examples). While driving, gaze is positioned at points in the visual field that are the best for spatio-temporal demands of the task [22]. A monitoring system could analyse a driver’s eye movement patterns to assess mental states such as engagement or fatigue, or to detect distractions to prevent from potentially dangerous traffic situations. Another interesting application area for eye movement models is robotics. Models derived from people performing certain activities or interacting with each other may be used to synthesise eye movements for humanoid robots. In contrast to manually programmed eye movements, a model-based synthesis promises to generate more natural visual behaviours. This may address the so-called uncanny valley problem in research on human-robot interaction and embodied conversational agents (see [23] and [24] for details on the uncanny valley problem).

In the same manner as activities, different locations or environments influence our eye movements. Daily life situations such as being in a cinema, in the office, in a restaurant, or in a shop involve different visual tasks and therefore require specific eye movement behaviours. By analysing differences in the dynamics of these behaviours over several hours, days, or even weeks, context-aware systems may use eye movement analysis to infer these situations and provide situation-dependent support. Application areas for such systems are in home care for monitoring elderly [25, 26], lifestyle monitoring or automatic life-logging [13], or as a memory aid for people with memory impairments [27, 28].

Finally, eye movements have the particularity that they can be consciously controlled but that they are also unconsciously generated by the brain. Unconscious eye movements are strongly linked to a number of cognitive processes of visual perception such as attention [29], saliency determination [30], visual memory [31, 32], or learning [33]. This link to cognition makes eye movement analysis a distinct source of information on a person’s context beyond physical activity and location. Analysing eye movements may provide an online assessment of these underlying cognitive processes. Eventually, this may allow us to extend the current notion of context with a cognitive dimension, leading to so-called cognitive-aware systems. Such systems - based on eye movement analysis and additional modalities to analyse brain activity such as portable electroencephalography (EEG) or functional near infrared spectroscopy (fNIRs) - will enable novel types of user adaptation and interaction not possible today.
1.3. Eye Movement Analysis

Eye movements have been extensively studied in clinical ophthalmology and cognitive science. A growing number of researchers use video-based eye tracking to study the movement of the eyes in natural environments. This has led to important advances on our understanding of how the brain processes tasks, and of the role that the visual system plays in this [34]. Eye movement analysis has a long history as a tool to investigate visual behaviour. Hacisalihzade et al. used Markov processes to model visual fixations of observers recognising an object [35]. They transformed fixation sequences into character strings and used the string edit distance to quantify the similarity of eye movements. Elhelw et al. used discrete time Markov chains on sequences of temporal fixations to identify salient image features that affect the perception of visual realism [36]. They found that fixation clusters were able to uncover the features that most attract an observer’s attention. Dempere-Marco et al. presented a method for training novices in assessing tomography images [37]. They modelled the assessment behaviour of two domain experts based on the dynamics of their saccadic eye movements. Salvucci et al. evaluated means for automated analysis of eye movement protocols [38]. They described three methods - based on sequence-matching and hidden Markov models (HMM) - that were able to interpret eye movements as accurately as experts but in significantly less time.

1.4. Eye Tracking and Applications in Human-Computer Interaction

The development of eye trackers to record eye movements in daily life is still an active area of research. A wide range of hardware allows for accurate gaze tracking in stationary settings. Mobile settings call for miniaturised, low-power eye trackers with real-time processing capabilities. These requirements are increasingly addressed by commonly used video-based systems. A number of commercial eye trackers are available of which some are targeted at mobile use. These systems can now be worn as relatively light headgear, such as the Mobile Eye from Applied Science Laboratories (ASL) or the iView X HED from SensoMotoric Instruments (SMI). However, the demanding video processing requires auxiliary equipment such as laptops or digital video recorders. Without additional batteries, mobile eye movement recordings with such systems are therefore limited to a couple of hours.

Gaze recorded using video cameras has long been investigated as a means to interact with a computer. Zhai et al. proposed a new way of using eye gaze for computer input [39]. Results of an early-stage experiment indicated that their method potentially reduces physical effort and fatigue. Qvarfordt et al. explored an interactive human-computer dialogue system for touristic city trip planning [40]. They showed that it was possible to sense the users’ interests based on eye-gaze patterns and adapt the system’s output accordingly. Drewes et al. proposed to use eye gestures as an alternative to gaze-based HCI [41].
1.5. Aims of the Work

They argue that these gestures are insensitive to accuracy problems and do not exhibit the “Midas touch” problem (for details see [42]).

Efforts to miniaturise video-based eye trackers led researchers to investigate other measurement techniques such as electrooculography (EOG). Compared to video, EOG is an inexpensive method for mobile eye movement recordings. It is computationally light-weight and can be implemented using wearable sensors. This is crucial with a view to long-term recordings in mobile real-world settings. Initially, EOG signals were mostly recorded in stationary settings using standard measurement equipment [43, 44]. Later, researchers investigated novel electrode configurations for implementing EOG-based eye trackers for mobile use. While these systems allow for mobile eye movement recordings, they are not suitable for eye-based activity and context recognition. For example, Manabe et al. proposed a system that uses EOG electrode arrays mounted on ordinary headphones [45]. While this approach is less obtrusive than electrodes stuck to the face it raises other issues such as a low signal-to-noise ratio. Vehkaoja et al. presented a light-weight head cap for EOG and facial EMG measurements with electrodes embroidered of silver coated thread [46]. A small device integrated into the cap allows for wireless data transmission. EOG signal processing is performed offline on a standard desktop computer.

Eye movement sensing based on EOG has found a range of different applications, particularly in HCI. Basic eye movement characteristics detected from EOG signals such as saccades, fixations, blinks, and deliberate movement patterns have been used for hands-free operation of stationary human-computer [47, 48] and human-robot [49, 50] command interfaces. As part of a hospital alarm system, EOG-based switches provided immobile patients with a safe and reliable way of signalling an alarm [51]. In mobile settings, EOG-based interfaces have been developed for assistive robots [52] and as a control for an electric wheelchair [53, 54]. These systems are intended to be used by physically disabled people who have limited peripheral mobility but still retain eye motor coordination. Basic characteristics of eye motion were also used to operate a wearable system for medical caregivers [55].

1.5. Aims of the Work

The aim of this thesis was to investigate and evaluate eye movement analysis as a sensing modality for context-awareness and activity recognition. With a view to mobile settings, first an eye tracker for wearable eye movement recordings was developed. The work then investigated the feasibility of using eye movement analysis for recognising physical activity, and for assessing memory recall processes. Specifically, the following topics were investigated in this thesis.

**Development of a Wearable Eye Tracker**

Everyday environments call for eye trackers that provide robust long-term recordings and embedded processing for real-time recognition of human ac-
tivity and context inference. As an alternative to well-established video-based eye tracking, the current work investigated EOG for mobile eye movement recordings. The work designed and implemented a wearable, EOG-based eye tracker that allowed for long-term recordings and embedded eye movement analysis. Challenges associated with wearability and EOG signal artefacts were addressed with a special mechanical mounting and optimised algorithms for adaptive EOG signal filtering and eye movement detection.

**HCI in Stationary Settings Using Relative Eye Movements Detected from EOG**

Research in eye-based HCI mostly focuses on direct manipulation of user interfaces using video-based eye trackers. As mentioned before, mobile settings require eye movement measurement techniques that do not necessarily allow for accurate gaze tracking. EOG - the method used in this work - provides information about relative eye movements of different size in different directions. This raises several questions, particularly on how such relative eye movements - without any information on gaze - can be used for HCI purposes. This thesis considered so-called eye gestures, sequences of an arbitrary number of consecutive relative eye movements. The work investigated how such eye gestures could be detected in real-time from continuous EOG data. To emphasise the technical system development at this early stage, this work focused on a stationary setting. With a view to a future application in mobile settings, this work also investigated means to remove signal artefacts caused by physical activity. This required to develop algorithms for adaptive real-time EOG signal processing, eye movement analysis, and eye gesture encoding and recognition.

**Development of an Eye-Based Activity Recognition Architecture**

Sensors for activity recognition commonly employed today, such as accelerometers or gyroscopes, have been used for more than a decade. A large number of studies have identified signal features and machine learning techniques most suitable for recognising activities in stationary and mobile settings. A similar knowledge base has not yet been created for eye-based activity and context recognition. In psychology literature, a limited number of eye movement features, such as fixation duration or blink rate, are known to be linked to certain visual tasks. Another goal of this work was to develop a feature set that covers a broader range of eye movement dynamics than those described in psychology. The features were derived from three basic eye movement characteristics detected from EOG signals: saccades, fixations, and blinks. This feature set only represents one building block of a complete software architecture for eye-based activity recognition. A second goal of this work was to complement the features with algorithms for signal processing as well as activity recognition and evaluation. The features and algorithms were designed to be generic, i.e. to not depend on the particular choice of EOG as a measurement technique, the chosen set of activities, or the selection of settings.
1.6. Thesis Outline

Recognition of Daily Life Activities from Eye Movement Data

Several studies aimed to model visual behaviour during specific (visual) tasks. These studies explored the link between these tasks and a number of eye movement features but they did not aim to recognise the task or activity using this information. Another aim of this thesis was to infer activity from eye movement data. The approach taken in this work was to use machine learning techniques to map eye movement data to a set of context classes. In a training phase, eye movements were recorded while a person performed a certain activity. During operation, eye movements were then compared to those observed during training to decide for the most similar context class. Following this scheme, the recognition architecture was applied and evaluated on two example activity recognition problems. The first study focused on spotting and recognising reading activity from a large amount of eye movement data. The data was recorded from several people in different natural mobile daily life situations. The second study investigated the problem of recognising several common desk-based activities in a stationary office setting in a natural work environment.

Towards Cognitive-Awareness Using Eye Movement Analysis

As described above, eye movements are not only influenced by physical activity but they are also strongly linked to a number of cognitive processes of visual perception. In memory recall tasks, significant differences in eye movement patterns were found for people looking at familiar and unfamiliar faces [32]. This raises the question whether a machine can automatically assess if a person looks at familiar or unfamiliar pictures by analysing the person’s eye movements. The last aim of this thesis was to perform an initial step towards such a system and thus the vision of cognitive-awareness. The work investigated whether the measurement technique and the eye movement features developed for eye-based activity recognition yielded similar results to those described in [32] using a video-based gaze tracker.

1.6. Thesis Outline

This thesis comprises six scientific publications that address the aims described in the last section. Figure 1.1 visualises the thesis’s aims and the chapters that cover them; arrows indicate relations in terms of results. Table 1.1 links each of these chapters to the corresponding publications.

Chapter 2 provides a summary of contributions, discusses the relevance of the findings and their limitations, and gives an outlook to future work. Chapter 3 details the eye tracker specifically developed for wearable eye movement recording based on EOG, the so-called wearable EOG goggles. Chapter 4 compares the system’s performance to a common video-based eye tracker on the example of a stationary eye gesture human-computer interface. Chapter 5 extends on this by outlining the potential of using the wearable EOG goggles for context-aware gaming. Chapters 6 and 7 describe and evaluate the developed EOG signal processing algorithms, the eye movement fea-
Chapter 1: Introduction

Figure 1.1: Outline of the chapters included in this thesis according to the aims presented in Section 1.5 (see Table 1.1 for the corresponding publications). Arrows indicate relations in terms of results.

...ture set, and the recognition architecture on two example eye-based activity recognition problems. Finally, as a first step towards the vision of cognitive-awareness, Chapter 8 presents results from a study on assessing visual memory recall processes from eye movements.
### Table 1.1: Publications and corresponding chapters included in this thesis.

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Publication</th>
</tr>
</thead>
</table>
| 3       | Wearable EOG goggles: Seamless sensing and context-awareness in everyday environments  
Andreas Bulling, Daniel Roggen and Gerhard Tröster  
Journal of Ambient Intelligence and Smart Environments, 1(2), 157-171,  
IOS Press, April 2009 |
| 4       | It’s in Your Eyes - Towards Context-Awareness and Mobile HCI Using Wearable EOG Goggles  
Andreas Bulling, Daniel Roggen and Gerhard Tröster  
| 5       | EyeMote - Towards Context-Aware Gaming Using Eye Movements Recorded From Wearable Electrooculography  
Andreas Bulling, Daniel Roggen and Gerhard Tröster  
| 6       | Robust Recognition of Reading Activity in Transit Using Wearable Electrooculography  
Andreas Bulling, Jamie A. Ward, Hans Gellersen and Gerhard Tröster  
| 7       | Eye Movement Analysis for Activity Recognition Using Electrooculography  
Andreas Bulling, Jamie A. Ward, Hans Gellersen and Gerhard Tröster  
| 8       | What’s in the eyes for context-awareness?  
Andreas Bulling, Daniel Roggen and Gerhard Tröster  
1.7. Additional Publications

The following publications have been written in addition to those presented in this thesis:


Bibliography


Chapter 1: Introduction


1.7. Additional Publications


Chapter 1: Introduction


Chapter 1: Introduction


This chapter summarises the contributions of the thesis. Specifically, the chapter introduces the eye tracker developed for wearable eye movement recordings - the wearable EOG goggles. It presents the results of a study to evaluate this system for a stationary eye-based HCI application. The chapter then summarises the developed eye movement features and the recognition architecture for eye-based activity recognition. The chapter presents the results of three studies that applied these methods to eye-based activity recognition and the problem of assessing memory recall processes. The chapter closes by discussing limitations to these contributions and by identifying open research questions that may lead to further advances in eye movement sensing and analysis, and eye-based activity recognition.
Chapter 2: Thesis Summary

2.1. Summary of Contributions

This section presents the most important results and contributions that advance the state-of-the-art in eye movement sensing and activity recognition. The summary is structured according to the aims of the thesis introduced in Section 1.5 on page 5 and illustrated in Figure 1.1 on page 8. Detailed descriptions and discussions can be found in the particular publication chapters referenced in this summary (cf. Table 1.1, page 9).

2.1.1. Development of a Wearable Eye Movement Sensor

This work investigated EOG as an alternative measurement technique for wearable eye movement tracking using body-worn sensors. This investigation required to develop a wearable EOG-based eye tracking sensor, and to investigate means for adaptive compensation of signal artefacts caused by physical activity. With regard to the first aim (see Chapter 1.5, page 5) the following contributions were made.

Wearable EOG Goggles

First, requirements for a wearable context-aware eye tracking device based on EOG were identified (see Section 3.3.1, page 38). These requirements included: 1) a light-weight and low-power design, 2) real-time EOG signal processing, 3) on-board data storage and wireless transmission capabilities, and 4) additional sensors such as an accelerometer and a light sensor for compensating EOG signal artefacts caused by physical activity and changes in ambient light.

The final design was a goggle-type eye tracker consisting of two components: a goggles frame with integrated EOG electrodes and a processing unit (see Figure 3.2 on page 40, Figure 3.3 on page 41, and Figure 3.4 on page 42). During the design process, trade-offs related to wearability and signal quality, and the type of electrodes were investigated (see Section 3.3.2, page 39). For increased wearability, only the analogue amplification circuitry was integrated into the goggles frame. Dry electrodes were chosen because they are more convenient and allow for easy removal and reattachment of the goggles to the head. To increase the common-mode rejection ratio, similar to state-of-the-art electrocardiogram (ECG) systems, a driven right leg circuit was implemented on the device. The core signal processing unit of the system has a credit card size of 82x56 mm. It is based on a 16-bit digital signal processor and contains dedicated 24-bit analog-digital converters for each EOG channel, a Bluetooth, and a MMC module. The device runs freeRTOS, a real-time operating system devised for embedded systems. The complete system weighs 188 g and is powered by a 1500 mAh Li-polymer battery. This allows for up to 7.2 hours (MMC) and 6.7 hours (Bluetooth) of eye movement recording.

Adaptive Motion Artefact Compensation
As EOG is measured with body-worn sensors, body motion causes artefacts in the EOG signals which affects eye movement detection (see Section 3.4.1, page 44). This thesis identified a number of EOG signal characteristics that need to be preserved while removing these artefacts. First, the signal edges needed to be retained to allow for the detection of blinks and saccades. Second, EOG signal amplitudes needed to be preserved to be able to distinguish between different types and directions of saccadic eye movements. In addition, denoising filters should not introduce signal artefacts that may have been misinterpreted in subsequent signal processing steps.

Walking is a common activity in everyday life and was therefore selected in this thesis as a test bench for investigating artefacts induced by body motion. Analyses showed that walking introduced periodic EOG signal artefacts depending on the current step frequency. This means that the optimal window size of any filter to remove these artefacts depends on the temporal step length. Therefore, this work developed an adaptive filter that exploited the repetitive characteristics of walking and adapted the window size of a median filter to the step length (see Figure 3.7, page 47). The algorithm for detecting walking used the goggles-mounted accelerometer (see Figure 3.4, page 42). The adaptive filter was implemented on the device and evaluated in a mobile setting. The experimental scenario involved five participants to perform different eye movements while standing and walking down a corridor (see Table 4.5, page 78). The expected eye movements were shown on a head-up display (HUD) with a defined order and timing (see Figure 4.7, page 79). Compared to a median filter with fixed window size, the adaptive filter reduced signal artefacts caused by walking activity by up to 80% in the horizontal, and up to 60% in the vertical EOG signal component (see Figure 4.8, page 84).

### 2.1.2. HCI in Stationary Settings Using Relative Eye Movements Detected From EOG

This work investigated how relative eye movements detected using the wearable EOG goggles (see Section 2.1.1, page 18) could be used as a means to interact with a computer. With regard to the second aim (see Chapter 1.5, page 5) the following contributions were made.

**Eye Movement Encoding and Eye Gesture Recognition**

Eye gestures were so far only presented as an input technique for stationary video-based eye trackers. Eye gesture recognition is based on the detection of consecutive saccades which by their order and direction define the type of eye gesture (see Figure 4.1, page 73). For using EOG, new algorithms for eye movement encoding and eye gesture recognition needed to be developed. First, saccades were detected in the continuous vertical and horizontal EOG signal streams. These saccades were then encoded into distinct eye movements. Both processing steps were implemented as embedded algorithms for real-time execution on the wearable EOG goggles.

The thesis introduced the idea of discretising eye movements based on their direction and EOG signal amplitude. The discretisation encoded sac-
cades of both EOG signal components into an alphabet of characters representing the different eye movements. The number of characters in the alphabet was application dependent. For HCI, an alphabet of 16 characters was used (see Figure 4.4, page 71, top right corner). For eye-based activity recognition, two different alphabets were chosen. For recognising reading activity, only two characters - representing the characteristic small right and the large left eye movements during reading - were used (see Figure 6.4, page 116); for recognising less structured office activities, the alphabet was extended to 24 discrete characters (see Figure 7.5a, page 143). This discretisation procedure merged simultaneous saccades into one eye movement character sequence that could be more efficiently processed and analysed (see Figure 7.5b, page 143).

For continuous eye gesture recognition, the character sequence was scanned for eye movement patterns following a string matching approach. For matching, the current string sequence was continuously compared with string templates representing all possible gestures (see Table 4.1, page 73). For each template, the edit distance between the templates and the segment was calculated. If one of the templates matched the current segment (i.e. the edit distance is zero), the corresponding eye gesture was recognised.

Eye Gestures from EOG for Stationary HCI

The eye gesture recognition algorithm was evaluated using the wearable EOG goggles for stationary human-computer interaction. As an example application, the work developed a computer game consisting of eight different game levels. In each game level, participants had to perform one defined eye gesture. 11 participants performed three runs with all eight game levels being played in each run. Participants were seated in front of the screen facing its centre (see Figure 4.6, page 74). In contrast to state-of-the-art systems using video-based eye trackers, no head stand was used, i.e. movements of the head and the upper body were allowed at any time during the experiments. The eye gestures were selected to be of increasing complexity (see Table 4.1, page 73).

The results showed that the concept of playing a computer game using eye gestures was quickly understood by all participants. All participants achieved an eye gesture accuracy of around 90% and often managed to perform the various gestures at the first try (see Table 4.2, page 75). In the course of the three experimental runs, the participants’ performance increased by about 8% accuracy (see Figure 5.4, page 96). The wearable EOG goggles proved to be a robust system for detecting the example set of eye gestures. The average accuracy over all participants in performing the eye gestures was between 83% and 93% (see Table 4.3, page 77). The results confirmed comparable performance to a video-based system on a subset of these gestures (see Table 4.4, page 77).
2.1.3. Development of an Eye-Based Activity Recognition Architecture

This thesis developed a software architecture for activity recognition based on eye movement analysis. Input to the recognition architecture were the two EOG signals capturing the horizontal and the vertical eye movement components. In the first stage, these signals were processed to remove signal artefacts that might hamper eye movement analysis. In the case of EOG signals, algorithms for baseline drift and noise removal were applied. Only this initial processing depended on the particular eye tracking technique used; all further stages were independent of the underlying type of eye movement data.

In the next stage, three different eye movement types were detected from the processed eye movement data: saccades, fixations, and blinks. These eye movement types, together with wordbooks describing repetitive eye movement patterns, formed the basis of the 90 features developed in this work for eye-based activity recognition. The robustness of the algorithms for detecting saccades, fixations, and blinks were therefore key to achieving good recognition performance. All features were extracted using a sliding window. In the last stage, a hybrid method selected the most relevant of these features, and used them for activity recognition (see Section 2.1.4, page 23). With regard to the third aim (see Chapter 1.5, page 5) the following contributions were made.

Saccade and Fixation Detection

For saccade detection, the thesis developed the so-called Continuous Wavelet Transform - Saccade Detection (CWT-SD) algorithm (see Figure 7.3, page 140). Using thresholding on a wavelet coefficient vector calculated from the raw EOG signal, the signal was divided into saccadic and non-saccadic (fixational) segments (see Section 7.4.3, page 139, for details). Humans typically alternate between saccades and fixations. This allowed us to use CWT-SD for detecting fixations as well. The algorithm for fixation detection exploited the fact that gaze remains stable during a fixation. This results in the corresponding gaze points (the points in visual scene gaze is directed at) to cluster together closely in time. Fixations were identified by thresholding on the dispersion of these gaze points. To increase robustness, both algorithms incorporated checks using thresholds that approximate the typical saccade and fixation characteristics described in eye physiology literature.

The thesis evaluated the CWT-SD algorithm in an experiment with five participants. A total of twenty recordings were made per participant, 10 on two different days. The experiment involved tracking the participants’ eyes while they followed a sequence of flashing dots on a computer screen. The sequence was comprised of a total of 591 horizontal and 855 vertical saccades. The thesis evaluated the F1 score across a sweep on the CWT-SD threshold separately for each EOG signal component. CWT-SD achieved a maximum F1 score of 0.94 averaged over all participants (see Figure 7.4, page 142). The standard deviation across all participants reached a minimum for a whole range of threshold values around this maximum. This suggested
that also thresholds close to this point can be selected that still achieve robust detection performance.

**Blink Detection**

The thesis investigated different methods for blink detection with a focus on real-time algorithms. It developed and compared two new methods for blink detection - template matching and a method based on continuous wavelet transform (CWT-BD) - with common methods, namely velocity threshold analysis and Haar wavelet decomposition (see Section 3.4.2, page 47). The thesis first evaluated each of these algorithms on a multi-participant EOG dataset containing a total of 105 blinks. In this evaluation, template matching and CWT-BD showed the best performance. Template matching achieved an overall precision in detecting blinks of 99% and a recall of 96%, CWT-BD 95% precision and 99% recall. The other methods showed worse performance with recall values of only 72% and 87% (see Table 3.1, page 49).

The thesis further evaluated the CWT-BD algorithm on EOG signals recorded from five participants looking at different pictures. A total of 706 blinks was labelled by visual inspection of the vertical EOG signal component. With an average blink rate of 12 blinks per minute, this corresponded to about 1 hour of eye movement data. CWT-BD was evaluated over sweeps of its two main parameters, the blink threshold and the maximum blink duration. CWT-BD performed best for durations between 400 ms and 600 ms and reached a top performance with a F1 score of 0.94 for 500 ms (see Figure 7.6, page 144). Durations outside this range were subject to a considerable drop in performance. These findings reflected the values for the average blink duration known from literature.

For HCI purposes only, detected blinks were removed from the vertical EOG signal depending on the type of the blink: presaccadic, intersaccadic, or postsaccadic (see Section 4.4.3, page 70).

**Analysis of Repetitive Eye Movement Patterns**

Structured activities such as reading involve characteristic patterns of eye movements that are repeatedly performed. To be able to analyse these patterns, the thesis developed a method to encode eye movements into a discrete, character-based representation (see Section 2.1.2, page 19). An eye movement pattern was defined as a string of an arbitrary number of successive characters. As an example for the pattern length four, the string “LrBd” translates to large left (L) → small right (r) → large diagonal right (B) → small down (d). These strings were collected in different wordbooks using a sliding window approach (see Figure 7.7, page 145). Each wordbook contained the type and occurrence count for each pattern of a particular length. Each newly found pattern of a certain length was added to the corresponding wordbook. For a pattern that was already included in the wordbook, the occurrence count of the pattern was increased by one.

**Eye Movement Features**
The thesis investigated four groups of features based on the detected saccades, fixations, blinks, and the wordbooks of eye movement patterns (see Table 7.1, page 146, for the naming scheme). These four groups comprised 90 features that were calculated using a sliding window. Some of the features were directly derived from particular eye movement characteristics, others were devised to capture additional aspects of eye movement dynamics (see Section 7.4.5, page 145).

The thesis first investigated the features’ ability to discriminate between different activity classes (see Section 7.5, page 147). The top 15 features covered three of the four proposed groups with saccade, fixation, and wordbook features all prominently represented (see Figure 7.11, page 159). Each feature group was found to cover specific aspects of eye movements that complemented each other. This is a promising finding for the general utility of these groups for other activity recognition problems. The analysis also revealed that the top three features were all based on horizontal saccades. The fixation rate was used for all participants and highly ranked for seven of them. The blink rate was selected for five out of the eight participants but was assigned a high rank for only one. Wordbook features were not used for only one of the participants, but they were highly ranked for the other seven.

An additional analysis revealed that the selection of the most important features highly depends on the nature of the activity (see Table 7.3, page 152).

Reading is a regular pattern characterised by a specific sequence of saccades and short fixations of similar duration. Consequently, reading activity could be best described by using wordbook features that reflect eye movement sequencing. Writing was similar to reading, but required greater fixation duration and greater variance. It was best described using average fixation duration and its variance as well as wordbook features. Copying involves regular back and forth saccades between screens. This was reflected in the selection of a mixture of small and large horizontal saccade features, as well as variance in horizontal fixation positions.

In contrast, watching a video and browsing are highly unstructured and - depending on the video or website being viewed - may be comprised of different activities, e.g. watching a video, reading text, or looking at a still picture. The lack of wordbook features reflected this, as did the mixed selection of features: variance of both horizontal and vertical fixation positions, small positive and negative saccadic movements, and blink rate.

These results suggest that for tasks that involve a known set of specific activity classes, recognition can be optimised by only choosing eye movement features known to best describe these classes.

### 2.1.4. Recognition of Daily Life Activities From Eye Movements

The thesis investigated two activity recognition problems using the developed eye movement features and eye-based activity recognition architecture (see Section 2.1.3, page 21). With regard to the fourth aim (see Chapter 1.5, page 5) the following contributions were made.
Chapter 2: Thesis Summary

Recognition of Reading Activity in Transit

The eye-based activity recognition architecture was first evaluated on the problem of spotting and recognising reading of people in transit in mobile daily life environments. The thesis considered a scenario of travelling to and from work (see Table 6.1, page 113). The experiment involved eight participants occasionally reading text during different modes of locomotion including sitting at a desk, walking along a street, waiting at a tram stop and riding a tram (see Figure 6.2, page 112). Each participant was tailed by an assistant who annotated both the current mode of locomotion, and whether the participant was reading. To avoid distractions, a wireless “Wii Remote” controller from Nintendo was used for labelling (see Figure 6.3, page 114). A total EOG dataset of roughly six hours was recorded with reading occurring about half of the time. This required spotting reading in a dataset with more than 50% of other types of eye movements.

Reading involves frequent, short scans and less frequent, longer newline movements (see Figure 6.1, page 109). Therefore, only the left and right saccadic movements of the eyes were analysed (see Figure 6.4, page 116). Three approaches for continuous recognition of reading activity were compared: string matching and two variants of Hidden Markov Models - mixed Gaussian and discrete. Using person-independent training and parameters, a top accuracy of 80.2% (precision: 84.6%, recall: 71.0%, 11.6% false positives) over all participants was achieved using string matching (see Figure 6.6, page 120). The results for each participant showed a range of differences in recognition performance using the string matching algorithm (see Table 6.2, page 123). These differences did not correlate to the gender of the person. The highest recall result was 92.9% with a false positive rate (FPR) of 15.5%. The worst result was 25.9% recall and 3.8% FPR. This was to be expected as the raw EOG signal quality for that particular participant was very low. An analysis of the different modes of locomotion uncovered that walking produced the worst results but only by a small margin (precision: 85.2%, recall: 64.9%). Recognition was better while participants sat (precision: 89.2%, recall: 73.9%) than when they stood (precision: 81.0%, recall: 72.8%; see Figure 6.7, page 121).

The main finding was that EOG is a robust measurement technique to recognise reading in daily life scenarios across different participants and modes of locomotion. In contrast to video-based systems, the participants only had to wear light-weight equipment. This small weight contributed to the participants feeling unconstrained and allowed for natural reading behaviour.

Recognition of Office Activities

A second study investigated the recognition of a set of typical office activities from eye movements recorded using EOG. Eight participants were involved in two continuous activity sequences each lasting for about 30 minutes. This resulted in a total EOG dataset of about eight hours. Each sequence was comprised of five different office activities performed in random order: copying a text between two screens, reading printed texts, taking hand-written notes, watching a video and browsing the web (see Figure 7.8, page 148). In ad-
2.1. Summary of Contributions

In addition, a period of rest was included (the NULL class). For this period, the participants were asked not to engage in any of the previous activities.

For classification, the study evaluated the eye-based activity recognition architecture (see Section 2.1.3, page 21) in combination with a support vector machine (SVM) classifier. Using person-independent training and parameters, an average precision of 76.1% and recall of 70.5% over all classes and participants were achieved. The number of features used for each participant to achieve this result varied from only nine features up to 81 features (see Table 7.2, page 151). Reading was a pervasive activity also in this study. Consequently, this led to confusions with browsing, which involves a variety of sub-activities including reading (see Figure 7.10, page 158). An analysis of the recognition performance for each activity revealed that six out of the eight participants returned best average precision and recall values of between 69% and 93%. Two participants, however, returned results that were lower than 50% (see Figure 7.9, page 158). It turned out that, similar to the first study, in both cases the EOG signal quality was worse compared to the other participants. Changes in signal amplitude for saccades and blinks - upon which feature extraction and thus recognition performance directly depend - were not distinctive enough to be reliably detected.

The study revealed several findings for the general problem of context recognition using eye movement analysis. First, eye movement analysis can serve as a sensing modality for recognising human activity. Second, good recognition performance required to use a mixture of several eye movement features. Finally, information on repetitive patterns of eye movements also proved to be useful. As different recognition tasks likely require different combinations of features, this work recommends that a mixture of feature types be considered for each new task.

2.1.5. Towards Cognitive-Awareness Using Eye Movement Analysis

Finally, the thesis investigated the link between eye movements and cognition. Specifically, it focused on the problem of assessing memory recall processes from eye movements. With regard to the fifth aim (see Chapter 1.5, page 5) the following contribution was made.

Assessment of Memory Recall Processes

The thesis introduced eye movements detected from EOG signals as a promising sensing modality for cognitive-aware systems. A study with six participants investigated the feasibility of assessing memory recall processes of people looking at familiar and unfamiliar pictures by analysing eye movement characteristics. Participants were exposed to four continuous sequences of random pictures selected from four categories: abstract images, pictures of buildings, faces, and landscape photographs (see Figure 8.5a, page 180).

This study revealed that the number of fixations was one of the most discriminative eye movement features to assess memory recall processes of a person. The so-called fixation count decreased significantly with the num-
ber of prior exposures at a significance level of $p < 0.05$ (see Figure 8.5b, page 180). While a more detailed analysis is required, EOG showed similar performance to a video-based eye tracker used in a previous study. This finding supports EOG as a measurement technique for capturing eye movement characteristics that reflect memory recall processes.
2.2. Conclusion

Activity recognition has become a key area of research for context-awareness in ubiquitous computing and HCI. Current context-aware systems are limited in the type and complexity of activities they are able to recognise. They also have a hard time to deal with the subtleness with which some of these activities occur. In addition, activity sensing using subtle cues remains largely unexplored. This calls for alternative sensing modalities that complement those modalities commonly used today.

This thesis introduced and demonstrated the use of eye movements as a novel sensing modality for context-awareness and activity recognition. Specifically, this work investigated methods for wearable eye tracking, the real-time detection of and feature extraction from eye movement characteristics, eye-based activity recognition, and the assessment of memory recall processes from eye movements. Based on the summary presented in Section 2.1 (page 18), the following conclusions can be drawn:

- Eye movement dynamics - without any information on gaze - provide vital context information that is not yet available using common sensing modalities. In particular, eye movements reflect the physical activities during which they are performed. Eye movements also hold great promise for providing means to assess several cognitive processes of visual perception.

- To use the eyes as a sensing modality requires to employ an architecture that is particularly geared towards eye-based context inference. Such an architecture comprises methods for wearable eye movement sensing and the detection of different eye movement characteristics, a discriminative set of eye movement features, as well as machine learning techniques for classification.

- EOG proved to be a robust measurement technique for wearable eye movement sensing. Results for eye-based HCI showed that EOG can be used as an alternative to well-known video-based eye tracking. EOG is particularly suited for mobile settings as it is cheap, only requires light-weight signal processing, and can be implemented using body-worn sensors. The feasibility of a wearable EOG-based eye tracker was demonstrated in the form of goggles in this work.

- Three of the main eye movement characteristics particularly useful for eye-based activity recognition - saccades, fixations, and blinks - can be robustly detected from EOG signals across different people. The current work has proposed 90 different eye movement features that can be extracted from these characteristics. These features are able to discriminate between different visual behaviours as they cover a broad range of eye movement dynamics that reflect the specific nature of the underlying physical activities.
Chapter 2: Thesis Summary

- The feasibility of eye-based activity recognition was shown on two example activity recognition problems: the spotting and recognition of reading activity in mobile daily life settings (precision: 84.6%, recall: 71.0%, eight participants), and the recognition of five office activities in a stationary office setting (precision: 76.1%, recall: 70.5%, eight participants). The best recognition results were achieved using a mixture of eye movement features. Sequence information on eye movement patterns, in the form of a wordbook analysis, also proved particularly useful.
2.3. Limitations and Relevance

This work has introduced EOG as a measurement technique for wearable eye tracking. EOG signal artefacts, such as noise or baseline drift, are an issue particularly for eye movement recordings in mobile settings. It is for this reason that some eye movement characteristics potentially useful for activity recognition, such as microsaccades or smooth-pursuit movements, were not investigated in this work. Characteristics such as the pupil diameter will most likely remain inaccessible using EOG.

The wearable EOG goggles proved to be a robust eye tracker for stationary eye-based human-computer interaction using a standard computer screen. This work did not investigate how the system performs in mobile settings, for example using a much smaller head-up display. One of the main issues with the current prototype is the mechanical mounting. Poor placement of electrodes and signal artefacts due to electrode movements on the skin were the reason for many of the problems in this work. For long-term use, wearability and comfort will become even more important. The current prototype has not yet been optimised to match these requirements. For example, several users reported a rather high electrode pressure below the eye.

The developed feature set and the recognition architecture were designed to not depend on the measurement technique or the particular choice of activities and settings. However, the two activity recognition problems investigated in the current work only cover a subset of the activities observable in daily life. It remains to be investigated which other activity recognition problems the architecture can be applied to. In particular, it is an open question how the architecture performs for a more diverse range of physical and visual activities, settings, and a larger number of people.

Precise ground truth annotation is an open issue in activity recognition. Eye movement annotation is particularly challenging as the eyes move constantly, fast, and only subtle at times. In this work, issues with annotation were not explicitly addressed. Instead, these issues could only be minimised by focusing on very characteristic visual tasks (such as reading), and by following well-structured experimental procedures that were easier to annotate.
2.4. Outlook

This thesis has opened up new and promising research perspectives that may lead to further advances in EOG signal processing, eye movement sensing and analysis, eye-based activity recognition, and eye-based interaction.

Further research on these topics should address the following challenges:

- **EOG signal quality** - Artefacts in the EOG signals were one source of errors in this work. Future research should focus on improving EOG signal quality, for example by developing novel electrodes or by improving the electrode mounting and the algorithms for EOG signal artefact removal. Improvements in electrode mounting will also contribute to unobtrusiveness and thus the general applicability of EOG for wearable eye movement recordings.

- **Eye movement characteristics** - Additional eye movement characteristics potentially useful for eye-based activity recognition - such as smooth pursuit movements, pupil dilation, microsaccades, or the vestibulo-ocular reflex - are difficult to detect given the presence of noise and baseline drift in the EOG signals. Video-based eye trackers are less susceptible to such artefacts and will provide access to at least some of these characteristics. It should be considered to combine both sensing modalities to extend the capabilities of the recognition architecture presented here.

- **Generalisability of the recognition architecture** - The developed recognition architecture was evaluated on a selected set of common daily life activities. It is necessary to extend this selection with a broader range of activities observable in other daily life situations. This will lead to new insights into the generalisability of the methods, particularly with respect to the developed eye movement features. In addition, future research should use long-term recordings to investigate the robustness and stability of these features over time.

- **Fusion with additional modalities** - This work has introduced eye-based activity recognition on predominantly visual daily life activities. For arbitrary physical activities, e.g. in sports, the combination of eye movements with other modalities such as head movements or hand gestures will prove beneficial. Sensor fusion will also allow us to analyse hand-eye coordination or secondary task performance in addition to eye-based activity recognition.
3

Wearable Eye Movement Sensing

Andreas Bulling, Daniel Roggen and Gerhard Tröster

Full publication title: Wearable EOG goggles: Seamless sensing and context-awareness in everyday environments

DOI: 10.3233/AIS-2009-0020
Chapter 3: Wearable Eye Movement Sensing

Abstract

In this article we introduce the analysis of eye motion as a new input modality for activity recognition, context-awareness and mobile HCI applications. We describe a novel embedded eye tracker that, in contrast to common systems using video cameras, relies on Electrooculography (EOG). This self-contained wearable device consists of goggles with dry electrodes integrated into the frame and a small pocket-worn component with a DSP for real-time EOG signal processing. It can store data locally for long-term recordings or stream processed EOG signals to a remote device over Bluetooth. We show how challenges associated with wearability, eye motion analysis and signal artefacts caused by physical activity can be addressed with a combination of a special mechanical design, optimised algorithms for eye movement detection, and adaptive signal processing. In two case studies, we demonstrate that EOG is a suitable measurement technique for the recognition of reading activity and eye-based human-computer interaction. Eventually, wearable EOG goggles may pave the way for seamless eye movement analysis in everyday environments and new forms of context-awareness not possible today.

3.1. Introduction

Fifteen years ago, ubiquitous computing was introduced as a vision of technology fading into the background and always ready to interact and transparently respond to the users’ needs. To realise this vision it is essential to be able to recognise and react according to the users’ context [1, 2]. This is primarily a pattern recognition problem in which context is inferred from specific sensor signal patterns. Sensing the users’ context may thereby rely on ambient infrastructure - so-called smart environments - or on wearable sensing and computing [3]. Wearable sensing allows for permanent context recognition and enables access to aspects that are difficult to measure with ambient sensors. By combining both, ubiquitous systems can be developed that provide seamless sensing across different environments the user visits in his daily routine.

Human activity is one of the key aspects of user context. The recognition of physical activity both using ambient [4, 5] and body worn [6] sensors has been extensively studied. However, context-awareness encompasses more than mere physical activity. The inclusion of social and affective aspects may help to paint a more accurate view of the user’s context [7]. Cognitive aspects such as attention and intentionality may enhance activity recognition systems and benefit proactive wearable assistants. So far, most of these aspects remain unexplored as they cannot be picked-up by sensors usually deployed in today’s wearable and ubiquitous computing scenarios.

A rich source of information about the state of the user can be found in the movement of the human eyes. This includes information related to
the users’ activities, such as reading [8], but also to cognitive processes of visual perception such as attention [9], saliency determination [10], visual memory [11] and perceptual learning [12]. Deliberate eye movements can be implemented for human-computer interaction (HCI) to provide feedback that can be used as explicit contextual information [13]. Attentive user interfaces (AUI) may also infer user intention and activity by analysing unconscious eye movements. As part of an ongoing project we investigate up to which extent deliberate and unconscious eye movements can be exploited to enable new kinds of context-aware applications.

The investigation of eye motion may focus on eye tracking or gaze tracking. Eye tracking refers to the analysis of general characteristics of relative eye movements, long-term movement dynamics or statistical properties of eye motion. In contrast, gaze tracking refers to the estimation of absolute gaze direction.

For stationary settings, a wide range of hardware for accurate gaze tracking is available. However, state-of-the-art mobile systems still do not meet the requirements for unobtrusive long-term recordings and real-time analysis in everyday environments. To our knowledge, at this stage no wearable solution exists that is self-contained and unobtrusive for use in daily-life activities. Furthermore, current video-based systems are not geared to embedded online analysis of eye motion to recognise human activity and infer user context.

3.1.1. Contributions

In this work, we address these issues by describing the design and implementation of a highly-integrated wearable eye tracker for context-awareness and mobile HCI based on Electrooculography (EOG). In contrast to well-established vision-based gaze tracking, EOG is measured with body-worn sensors and can be implemented as a low-cost and low-power embedded system. The device allows for unobtrusive recordings of EOG signals and their real-time processing and enables online inference of activity and context. It consists of glasses and a light-weight device worn on the body and is particularly designed for long-term use in daily life with simultaneous physical activity.

The specific contributions of this work are (1) the design and implementation of a wearable EOG-based eye tracker implemented as goggles, (2) the development of a software framework for continuous EOG signal analysis and the detection of eye movement events, (3) the evaluation and implementation of a set of algorithms for robust EOG signal processing within this framework and (4) the characterisation of wearable EOG and the goggles in two case studies.

3.1.2. Paper Organisation

Section 3.2 provides information on the physiology of eye motion and the state-of-the-art in eye movement research with particular emphasis on sensors and applications. Based on the definition of the requirements for a wearable
Chapter 3: Wearable Eye Movement Sensing

EOG-based eye tracking device, in Section 3.3, we describe the implementation challenges and the final design of the wearable EOG goggles. Section 3.4 gives detailed information on the signal processing algorithms implemented on the device for continuous recognition of eye movement sequences. Two case studies enabled by wearable EOG and the goggles are outlined in Section 3.5. In Sections 3.6 and 3.7, we discuss the current status of the system, comment on the case studies and give an outlook to future work, respectively.

3.2. Eye Movements: Physiology, Sensors and Applications

3.2.1. Physiology of Eye Motion

To be able to take advantage of the typical characteristics of eye movements to perform context recognition and implement eye-based HCI, it is important to understand its two main types, namely saccades and fixations.

**Saccades:** Humans do not look at a wider scene in a steady way. Instead, their eyes move around constantly to locate interesting parts and combine them into one mental representation. The main reason for this is that only a small central region of the retina, the fovea, allows to perceive the scene with high acuity. Simultaneous movements of both eyes in the same direction are called saccades. Typical characteristics of saccadic eye movements are 400°/s for the maximum velocity, 20° for the amplitude and 80ms for the duration [14].

**Fixations:** A fixation is the static state of the eyes during which gaze is held upon a specific location. Humans typically alternate saccadic eye movements and fixations (see Figure 3.1). However, visual fixation is never perfectly steady and fixational eye movements can also occur involuntarily. The term “fixation” can also be referred to as the time between two saccades during which the eyes are relatively stationary. Certain activities involve characteristic fixation patterns. For example during reading, the eyes fixate on successive locations within a line but also across a page to reach different sections of the text [8].

3.2.2. Eye Movements in Daily Life

A growing number of researchers investigate movements of the eyes during daily activities. Important advances have been made to understand how the human brain processes visual tasks [15], how vision contributes to the organisation of active tasks in everyday life [16] and how eye, head, and hand movements are coordinated temporally [17]. In a recent study, Logan et al. studied activity recognition with a person living in a smart environment instrumented with a large number and variety of common sensors [4]. They found that among all activities, reading was one of the most difficult to detect.
3.2. Eye Movements: Physiology, Sensors and Applications

![Figure 3.1: Denoised horizontal (EOG_h) and vertical (EOG_v) EOG signals with marked segments showing the two main eye movement characteristics: saccades (S) and fixations (F).](image)

They concluded that in order to catch all types of physical activity in daily-life scenarios, novel sensors and algorithms need to be developed.

Although eye movements have been investigated in daily routine, none of these studies used eye movement patterns to perform activity and context recognition. Furthermore, other aspects of visual perception such as attention [9], saliency determination [10], visual memory [11] or perceptual learning [12] so far remain unexplored as a novel input for context-aware systems.

3.2.3. Video-Based Eye Tracking

**Devices**

The common method to track eye gaze in natural environments are systems based on video cameras. A number of commercial gaze trackers are available of which some are targeted at mobile use, for example the Mobile Eye from Applied Science Laboratories (ASL) [18] or the iView X HED from SensoMotoric Instruments (SMI) [19]. Nevertheless, they still require bulky headgear and additional equipment to process the video streams. This does not allow for unobtrusive recordings and constrains the user in his physical activities. Furthermore, as the video processing is performed in real-time requiring considerable computational power, these systems are limited to only very few hours of mobile gaze tracking.

**Eye-Based Interaction**

Eye-gaze recorded using vision has long been investigated as a means to interact with a computer. However, due to the lack of appropriate hardware, mo-
bile eye-based interaction is a so far barely regarded field of research. Most HCI work has focused on direct manipulation of user interfaces in stationary settings. Zhai et al. proposed a new way of using eye gaze for computer input [20]. Results of an early-stage experiment indicated that their method potentially reduces physical effort and fatigue. Qvarfordt et al. explored an interactive human-computer dialogue system for touristic city trip planning [21]. They showed that it was possible to sense the users’ interests based on eye-gaze patterns and adapt the system’s output accordingly. Drewes et al. proposed to use eye gestures to implement new ways of HCI [22]. They argue that these gestures are insensitive to accuracy problems and do not exhibit the “Midas touch” problem (for details see [23]).

3.2.4. EOG-Based Eye Tracking

Electrooculography (EOG)

The eyes are the origin of a steady electric potential field, which can also be detected in total darkness and if the eyes are closed. It is generated by a dipole with its positive pole at the cornea and its negative pole at the retina. The magnitude of this so-called corneo-retinal potential difference (CRP) lies in the range of $0.4 \text{mV}$ to $1.0 \text{mV}$. On the assumption of an unchanging CRP, the electric signal that can be derived using two pairs of skin electrodes placed at periorbital positions around one eye is called Electrooculogram (EOG). EOG typically shows signal amplitudes ranging from $5 \mu \text{V/°}$ to $20 \mu \text{V/°}$ and an essential frequency content between $0 \text{Hz}$ and $30 \text{Hz}$ [24]. If the eyes move from the centre position towards the periphery, the retina approaches one electrode while the cornea approaches the opposing one. This change in the orientation of the dipole and the electric potential field results in a change in the measured EOG signal. Inversely, by analysing these changes, eye movements can be tracked.

Baseline Drift: Baseline drift is a slow signal change mostly unrelated to the actual eye movements but superposing the EOG signal. Baseline drift has many possible sources as for instance interfering background signals, electrode polarisation [25] or physical influences such as varying contact pressure of the electrodes. The differences in EOG signal amplitude during saccadic eye movements can be assumed to be drift-free as saccades are performed in a very short period of time. All other signals can become subject to changes caused by baseline drift. In a four electrode setup, baseline drift can be different for the horizontal and vertical EOG signal components.

Several approaches to remove baseline drift from electrocardiographic signals (ECG) have been proposed in recent literature (for example see [26, 27]). As ECG shows repetitive characteristics, some of the algorithms perform sufficiently well at removing baseline drift from these signals. However, they perform worse for signals with non-repetitive characteristics such as EOG. Thus, the development of robust algorithms for baseline drift removal from EOG signals is still an active field of research.
3.2. Eye Movements: Physiology, Sensors and Applications

Devices

Efforts to miniaturise video-based eye trackers led researchers to consider EOG signals recorded with standard equipment for eye tracking in stationary settings [28, 29]. Others investigated novel electrode configurations for implementing EOG-based eye trackers for mobile use. Manabe et al. proposed a system that uses EOG electrode arrays mounted on ordinary headphones [30]. While this approach might be less obtrusive than electrodes stuck to the face it raises other issues - namely, low signal-to-noise ratio (SNR) and poor separation of the horizontal from the vertical component of eye motion. Vehkaoja et al. presented a light-weight head cap for EOG and facial EMG measurements with electrodes embroidered of silver coated thread [31]. A small device integrated into the cap allows for wireless data transmission. Signal processing is performed offline using a standard desktop computer and the system is still to be evaluated in operation.

Although novel devices for recording eye motion using EOG have been developed, none of them allows for combined, embedded signal acquisition and activity and context recognition in a standalone wearable device. The system described in this work is a highly miniaturised, autonomous wearable sensor particularly designed for both tasks. Low-power and light-weight implementation, real-time signal processing capabilities and additional sensors for artefact compensation make this embedded system a unique solution for robust recordings in mobile daily life settings.

EOG-Based Interfaces

Basic eye movement characteristics detected from EOG signals such as saccades, fixations, blinks and deliberate movement patterns have been used for hands-free operation of stationary human-computer [32, 33] and human-robot [34, 35] command interfaces. As part of a hospital alarm system, EOG-based switches provided immobile patients with a safe and reliable way of signalling an alarm [36]. All of these studies show that EOG is a measurement technique that is easy to operate, reliable and can also be made cosmetically acceptable.

For mobile settings, EOG-based interfaces have been developed for assistive robots [37] and as a control for an electric wheelchair [38, 39]. These systems are intended to be used by physically disabled people who have extremely limited peripheral mobility but still retain eye motor coordination. Mizuno et al. used basic characteristics of eye motion to operate a wearable computer system for medical caregivers [40]. Although these studies target mobile settings the people themselves are still constrained in their movements.

3.2.5. Summary

In this section, we have shown that the eyes are a rich source of information which has not yet been used for activity and context recognition or mobile HCI applications. A review on the state-of-the-art in eye tracking devices and related applications revealed that the main reasons for this is the lack of
an appropriate sensor and missing processing techniques for online context recognition based on eye motion.

We have shown that EOG provides several advantages over common systems based on video in particular in terms of embedded implementation and long-term recordings in daily life. However, in current work the information obtained from EOG remains coarse, the users are static, and signal processing is done offline using desktop computers. In this work we demonstrate how complex contexts can be recognised from EOG in mobile scenarios using an autonomous wearable device without the need for such additional infrastructure.

### 3.3. Design and Implementation of a Wearable Context-Aware EOG Sensor

#### 3.3.1. Requirements

A wearable context-aware eye tracking device based on EOG that is robust to simultaneous physical activity and changing environments has to meet the following requirements:

1. To achieve a convenient and unobtrusive implementation and minimise user distraction the device needs to be wearable and lightweight.

2. To allow for autonomous long-term recordings the device needs to be low-power and support on-board data storage.

3. The device needs to provide real-time signal processing capabilities to allow for context-aware interaction.

4. To compensate for EOG signal artefacts caused by physical activity and changes in ambient light [24] an accelerometer and a light sensor need to be added.

Given the fact that EOG is recorded using electrodes placed around the eye, the natural choice for the basic form of the device was that of goggles: A goggles frame is built to minimise distraction to the user. It is very close to the face and covers the lateral positions on each side of the head commonly used for placing the electrodes. Only for the electrodes recording the vertical signal component above and below the eye an extension to the frame needs to be made. Furthermore, a goggles frame provides enough space to carry a small component that contains the amplification circuits for the analogue signals, the accelerometer and a connection to the light sensor. The latter can be fixed to the frame in between both eyes to provide measurements of incident light.
3.3. Design and Implementation of a Wearable Context-Aware EOG Sensor

3.3.2. Challenges

The first trade-off we had to deal with was the one between wearability and signal quality: To reduce noise, the analogue amplification circuit and conversion from analogue to digital signals must occur close to the electrodes integrated into the frame. However, this results in increased weight and size and therefore reduced wearability. To optimise the design for weight, only the light sensor, the accelerometer and the amplification circuits can be attached to the glasses frame. On the downside, due to a longer wire up to the processing unit, in this case the analogue high-impedance EOG signals pick up an increased amount of noise compared to the first option.

A second challenge was the decision for the type of electrodes and their mounting to the glasses frame. Wet electrodes are commonly used for EOG recordings and provide high quality trace pickup. They come with a built-in layer of conductive gel, which assures good conductivity between the skin and the electrodes to minimise signal artefacts. However, because of the gel layer, they need to be stuck to the face with an adhesive brim, which may be uncomfortable and potentially irritate the skin.

In contrast to wet electrodes, dry electrodes are more convenient as they allow for easy removal and reattachment to the skin. However, they can move on the skin or can loose contact. Proper signal acquisition therefore requires a mechanical mounting that assures permanent contact and constant pressure on the skin at any time. To accommodate for different head sizes the electrode mounting also needs to be adjustable. Even if good electrode-skin contact can be guaranteed, dry electrodes usually show lower signal stability. This is caused by a higher initial contact resistance, which decreases considerably over time thus resulting in increased signal drift (see Section 3.2.4).

The described challenges emphasise that a proper design of an integrated wearable sensor based on EOG is subject to a variety of sometimes contradictory technical constraints. We believe, however, that most of these can be solved in the future and therefore decided to optimise the first prototype of the wearable eye tracker for wearability and comfort accepting a mechanically more complex device and potentially lower signal quality.

3.3.3. Hardware Design

The final design consists of two components (see Figures 3.2, 3.3, and 3.4): Goggles with integrated electrodes and a signal processing unit (called WEPU, Wearable EOG Processing Unit).

The complete system weighs 188g (Goggles: 60g, WEPU: 78g, cable: 50g) and is powered by a 3.7V / 1500mAh Li-polymer battery attached to the WEPU. The total power consumption is 769mW when storing data on the MMC card and 828.4mW when streaming data to a remote computer using Bluetooth. This allows for up to 7.2 hours (MMC) and 6.7 hours (Bluetooth) of autonomous eye movement recording, respectively.
Figure 3.2: Two-part hardware architecture of the EOG-based wearable eye tracker with EOG amplification circuitry (EOG_h, EOG_v), accelerometer (ACC), light sensor (LIGHT), analog-digital converters (ADC), DSP, EEPROM, Bluetooth module (BT) and MMC card holder for data transmission and storage.

Mechanics

The Goggles contain the light sensor, dry EOG electrodes and a small analogue amplification board (see Figure 3.3). The light sensor is attached at the front of the frame in between both eyes pointing forward in line of incident light. Although a minimum of five EOG electrodes is required for recording eye motion (four for the two signal components and one for the reference signal), six electrodes are arranged around both eyes. This allows for flexibility if one of the eyes can not be measured due to poor signal quality. The electrodes are mounted on spring steel to ensure permanent skin contact and constant contact force. The spring steel is bent in such a way that the electrodes are placed flat on the skin. Because of the anatomy of the skull, this is particularly challenging for the electrodes above and below the eye, which need to be placed further away from the frame. Each electrode is connected to the amplification board with a shielded one core cable following the frame’s shape. The amplification board has a size of 42x15mm and is screwed onto the left side of the frame. Besides the analogue amplification circuit this board also contains an accelerometer for measuring head movements. The Goggles are connected to the WEPU with a shielded 14 core cable. The WEPU can be worn on the body, e.g. in a cloth bag fixed to one of the upper arms.

Electronics

The EOG signal is composed of a small voltage superimposed by a large offset voltage relative to the ground electrode above the right eye. The offset is mostly caused by stray electrical signals on the leads and therefore referred to as common-mode interference. If an electric circuit is able to efficiently reject this interference it has a high common-mode rejection ratio (CMRR). For signal amplification on the Goggles, we decided for low-power, medical
3.3. Design and Implementation of a Wearable Context-Aware EOG Sensor

Figure 3.3: Detailed view of the Goggles component: Dry electrodes mounted on spring steel (1) and screwed to the frame (2). Shielded one core cables (3) connect each electrode to a small analogue amplification board (4), which is connected with a shielded 14 core cable (5) to the WEPU. The pictures also show the light sensor attached at the front of the frame (6).

grade instrumentation amplifiers. They provide a very high CMRR rating of $110dB$ and $10^{10}\Omega$ input impedance and thus meet the requirements for a potentially robust signal acquisition. To further increase the CMRR, a Driven Right Leg (DRL) circuit [41] is implemented on the Goggles. This circuit measures the common mode potential and feeds its negative back into the body to actively cancel signal interference. Using this approach, we are able to achieve a CMRR of more than $105dB$.

The WEPU is the core signal processing unit of the system with a credit card size of $82\times56\,mm$ (see Figure 3.3). It is based on a 16-bit dsPIC from Microchip and contains dedicated 24-bit Delta-Sigma ADCs for each EOG channel, a Bluetooth and a MMC module and an EEPROM. The main advantage of the dsPIC compared to other microcontrollers commonly used on wearable sensors is its suitability for efficient real-time signal processing. It runs at $3V$ with 40 Million Instructions Per Second (MIPS). The ADCs are critical for fast analog-digital conversion with high resolution. To achieve a
Chapter 3: Wearable Eye Movement Sensing

Figure 3.4: Components of the EOG-based wearable eye tracker: the WEPU with credit card size (1), the Goggles (2) and the shielded 14 core cable (3). The pictures at the bottom show the Goggles worn by a person with the positions of the two horizontal (h) and vertical (v) dry electrodes, the light sensor (l) and the accelerometer (a) with direction of its axes (ACC_Y, ACC_Z).

resolution of 2.5\(\mu\)V/° for EOG signals with a dynamic range of 600mV the ADCs have to provide a minimum resolution of 18 bit. On the wearable eye tracker, the two ADCs allow the raw EOG signals to be processed with a sampling rate of up to 250Hz and a resolution of 20 bits noise-free, i.e. that can be distinctly resolved. Processed data can either be transmitted using Bluetooth or stored on the MMC for offline analysis. The EEPROM is used to store configuration data and parameters for the signal processing algorithms described in Section 3.4. Four LEDs and two buttons allow the user to access the functionality of the device.

3.3.4. Firmware

The dsPIC on the WEPU runs freeRTOS, an open-source real-time operating system devised for embedded systems. freeRTOS is configured to run in preemptive mode using predefined task priorities. Using an operating system does not only contribute to clean and well-structured code but also provides
3.3. Design and Implementation of a Wearable Context-Aware EOG Sensor

![Three-tier software architecture of the EOG-based wearable eye tracker](image)

**Figure 3.5:** Three-tier software architecture of the EOG-based wearable eye tracker with layers for hardware abstraction and the operating system (freeRTOS), access to external components (Device Layer), common routines (Library) and core functionality (Task Layer).

services such as interrupt handling and a scheduler, which eases development and later code maintenance. In addition, this allows us to eventually run a context recognition middleware and opens the possibility to integrate the device into multi-modal context recognition systems, physiological sensor networks or smart sensing environments. As a first step in this direction, we have implemented a driver for the Context Recognition Network (CRN) Toolbox [42], which enables the recording of EOG for eye movement analysis to be automatically synchronised with signals coming from body-worn accelerometers for activity recognition.

The firmware is composed of three layers (see Figure 3.5). Among these layers, the Hardware Abstraction Layer (HAL) accesses the hardware. It provides a number of interfaces to the upper layers thus hiding all low-level hardware access. The Device Layer (DEL) uses the HAL to provide functionality for components external to the DSP such as the Bluetooth and the MMC module. For the MMC, a custom file system has been implemented to allow for efficient data storage. The binary EOG data can afterwards be converted into a standard two-column format using a custom software. The core functionality of the firmware is provided by the following five freeRTOS tasks implemented in the Task Layer (TAL):

The Controller Task is in charge of the interface components, power management and task control. It processes events coming from the two push buttons and the LEDs, monitors the output of the charger IC and activates the stand-by mode. In this mode, it reduces the DSP clock speed, holds the Bluetooth module and MMC in reset state and powers down the ADCs.

The Packet Reader Task reads packets received from the Bluetooth module, calculates a checksum and stores valid packets into a receive buffer. If the task receives a configuration packet, it initiates the reconfiguration process during which the Sampler Task and Processing Task may select other types or a different order of the algorithms for EOG signal processing.

The Sampler Task retrieves data samples from the external 24-bit ADCs and uses the 12-bit ADC internal to the DSP to sample the accelerometer and light sensor signals. The data is then passed on to the Processing Task.

The Processing Task implements the core functionality of the eye tracker
as it takes samples from the Sampler Task, executes each building block of the signal processing cascade and passes processed samples on to the Communicator Task. The type and order of the processing algorithms applied to the signals within the cascade can freely be chosen by the user (see Figure 3.6 for the default).

The Communicator Task reads data from the sample queue, builds a data packet and sends the packet either to the MMC controller or the Bluetooth module depending on the output medium selected by the user. It is also in charge of operating these peripherals.

Additionally, a separate Library contains functionality that is shared by these tasks such as the CRC routines.

3.4. EOG Signal Processing for Context Sensing

In this section, we describe the signal processing cascade implemented on the wearable eye tracker (see Figure 3.6). The processing is tailored to the particular needs of recognising context information from EOG signals. It aims at removing artefacts and noise from the EOG signals. Additionally, it aims at providing robust detection of saccades and blinks for eye movement event encoding and the recognition of so-called eye gestures, which consist of several consecutive movements. We first describe the denoising filter and the algorithm for compensation of EOG signal artefacts caused by walking. We continue with a description of the algorithms for the detection of blinks and saccades, removal of blinks, eye movement event encoding and eye gesture recognition.

3.4.1. Adaptive Filtering

Denoising

Raw EOG signals are corrupted with noise from the following sources:

- Noise caused by the residential powerline, usually referred to as mains hum
- Noise introduced by the measurement circuitry, the electrodes, wires, etc.
- Noise from other physiological sources interfering with EOG such as electromyographic (EMG) signals
- Noise due to simultaneous physical activity, which may cause the electrodes to lose contact or move on the skin

As a first step in the signal processing cascade, noise reduction is necessary to improve the signal for the following processing blocks. In contrast to other physiological signals such as those from electrocardiography (ECG),
3.4. EOG Signal Processing for Context Sensing

Figure 3.6: Flowchart showing the building blocks of the signal processing cascade implemented on the wearable eye tracker: horizontal and vertical EOG (EOG$_h$, EOG$_v$) and acceleration (ACC$_Y$, ACC$_Z$, cf. Figure 3.4) signals as input, adaptive filtering using a median filter and eye movement event detection. As output, the system returns the detected eye movement events as well as the processed EOG data (EOG$_{hp}$, EOG$_{vp}$).

Eye movements are usually non-repetitive, which make the generated EOG signals unpredictable. This mostly prohibits to apply optimised algorithms that make use of structural and temporal knowledge about the expected signal to improve denoising quality. Furthermore, EOG signals exhibit characteristics that need to be preserved for further signal analysis. The signal edges need to be retained to allow for saccade detection and analyse eye movement dynamics. Signal amplitudes need to be preserved to allow to distinguish between different types and directions of saccadic eye movements. Denoising filters must not introduce artificial signals that may accidently be interpreted as eye movements in subsequent signal processing steps.

To identify a suitable approach for denoising we evaluated different algorithms, such as a standard low-pass filter, a filter based on wavelet shrinkage denoising (see [43] for details) and a standard median filter applied on a signal window with fixed length. From our experiments we found that the median fil-
ter performed best as it preserved edge steepness of saccadic eye movements, retained signal amplitudes and did not introduce any artificial signals. Furthermore, as the median filter is computationally light-weight it is well suited for online signal processing on a DSP. However, it is crucial to choose an appropriate window size to reduce noise without removing important EOG signal parts.

**Motion Artefact Compensation**

As EOG is measured with body-worn sensors, motion causes artefacts in the signals and affects eye movement detection. Walking is a common activity in everyday life. Thus, walking serves as a good test bench for investigating artefacts induced by body motion. Analyses showed that artefacts in the EOG signals occur periodically according to the step frequency. A median filter with fixed window size fails to eliminate these artefacts for different persons and walking speeds. A parameter sweep on the window size using example data recorded from several participants revealed that the optimal size is strongly related to the temporal step length. Therefore, we use an algorithm implementing an adaptive filter. The idea is to exploit the repetitive characteristic of walking and adapt the window size of the median filter to the step length as long as walking activity is detected (see Figure 3.7).

An approach for detecting footsteps using data from a head-mounted accelerometer has been described in [44]. To detect walking activity, we implemented an extended version of the algorithm. Our algorithm first analyses the vertical axis of the goggle-mounted accelerometer (ACC\textsubscript{Y}, cf. Figure 3.4). If the corresponding signal exceeds a defined threshold, the algorithm tries to detect steps by searching for zero-crossings of the first derivative of the low-pass-filtered acceleration data of the horizontal axis (ACC\textsubscript{Z}, cf. Figure 3.4). Walking is assumed as long as such steps are detected. In order to smooth out variations in walking style for different participants, the step length is calculated on the basis of three consecutive step movements (e.g. right - left - right) separately for the left and the right leg. By calculating the length continuously for each step, the algorithm can adapt to different persons and walking speeds. For softer adaptation, only small increments are applied (see Figure 3.7). If walking activity is not detected anymore, the window size is progressively set towards its default value.

In [13] we evaluated the adaptive filter in a mobile setting. The experimental scenario involved participants to perform different eye movements while standing and walking down a corridor. The expected eye movements were shown on a head-up display (HUD) with a defined order and timing. We recorded five male participants between the age of 21 and 27 totalling roughly 35 minutes of recording with walking activity accounting for about 22 minutes. As the mobile setting did not allow to record a ground truth, we decided to do a comparison to a standard median filter with fixed window size to assess a relative performance measure. Figure 3.8 shows a boxplot for the total number of detected saccades in the horizontal EOG signal component. Each box summarises the statistical properties of the data of the participants: The horizontal red lines in each box indicates the median and the upper and
Figure 3.7: Adaptive filter for artefact compensation while walking slowly (a) and fast (b). Vertical acceleration signal ($\text{ACC}_Y$) and threshold (dashed line) for detecting walking activity. Horizontal acceleration signal ($\text{ACC}_Z$) and first derivative for calculating the step length. The window size used by the median filter and tuned to the walking pace is shown at the bottom.

lower quartiles. The vertical dashed lines indicate the data range, points outside their ends are outliers. Boxes are plotted for the following cases: stationary and raw signal, stationary and fixed median filter, stationary and adaptive filter, walking and raw signal, walking and fixed median filter, walking and adaptive filter. The single solid horizontal line indicates the expected number of saccades defined by the experimental procedure. We found that the adaptive filter was able to reduce signal artefacts caused by walking activity by up to 80% in the horizontal, and up to 60% in the vertical signal component.

3.4.2. Blink Detection

Blinks need to be detected in the vertical EOG signal component for two reasons: For certain HCI applications they may provide important input and a more versatile input alphabet. For applications focused on eye movement detection, blinks need to be detected and removed because their characteristics are very similar to those of saccadic eye movements. This can affect subsequent signal processing steps and eventually render robust eye movement analysis impossible.

We evaluated different algorithms with special attention to a real-time implementation on the DSP. The methods considered included continuous
wavelet transform (CWT-BD), velocity threshold analysis \cite{34}, Haar wavelet decomposition \cite{45} and template matching. Similar to the algorithm described in the following section, CWT-BD uses thresholding of wavelet coefficients for blink detection. The approach based on template matching works as follows: First, a blink template is created using manually cut equally-sized raw signal segments of 10 blinks from different persons, vertically shifted by their median and aligned at their peaks. To create the template, the mean at each sample point over all segments is calculated. Afterwards, blinks are detected by shifting this template over the vertical EOG signal component by following a sliding window approach. In each step, the Euclidean distance between the template and the signal segment of the current window is computed as a similarity metric. If the distance is below a defined threshold, i.e. the similarity between the template and the current segment is high, a blink event is recorded.

We evaluated each of these algorithms on nine EOG signals from different participants (see Table 3.1). The signals contained a total of 105 blinks with different blink amplitudes (80 large, 25 small). For each algorithm, true positive ($TP$), false positive ($FP$) and false negative ($FN$) counts were taken to calculate precision ($TP/TP+FP$) and true positive rate (recall) ($TP/TP+FN$) values. From the results we found that both CWT-BD and template matching were able to recognise all large blinks while the velocity algorithm and the

---

**Figure 3.8:** Boxplot for the total number of detected saccades in the horizontal EOG signal component with fixed thresholds over all participants: stationary/ raw (a), stationary/ fixed median filter (b), stationary/ adaptive filter (c), walking/ raw (d), walking/ fixed median filter (e), walking/ adaptive filter (f).
### 3.4. EOG Signal Processing for Context Sensing

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>CWT-BD</td>
<td>104</td>
<td>6</td>
<td>1</td>
<td>95%</td>
<td>99%</td>
</tr>
<tr>
<td>Template Matching</td>
<td>101</td>
<td>1</td>
<td>4</td>
<td>99%</td>
<td>96%</td>
</tr>
<tr>
<td>Velocity Threshold</td>
<td>91</td>
<td>0</td>
<td>14</td>
<td>100%</td>
<td>87%</td>
</tr>
<tr>
<td>Haar Wavelet</td>
<td>76</td>
<td>6</td>
<td>29</td>
<td>93%</td>
<td>72%</td>
</tr>
</tbody>
</table>

**Table 3.1**: Combined true positive (TP), false positive (FP), false negative (FN), precision and true positive rate (recall) values for blink detection using different algorithms. The results were calculated based on nine EOG datasets from different participants containing a total of 105 blinks of which 80 had large and 25 small amplitudes.

Haar wavelet performed worse. Although CWT-BD performed slightly better on blinks with smaller amplitude, for the wearable eye tracker we decided to use the template matching algorithm due to its higher precision rate. The reason is that a small number of blinks that are not recognised correctly have less negative impact on eye movement detection than the same number of actual movement events that get detected as blinks and removed due to wrong classification.

### 3.4.3. Saccade Detection

For saccade detection we developed the so-called *Continuous Wavelet Transform - Saccade Detection* (CWT-SD) algorithm, which can be efficiently implemented for real-time processing on the DSP. The CWT-SD first computes the continuous 1-D wavelet coefficients from the signal at scale 20 using Haar wavelets. Saccades are detected for all samples where the absolute value of the coefficient vector exceeds a defined threshold. The direction and size of a saccade is given by the sign of the first derivative and the maximum value of the corresponding EOG signal amplitude.

### 3.4.4. Blink Removal

For blink removal, the streams of saccade and blink events are analysed in parallel. Three cases need to be distinguished to maintain essential signal characteristics such as saccade amplitude and slope required for eye movement detection.

*Presaccadic blinks* are caused by blinks that share their last edge with a saccade. Presaccadic blinks are removed by replacing the blink interval with the signal value at the beginning of the blink.

*Intersaccadic blinks* usually occur during slow eye movements or fixation periods. This type of blink is removed by replacing its interval with a linear interpolation between the value at the beginning and the value at its end.
Postsaccadic blinks are blinks that immediately follow a saccade and thus share their first edge with it. For removal, the blink interval is replaced with the signal value at the end of the blink.

### 3.4.5. Eye Gesture Recognition

The idea of combining a sequence of distinct relative eye movements to create more complex gestures was introduced in [22] for a video-based eye tracker. We follow a similar approach for the continuous recognition of eye gestures based on EOG (see Figure 3.9). Our algorithm takes the streams of saccade events for the horizontal and the vertical EOG signal component as its input. It maps these saccades to eye movements with basic, intermediate and diagonal directions and finally encodes them into a combined string sequence. Basic directions are left, right, up and down (L, R, U, D). Diagonal eye movements (1, 3, 7, 9) are characterised by simultaneous saccades with similar signal amplitudes. If the saccades have different signal amplitudes, the corresponding eye movements are called intermediate (e.g. V, M).

The algorithm for eye movement detection works as follows: It first checks for simultaneous saccade events in both EOG signal components within a time window of 0.06 s given by eye physiology. If no simultaneous saccade events are detected, the single saccade is directly mapped to the symbol of the corresponding basic eye movement. If two such events are detected within the time window, a non-basic eye movement has occurred. The algorithm then uses the corresponding saccades’ directions and amplitudes to combine both events into the appropriate symbol (c.f. Figure 3.9): Two saccades with equally large amplitudes are merged to the symbol exactly in between (e.g. symbols R and D are mapped to symbol 3). If the saccades’ amplitudes differ by more than 50% the saccades are merged to the closest neighbouring symbol (e.g. symbols r and D are mapped to symbol V). This scheme encodes each eye movement into a distinct event symbol, thus merges both EOG signal components into one string sequence.

To recognise eye gestures consisting of several consecutive movements, the resulting string sequence is scanned for eye movement patterns following a string matching approach: For matching, the sequence is continuously compared with templates representing all gestures required by the specific application. For each template, the edit distance between the templates and the segment is calculated. To allow for variability inherent in the eye gestures, the edit distance between two symbols of which one represents an intermediate direction is set to zero if the angle between them is smaller than or equal to 22.5°. If one of the templates exactly matches the current segment of the string sequence (i.e. the edit distance is zero), the corresponding eye gesture is recognised.
Figure 3.9: Eye movement encoding from horizontal (EOG_h) and vertical (EOG_v) EOG signals for gesture 3U1U: Windows marked in grey with distinct saccade events (R, L, D, U) detected in the horizontal and vertical signal component (a), mapping to basic (U) and diagonal (1, 3) eye movements (b) and final merging into a combined sequence of symbols (c). The circle in the middle shows all possible symbols for saccade event mapping.

3.5. Case Studies

3.5.1. Activity Recognition

The aim of the experiment conducted in [8] was to recognise the reading activity of people in transit in an everyday environment using a wearable EOG system. We defined a scenario of travelling to and from work containing a semi-naturalistic set of reading activities. It involved eight participants reading text while being engaged in a sequence of activities such as sitting at a desk, walking along a corridor, walking along a street, waiting at a tram stop and riding a tram. For this two-class classification problem, we evaluated three recognition algorithms - string matching and two variants of Hidden Markov Models (HMMs), mixed Gaussian and discrete - on a dataset of about six hours.
Chapter 3: Wearable Eye Movement Sensing

Figure 3.10: ROC curves showing a performance comparison for the recognition of reading activity between string matching (STR), Gaussian HMM (HMM) and discrete HMM (D-HMM). For the HMMs, both the participant-dependent and the participant-independent results are shown.

Two leave-one-out training schemes were used for the HMMs: participant-dependent only using the calibration data from the participant being tested and participant-independent using calibration data only from other participants. The different methods were compared across a sweep of their main parameters. The resulting Receiver Operating Characteristics (ROC) curves for the participant-dependent and the participant-independent case are shown in Figure 3.10. These plot true positive rate against false positive rate (FPR) \( \frac{TP}{FP+TN} \). Best case results approach the top left corner while worst case (which means random) follow the diagonal.

The ROC shows that string matching outperforms the HMMs. At its “best”, we were able to recognise reading activity over all participants using string matching with a recall of 71.0% and FPR of 11.6% (total accuracy 80.2%). The mixed Gaussian returns a lower best-case at recall 62.0%, FPR 24.0% and accuracy 68.9% while the worst performing algorithm is the discrete HMM. The experiment has shown that wearable EOG is a feasible approach for recognising reading in daily-life scenarios and is robust across an example set of activities for different participants.
3.5.2. Human-Computer Interaction

In [13] we evaluated the wearable eye tracker for human-computer interaction. To investigate the use of explicit eye gestures as a control input, we developed a desktop computer game consisting of eight different game levels. In each game level, participants had to repeatedly perform a defined eye gesture as fast as possible until the first successful try (see Table 3.2). To reach a high score, wrong eye movements, i.e. movements that were not part of the expected gesture, had to be minimised. In contrast to a previous study [22], movements of the head and the upper body were allowed at any time. We collected data from eleven participants - two female and nine male - between the ages of 24 and 64.

The average performance over all participants is given in Table 3.3 which shows the times $T_T$ and $T_S$, the time ratio $T_S/T_T$ and the accuracy $Acc$ to perform each of the eight gestures. $T_T$ denotes the total time the participants spent trying to complete each of the gestures while the success time $T_S$ only measures the time spent on all successful attempts. What can be seen from the table is that the participants were able to achieve an average eye gesture accuracy of 87%. Although the total time $T_T$ increased significantly for more complex gestures such as DDR7L9, the success time $T_S$ was comparable for all gestures. This highlights the fact that the eyes can be used as a versatile and fast input modality across gestures with different complexity.

Table 3.2: Eye gestures of increasing complexity and their string representations used for interaction in a computer game (cf. Figure 3.9). The grey dot denotes the start and the arrows the order and direction of each eye movement.

<table>
<thead>
<tr>
<th>Gesture</th>
<th>String Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1R</td>
<td>DRUL</td>
</tr>
<tr>
<td>RDLU</td>
<td>RLRLRL</td>
</tr>
<tr>
<td>3U1U</td>
<td>DR7RD7</td>
</tr>
<tr>
<td>1397</td>
<td>DDR7L9</td>
</tr>
</tbody>
</table>

3.6. Discussion

3.6.1. On Wearable Eye Tracking Using EOG

From our studies we found that the wearable eye tracker described in this work is a robust sensor for recording eye motion. In contrast to common sys-
Chapter 3: Wearable Eye Movement Sensing

<table>
<thead>
<tr>
<th>Gesture</th>
<th>(T_T[m\text{s}])</th>
<th>(T_S[m\text{s}])</th>
<th>(T_S/T_T)</th>
<th>Acc[%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1R</td>
<td>3370</td>
<td>2890</td>
<td>0.858</td>
<td>85</td>
</tr>
<tr>
<td>DRUL</td>
<td>4130</td>
<td>3490</td>
<td>0.845</td>
<td>90</td>
</tr>
<tr>
<td>RDLU</td>
<td>3740</td>
<td>3600</td>
<td>0.963</td>
<td>93</td>
</tr>
<tr>
<td>RLRLRL</td>
<td>6680</td>
<td>5390</td>
<td>0.807</td>
<td>90</td>
</tr>
<tr>
<td>3U1U</td>
<td>4300</td>
<td>3880</td>
<td>0.902</td>
<td>89</td>
</tr>
<tr>
<td>DR7RD7</td>
<td>12960</td>
<td>5650</td>
<td>0.436</td>
<td>83</td>
</tr>
<tr>
<td>1379</td>
<td>6360</td>
<td>3720</td>
<td>0.585</td>
<td>84</td>
</tr>
<tr>
<td>DDR7L9</td>
<td>25400</td>
<td>5820</td>
<td>0.229</td>
<td>83</td>
</tr>
<tr>
<td>Average</td>
<td>8368</td>
<td>4305</td>
<td>0.712</td>
<td>87</td>
</tr>
</tbody>
</table>

Table 3.3: Average performance for the different gestures over all participants. \(T_T\) is the total time spent to complete the gesture and \(T_S\) the success time spent only on successful attempts. The accuracy \(Acc\) is the ratio of eye movements resulting in a correct gesture to the total number of movements performed until success.

The device uses an EOG-based measurement technique that enables the unobtrusive implementation as goggles. The main advantage of EOG is the fact that the person only has to wear light-weight equipment. Our studies show that this contributes to the person feeling unconstrained and allows for natural behaviour and unrestricted physical activity in daily life settings. We plan to carry out additional experiments on human factors to further evaluate the goggles with respect to unobtrusiveness and user acceptance particularly for special user groups such as disabled or elderly people. Another advantage is that EOG processing requires less computational power than video due to lower data rates. This enables an embedded and low-power design and results in low data storage requirements. Both are crucial prerequisites for long-term data collection and real-time signal processing in mobile daily-life settings.

Our system particularly addresses challenges related to wearability, signal artefacts caused by physical activity and eye motion analysis. This is possible with a combination of a special mechanical design, adaptive signal processing and optimised algorithms for eye movement detection. An interesting question for future work is how the adaptive filter performs for people with different walking styles (e.g. elderly, persons limping, etc.) and how it can be extended to address different types of movement artefacts. A remaining issue with the current prototype is that dry EOG electrodes require permanent skin contact. Poor placement of electrodes and signal artefacts due to electrode movements were the reasons for many of the problems in our work. We believe that these may eventually be solved by developing special-purpose EOG goggles with a mechanically improved mounting that is tailored to assure permanent skin contact and proper electrode placement.
Baseline drift is an issue for wearable EOG recordings in particular if dry electrodes are used. It is for this reason that accurate eye-gaze tracking is difficult to achieve. Nevertheless, as we demonstrated in this paper, eye motion is a rich source of information on user activity and context that complements common pinpoint tracking. The development of novel electrodes is still a very active topic of research (for example see [46]). Eventually, dry electrodes that allow to record drift-free signals would allow EOG to be implemented for eye-gaze tracking.

3.6.2. On the Case Studies

The first case study shows that wearable EOG is a feasible approach for recognising reading activity in daily-life scenarios and is robust across an example set of simultaneous physical activities. We were able to detect reading activities over all subjects with a top recognition rate of 80.2%. This result raises the question of whether different reading behaviours can be detected automatically. A “reading detector” could enable novel attentive user interfaces that take into account aspects such as user interruptability and level of task engagement.

Results from the second case study show that EOG is a robust modality for HCI applications that can be efficiently processed to recognise eye gestures consisting of several consecutive eye movements. While using the eyes as a control input was quickly learned, 30% of the subjects reported of having had problems to stay concentrated during the game. However, fatigue is an intrinsic problem not only for eye gestures but also for common input modalities such as speech or hand gestures. Eye gestures outperform these modalities if the hands can not be used (e.g. during driving or while working on the computer) or if speech input is not possible (e.g. for privacy reasons or in very silent or very noisy surroundings).

3.7. Conclusion and Outlook

In this work, we have demonstrated an autonomous EOG-based eye tracker and context recognition system integrated into goggles. We have shown that this unobtrusive device is applicable to different people and works in a wide range of applications. Its embedded and self-contained design allows for wearable sensing and online analysis of eye motion and extends these to everyday environments. This enables context-aware feedback, which is a key aspect to smart wearable assistants and smart environments. Furthermore, we have shown that the main characteristics of eye motion can be captured in an efficient way, which promises a range of new context-aware applications. EOG-based eye input allows for versatile human-computer interaction and may eventually provide new means of light-weight interaction for mobile settings.

The movement patterns our eyes follow in daily routine reveal much about what we are doing as well as what we intend to do. Our long-term objective is to investigate how much information eye motion can provide about
Chapter 3: Wearable Eye Movement Sensing

the user’s activity and context. By connecting several eye trackers, concurrent eye movement recordings for a group of people and distributed activity recognition may become possible. In addition, we also plan to investigate unconscious eye movements, which are the result of cognitive processes in the human brain. These processes are related to external aspects such as the user’s activity or environment, but also to internal aspects of visual perception such as memory [11] and learning [12]. The analysis of eye motion thus may allow to deduce these aspects, which would give important input for future context-aware systems.

Eventually, eye motion may be used as a new sensing modality for activity recognition, context-awareness and mobile HCI applications, providing access to underlying cognitive processes not accessible with current sensing modalities.
Bibliography


Chapter 3: Wearable Eye Movement Sensing


3.7. Conclusion and Outlook


3.7. Conclusion and Outlook


4

Relative Eye Movements for Human-Computer Interaction

Andreas Bulling, Daniel Roggen and Gerhard Tröster

Full publication title: It’s in Your Eyes - Towards Context-Awareness and Mobile HCI Using Wearable EOG Goggles

DOI: 10.1145/1409635.1409647
Abstract

In this work we describe the design, implementation and evaluation of a novel eye tracker for context-awareness and mobile HCI applications. In contrast to common systems using video cameras, this compact device relies on Electrooculography (EOG). It consists of goggles with dry electrodes integrated into the frame and a small pocket-worn component with a DSP for real-time EOG signal processing. The device is intended for wearable and standalone use: It can store data locally for long-term recordings or stream processed EOG signals to a remote device over Bluetooth. We describe how eye gestures can be efficiently recognised from EOG signals for HCI purposes. In an experiment conducted with 11 participants playing a computer game we show that 8 eye gestures of varying complexity can be continuously recognised with equal performance to a state-of-the-art video-based system. Physical activity leads to artefacts in the EOG signal. We describe how these artefacts can be removed using an adaptive filtering scheme and characterise this approach on a five-participant dataset. In addition to explicit eye movements for HCI, we discuss how the analysis of unconscious eye movements may eventually allow to deduce information on user activity and context not available with current sensing modalities.

4.1. Introduction

Activity recognition is a key mechanism to devise context-aware systems for mobile and ubiquitous computing. The recognition of physical activity in mobile situations, for example motion from body worn sensors, has been extensively studied. However, context-awareness encompasses more than mere physical activity. Aspects like user attention and intentionality remain mainly unexplored as these cannot be picked-up by sensors usually deployed in today’s wearable and pervasive computing scenarios. A rich source of information about the state of the user can be found in the movement of the eyes. This includes information related to the users’ activities and environments and their cognitive processes of visual perception such as attention [1], saliency determination [2], visual memory [3] and perceptual learning [4]. As part of an ongoing project we seek to investigate up to which extent the effects of these processes on eye motion can be exploited to enable new kinds of context-aware applications. Explicit eye movements performed by the user can directly be used for HCI input. Mobile attentive user interfaces (MAUIs) may also infer user intention and activity or provide assistance by analysing implicit eye movements.

In earlier work, we proposed Electrooculography (EOG) as a novel measurement technique for wearable eye tracking and the recognition of user activity and attention in mobile settings [5]. EOG, in contrast to well established
vision-based eye tracking\(^1\), is measured with body-worn sensors and can be implemented as a wearable system. In this paper we describe how unobtrusive EOG recordings can be implemented using electrodes integrated into glasses and the signals processed in real-time on a light-weight device worn on the body. We further extend this work by demonstrating how “eye gestures” can be recognised from EOG for HCI purposes. The specific contributions of this work are (1) the design and implementation of a wearable EOG-based eye tracker for long-term recordings in daily life implemented as goggles, (2) an algorithm for continuous recognition of complex eye gestures from EOG signals, (3) a characterisation in a computer game where eye gestures are used for HCI and (4) the development and evaluation of a new method for artefact removal from EOG signals caused by walking.

4.2. Related Work

4.2.1. Activity Recognition

Logan et al. studied activity recognition with a person living in an environment instrumented with a large number and variety of common sensors \([6]\). They found that among all activities, reading was one of the most difficult to detect. They concluded that in order to catch all types of physical activity in daily-life scenarios, novel sensors and algorithms need to be developed.

A growing number of researchers investigate movements of the eyes during daily activities. Important advances have been made to understand how the human brain processes visual tasks \([7]\), how vision contributes to the organisation of active tasks in everyday life \([8]\) and how eye, head, and hand movements are coordinated temporally \([9]\). However, eye movements have not been used for activity or context recognition so far.

4.2.2. Eye-based Interaction

Eye gaze recorded using vision has long been investigated as a means to interact with a computer. Most HCI work has focused on direct manipulation of user interfaces (e.g. Zhai et al. \([10]\)). Qvarfordt et al. explored an interactive human-computer dialogue system which used eye gaze patterns to sense the users’ interests \([11]\). Drewes et al. proposed to use eye gestures consisting of several consecutive movements to implement new ways of human-computer interaction \([12]\). They showed that these gestures are insensitive to accuracy problems, immune against calibration shift and do not exhibit the “Midas touch” problem (for details see \([13]\)).

\(^1\)With “eye tracking” we understand the recording of eye movements to analyse general characteristics of eye motion such as movement dynamics or statistical properties. In contrast, the term “gaze tracking” describes the process of recording eye movements with the goal of calculating and tracking eye-gaze direction.
Chapter 4: Relative Eye Movements for Human-Computer Interaction

4.2.3. EOG-based Interfaces

Patmore et al. developed a system intended to provide a pointing device for people with physical disabilities [14]. Basic signal characteristics such as saccades, fixations and blinks have been used for controlling a robot with the person remaining stationary [15]. For mobile scenarios, similar characteristics were used to operate a wearable computer system for medical caregivers [16]. All of these studies show that EOG can be implemented as an easy to operate and reliable interface. While these systems use basic eye movements as an input, they do not make use of movement sequences to implement a more versatile input modality.

4.2.4. Eye Tracking Devices

The common method to track eye gaze in natural environments are systems based on video cameras. A number of commercial eye trackers are available of which some are targeted at mobile use, e.g. the Mobile Eye from Applied Science Laboratories (ASL) or the iView X HED from SensoMotoric Instruments (SMI). Nevertheless, these systems still require bulky headgear and additional equipment to process the video streams. To our knowledge, at this stage no solution for video-based eye tracking exists that is unobtrusive enough to allow for long-term recordings while leaving the wearer unaffected during physical activity.

Several researchers investigate novel electrode configurations for wearable EOG recordings. A gaze detector which uses EOG electrode arrays mounted on ordinary headphones was proposed by Manabe et al. [17]. While this approach might be less obtrusive than electrodes stuck to the face, it turned out to raise other issues: Low signal-noise ratio (SNR) and poor separation of the movement components. Vehkaoja et al. presented a light-weight head cap with electrodes embroidered of silver coated thread [18]. A small device integrated into the cap allows for wireless data transmission. As yet it is still to be evaluated in operation.

4.3. Wearable Electrooculography

4.3.1. Eye Movement Characteristics

The eyes are the origin of a steady electric potential field which can be described as a dipole with its positive pole at the cornea and its negative pole at the retina. The magnitude of this so-called corneoretinal potential (CRP) lies in the range of 0.4mV to 1.0mV. The CRP is the basis for a signal measured between two pairs of electrodes commonly placed above and below, and on the left and right side of the eye, the so-called Electrooculogram (EOG).

If the eyes move from the centre position towards the periphery, the retina approaches one of the electrodes while the cornea approaches the opposing one. This results in a change in the electric potential. Inversely, eye movements can be tracked by analysing these changes in the EOG signal. The electrode pairs capture the horizontal and the vertical component of eye
4.3. Wearable Electrooculography

motion. This requires good electrode placement, i.e. on the eyes’ horizontal and vertical axes of motion, as otherwise increased crosstalk between both components occurs. Usually, the signal amplitudes resulting from horizontal eye movements are larger than those from vertical movements. Therefore, crosstalk affects the vertical component more severely. Signal crosstalk poses problems on robust detection of eye movement events and eye gaze tracking for which both components need to be analysed simultaneously.

In the human eye, only a small central region of the retina, the fovea, is sensitive enough for most visual tasks. This requires the eyes to move constantly as only small parts of a scene can be perceived with high resolution. Simultaneous movements of both eyes in the same direction are called saccades. Typical characteristics of saccadic movements are \(400^\circ/s\) for the maximum velocity, \(20^\circ\) for the amplitude and 80ms for the duration. **Fixations** are static states of the eyes during which gaze is held at a specific location. Humans typically alternate saccadic eye movements and fixations while perceiving their environment.

### 4.3.2. Design and System Architecture

The wearable eye tracking device was designed to fulfil the following requirements:

- **Wearable and light-weight** to achieve a convenient and unobtrusive implementation and minimise user distraction.

- **On-board data storage and low-power** to allow for autonomous long-term recordings in daily life.

- **Real-time capability** to be able to perform online signal processing directly on the device.

- **Acceleration and light sensors** to compensate for artefacts caused by physical activity and changes in ambient light [19].

#### Hardware

The hardware is made of two components (see Figure 4.1): **Goggles** with integrated electrodes and a signal processing unit (called **Pocket**) with a credit card size of 82x56mm. The **Pocket** can be worn on the body, e.g. in a cloth bag fixed to one of the upper arms (see Figure 4.7). The system weighs 208g and is powered by a 3.7V / 1500mAh Li-polymer battery attached to the **Pocket** which allows for more than 7 hours of mobile eye movement recording. Raw EOG signals can be recorded on two channels with a sampling rate of up to 250Hz and a resolution of 20 bits noise-free².

---

²The noise-free resolution of an ADC is the number of bits of resolution beyond which it is impossible to distinctly resolve individual outputs.
Figure 4.1: Two-part hardware architecture of the eye tracker with EOG amplification (EOG_X, EOG_Y), accelerometer (ACC), light sensor (LIGHT), DSP, analog-digital converters (ADC), EEPROM, Bluetooth module (BT) and MMC card holder.

The Goggles contain dry EOG electrodes and a small analogue amplification circuit board with a size of 42x15mm attached to the glasses frame. Four electrodes are arranged around the left eye and mounted on flat springs to achieve good skin contact. The EOG signal is composed of a small voltage superimposed by a large offset voltage relative to the ground electrode above the right eye. The offset is mostly caused by stray electrical signals on the leads and therefore referred to as common-mode interference. If an electric circuit is able to efficiently reject this interference it has a high common-mode rejection ratio (CMRR). To increase the CMRR, a Driven Right Leg (DRL) circuit [20] is implemented on the Goggles. Briefly, this circuit measures the common mode noise and feeds its negative back into the body to actively cancel the interference. Finally, an accelerometer and a light sensor are attached to the component with the latter pointing forward in line of incident light (see Figure 4.2).

The Pocket is the core signal processing unit of the system. It is based on a dsPIC micro-controller and contains two 24-bit analog-digital converters (ADC), a Bluetooth and a MMC module and an EEPROM. EOG signals coming from the ADCs are processed in real-time and can either be transmitted using Bluetooth or stored on the MMC. The EEPROM is used to store configuration data and parameters for the signal processing algorithms. Four LEDs and two buttons provide a simple interface which allows the user to access the basic functionality of the device.

Software

The dsPIC on the Pocket runs freeRTOS, an open-source real-time operating system devised for embedded systems. freeRTOS is configured to run in preemptive mode using predefined task priorities. The firmware is composed of
4.4. EOG Signal Processing

Figure 4.2: Components of the EOG-based wearable eye tracker: armband with cloth bag (1), the *Pocket* (2), the *Goggles* (3) and dry electrodes (4). The pictures to the right show the *Goggles* worn by a person with the positions of the two horizontal (h) and vertical (v) electrodes, the light sensor (l) and the accelerometer (a).

Figure 4.3: Three-tier software architecture of the eye tracker with layers used for hardware abstraction and task management by the operating system, access to external components and core functionality.

three layers (see Figure 4.3). Among these layers, the *Hardware Abstraction Layer* (HAL) accesses the hardware. It provides a number of interfaces to the upper layers thus hiding all low-level hardware access. The *Device Layer* (DEL) uses the HAL to provide functionality for components external to the DSP such as the Bluetooth and the MMC module. The core functionality of the firmware is provided by 5 freeRTOS tasks which form the *Task Layer* (TAL). A separate *Library* (LIB) contains functionality which is shared by these tasks such as the CRC routines.

4.4. EOG Signal Processing

Eye gesture recognition is based on the detection of consecutive saccades which by their order and direction define the type of eye gesture. These saccades need to be detected in the continuous vertical and horizontal EOG signal streams. Blinks need to be removed because their characteristics are very
similar to those of vertical eye movements and would affect gesture recognition. In this section we describe the processing steps required for the detection of blinks and saccades and the removal of blinks. We then describe the algorithms for eye gesture recognition and compensation of EOG signal artefacts induced by walking.

4.4.1. Blink Detection

We detect blinks with a template matching approach: First, a blink template is created using manually cut equally-sized raw signal segments of 10 blinks from different persons, vertically shifted by their median and aligned at their peaks. To create the template, the mean at each sample point over all segments is calculated. Afterwards, blinks are detected by shifting this template over the vertical EOG signal component by following a sliding window approach. In each step, the Euclidean distance between the template and the signal segment of the current window is computed as a similarity metric. If the distance is below a defined threshold, i.e. the similarity between the template and the current segment is high, a blink event is recorded.

4.4.2. Saccade Detection

For saccade event detection we developed the so-called Continuous Wavelet Transform - Saccade Detection (CWT-SD) algorithm [21]: The CWT-SD first computes the continuous 1-D wavelet coefficients from the signal at scale 20 using Haar wavelets. A saccade event is detected for all samples where the absolute value of the coefficient vector exceeds a threshold. The direction and size of a saccade is given by its sign and amplitude.

4.4.3. Blink Removal

For blink removal, the streams of saccade and blink events are analysed in parallel. Blinks without simultaneous saccades are directly removed from the signal. In case of a simultaneous saccade, three cases need to be distinguished to maintain the essential signal characteristics required for eye gesture recognition:

*Presaccadic blinks* are caused by blinks which share their last edge with a saccade. Presaccadic blinks are removed by replacing the blink interval with the signal value at the beginning of the blink.

*Intersaccadic blinks* usually occur during slow eye movements or fixation periods. This type of blink is removed by replacing its interval with a linear interpolation between the value at the beginning and the value at its end.

*Postsaccadic blinks* are blinks which immediately follow a saccade and thus share their first edge with it. For removal, the blink interval is replaced with the signal value at the end of the blink.
Figure 4.4: Eye movement event encoding from horizontal and vertical EOG signals for gesture 3U1U: Windows marked in grey with distinct eye movement events detected in the horizontal and vertical signal component and final mapping to basic (U) and diagonal (3, 1) movements. The top right corner shows the symbols representing the possible directions for eye movement encoding.

4.4.4. Eye Gesture Recognition

The idea of combining distinct relative movements to more complex eye gestures was introduced in [12] for a video-based eye tracker. We follow a similar approach for continuous recognition of eye gestures based on EOG: Our algorithm takes the streams of saccade events for the horizontal and the vertical signal component as its input. It distinguishes between eye movements in basic, diagonal and intermediate directions (see Figure 4.4, top right corner): Basic directions are left, right, up and down (L, R, U, D). Diagonal and intermediate eye movements are characterised by simultaneous saccades in both signal components but different angles (e.g. I, 9, J).

For each saccade event the algorithm uses a time window of 0.06 s to check for a second event in the other component. If such an event is detected a diagonal eye movement has occurred. The algorithm then uses the saccades’ amplitudes and signs to combine and map them to the appropriate diagonal direction. If a second event is not found within the given time window the initial saccade is directly mapped to the corresponding basic direction. This scheme assigns each eye movement a distinct event symbol, thus merging both signal components into one event string sequence (see Figure 4.4).

To recognise eye gestures which consist of several movements, the event string sequence is scanned for eye movement patterns following a string matching approach. To make the recognition more robust, the symbols representing intermediate directions are recognised as the nearest neighbouring symbol (e.g. I as U or 9). For matching, the current string sequence is continuously compared with string templates representing all possible gestures (see Table 4.1). For each template, the edit distance between the templates and the segment is calculated. If one of the templates exactly matches the cur-
Figure 4.5: Adaptive filter for artefact compensation while walking slowly (a), moderate (b) and fast (c). Vertical acceleration signal and threshold (dashed line) for detecting walking activity. Horizontal acceleration signal and first derivative for calculating the step length. Resulting window size used by the median filter at the bottom.

4.4.5. Artefact Compensation

As EOG is measured with body-worn sensors, motion causes artefacts in the signals and affects eye movement detection. Walking is a common activity, e.g. on the way to work, during the day or in spare time at the weekend. Thus, walking serves as a good test bench for investigating artefacts induced by body motion. Analyses showed that artefacts in the EOG signals occur periodically according to the step frequency. A median filter with fixed window size fails to eliminate these artefacts for different persons and walking speeds. A parameter sweep on the window size using example data recorded from several participants revealed that the optimal size is strongly related to the temporal step length. Therefore, we use an algorithm implementing an adaptive filter. The idea is to exploit the repetitive characteristic of walking and adapt the window size of the median filter to the step length as long as walking activity is detected (see Figure 4.5).

To detect walking, the algorithm first analyses the vertical axis of the goggle-mounted accelerometer. If the corresponding signal exceeds a defined threshold, the algorithm tries to detect steps by searching for zero-crossings of the first derivative of the low-pass-filtered acceleration data of the horizontal axis (see [22] for details). Walking is assumed as long as such steps are detected. In order to smooth out variations in walking style for different participants, the step length is calculated on the basis of three consecutive step movements (e.g. right - left - right) separately for the left and the right leg. By calculating the length continuously for each step, the algorithm can adapt to different persons and walking speeds. For softer adaptation, only small incre-
The aim of this experiment is to assess the feasibility of using the wearable electrooculographic system for stationary human-computer interaction. To investigate the use of explicit eye gestures, we developed a computer game consisting of eight different game levels. In each game level, participants had to perform one defined eye gesture consisting of a changing number of consecutive eye movements. The gestures were selected and ordered to be of increasing complexity (see Table 4.1).

The first gesture was added to make it easier for the participant to familiarise with the game. Gesture 2 through 6 were inspired by [12] thus allowing to compare the results. The last two gestures contain shorter eye movements to assess if small-sized gestures can still be performed by the participants and recognised by the system.

### 4.5.1. Setup

The experiments were conducted using the wearable eye tracker running at 100Hz sampling rate, a standard desktop computer and a 17” flat screen with a resolution of 1024x768 pixels. As reference points for the calibration procedure red dots were put at the corners and edges of the screen. The participants were seated in front of the screen facing its centre (see Figure 4.6). In contrast to a previous study [12], no head stand was used, i.e. movements of the

<table>
<thead>
<tr>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
<th>Level 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>R1R</td>
<td>DRUL</td>
<td>RDLU</td>
<td>RLRLRL</td>
</tr>
<tr>
<td>Level 5</td>
<td>Level 6</td>
<td>Level 7</td>
<td>Level 8</td>
</tr>
<tr>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>3U1U</td>
<td>DR7RD7</td>
<td>1397</td>
<td>DDR7L9</td>
</tr>
</tbody>
</table>

Table 4.1: Eye gestures of increasing complexity and their string representations used in the eight levels of the computer game (cf. Figure 4.4). The grey dot denotes the start and the arrows the order and direction of each eye movement.
Chapter 4: Relative Eye Movements for Human-Computer Interaction

Figure 4.6: Experimental setup consisting of a desktop computer running the game (1), the Goggles (2), a flat screen with red dots used for calibration (3) and the Pocket (4). The screenshots on the right show the sequence of eye movements and the generated event symbols for gesture 3U1U (from top to bottom). The red dot denotes the start of the gesture and the blue dot its end point. Blue arrows indicate the order and the direction of each expected eye movement and are masked out after having been successfully performed.

head and the upper body were allowed at any time during the experiments. However, we encouraged the participants to sit upright with their eyes about 55cm to 65cm away from the screen. The expected movement order and their directions were shown as blue arrows with grey dots denoting the start and end point of a movement (see Figure 4.6 for an example).

4.5.2. Experimental Procedure

In a first step, the classification thresholds were calibrated: The threshold for blink detection was determined by asking the participants to blink 10 times and adjusting the threshold until the corresponding peaks in the EOG signal exceeded the threshold by about 30%. To calibrate for saccade detection, the participants were asked to look in alternation at two of the red dots on opposite edges of the screen. To improve the calibration the same was repeated with an extra stopover at the centre of the screen. Afterwards, the recognition was verified by asking the participants to focus on each red dot at the corners
Table 4.2: Accuracy for the different gestures for each individual participant without test run. The accuracy gives the ratio of eye movements resulting in a correct gesture to the total number of eye movements performed. The table also shows the participants’ gender (f: female, m: male) and vision aid usually needed (*: glasses, †: lenses).

<table>
<thead>
<tr>
<th>Gesture</th>
<th>S1 (m*)</th>
<th>S2 (f)</th>
<th>S3 (f)</th>
<th>S4 (m†)</th>
<th>S5 (m)</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1R</td>
<td>88</td>
<td>69</td>
<td>100</td>
<td>100</td>
<td>69</td>
<td>91</td>
</tr>
<tr>
<td>DRUL</td>
<td>100</td>
<td>71</td>
<td>100</td>
<td>100</td>
<td>86</td>
<td>90</td>
</tr>
<tr>
<td>RDLU</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>RLRLRL</td>
<td>100</td>
<td>100</td>
<td>95</td>
<td>93</td>
<td>100</td>
<td>94</td>
</tr>
<tr>
<td>3U1U</td>
<td>79</td>
<td>100</td>
<td>90</td>
<td>100</td>
<td>100</td>
<td>95</td>
</tr>
<tr>
<td>DR7RD7</td>
<td>90</td>
<td>95</td>
<td>78</td>
<td>71</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>1379</td>
<td>73</td>
<td>90</td>
<td>100</td>
<td>83</td>
<td>88</td>
<td>86</td>
</tr>
<tr>
<td>DDR7L9</td>
<td>95</td>
<td>91</td>
<td>93</td>
<td>71</td>
<td>89</td>
<td>92</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>S6 (m*)</th>
<th>S7 (m)</th>
<th>S8 (m)</th>
<th>S9 (m)</th>
<th>S10 (m*)</th>
<th>S11 (m*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>100</td>
<td>100</td>
<td>90</td>
<td>66</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>90</td>
<td>100</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>88</td>
<td>100</td>
<td>90</td>
<td>100</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>92</td>
<td>82</td>
<td>100</td>
<td>89</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>75</td>
<td>100</td>
<td>90</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>93</td>
<td>88</td>
<td>75</td>
<td>96</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>81</td>
<td>100</td>
<td>76</td>
<td>85</td>
<td>89</td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>100</td>
<td>73</td>
<td>76</td>
<td>76</td>
<td></td>
</tr>
</tbody>
</table>

| 92      | 94     | 86     | 93     | 89       |
| 92      | 94     | 86     | 93     | 89       |

of the screen in clockwise order. The assistant then checked the stream of eye movement events for errors and initiated a re-calibration if necessary.

Once the calibration was successfully completed the experiment was started. The participants performed three runs with all eight game levels being played in each run: The first was a test run to introduce the game and verify the blink and saccade thresholds for gesture recognition. No scores were recorded in this initial run. In two subsequent runs the participants played all levels of the game again. In these two runs, the participants were asked to concentrate on the game as performance measurements were taken to calculate the final game score.

In each game level, the corresponding eye gesture was to be repeatedly
performed as fast as possible by the participant until the first successful try. To reach a high score, wrong eye movements, i.e. movements which were not part of the expected gesture, had to be minimised. For each correct movement, the corresponding arrow was masked out on the screen to reduce visual distraction. For the same reason, each correct and incorrect eye movement was indicated by a distinct sound. If a wrong eye movement was recognised, the level was restarted and a penalty was rewarded on the game score. Once a whole eye gesture was successfully completed, the next game level was started showing the next gesture. All wrong and correct eye movements as well as the time required to complete each gesture were recorded for each level. To trigger these measurements, the participant had to press a button once before performing each gesture. The total experiment time for each participant was about 25 minutes. At the end of the experiment, the participants were asked on their experiences on the procedure in a questionnaire.

4.5.3. Results

We collected data from 11 participants - two female and 9 male - between the ages of 24 and 64. (Originally there were 14 participants, but three participants had to be withdrawn due to poor signal quality resulting in calibration problems.) Four participants usually needed spectacles which they could not use during the experiment. For the participants who completed the experiment, the average setup time was seven minutes including putting on the glasses as well as setting up and calibrating the recording system.

The results for each individual participant only show a small range of different accuracies (see Table 4.2). The results were calculated solely using data from the second and the third run as the first one was only for testing. The accuracy was calculated as the ratio of eye movements resulting in a correct gesture to the total number of eye movements performed in the level. The highest result is 95% (participant 3) while the worst result was for participant 8, with an accuracy of 86%. It can be seen from the table that performance does not correlate to the gender of the participant. Also the datasets recorded from persons which usually need a vision aid do not show significant differences to the others.

The average performance over all participants is given in Table 4.3 which shows the times $T_T$ and $T_S$, the time ratio $T_S/T_T$ and the accuracy $Acc$ to perform each of the eight gestures. $T_T$ denotes the total time the participants spent trying to complete each of the gestures while the success time $T_S$ only measures the time spent on all successful attempts.

Table 4.4 shows the average time required to perform five gestures in comparison to a video-based system used in a previous study [12]. The raw times for EOG ($T_R$ EOG) and video ($T_R$ Video) show that the latter performs much better. However, the experimental setups differed in one important aspect: In the previous study all gestures were performed using a static interface. In this work the arrows indicating the direction were successively masked out as soon as the movement was recognised correctly. While this approach reduced visual distraction, it visually emphasised a characteristic of the system: Due to the signal processing involved, recognising one distinct eye movement
4.5. Experiment I - Eye Gestures for Stationary HCI

Table 4.3: Average performance for the different gestures over all participants without test run. $T_T$ is the total time spent to complete the gesture and $T_S$ the success time spent only on successful attempts. The accuracy $Acc$ is the ratio of eye movements resulting in a correct gesture to the total number of movements performed until success.

<table>
<thead>
<tr>
<th>Gesture</th>
<th>$T_T$ [ms]</th>
<th>$T_S$ [ms]</th>
<th>$T_S/T_T$</th>
<th>Acc [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1R</td>
<td>3370</td>
<td>2890</td>
<td>0.858</td>
<td>85</td>
</tr>
<tr>
<td>DRUL</td>
<td>4130</td>
<td>3490</td>
<td>0.845</td>
<td>90</td>
</tr>
<tr>
<td>RDLU</td>
<td>3740</td>
<td>3600</td>
<td>0.963</td>
<td>93</td>
</tr>
<tr>
<td>RLLRLR</td>
<td>6680</td>
<td>5390</td>
<td>0.807</td>
<td>90</td>
</tr>
<tr>
<td>3U1U</td>
<td>4300</td>
<td>3880</td>
<td>0.902</td>
<td>89</td>
</tr>
<tr>
<td>DR7RD7</td>
<td>12960</td>
<td>5650</td>
<td>0.436</td>
<td>83</td>
</tr>
<tr>
<td>1379</td>
<td>6360</td>
<td>3720</td>
<td>0.585</td>
<td>84</td>
</tr>
<tr>
<td>DDR7L9</td>
<td>25400</td>
<td>5820</td>
<td>0.229</td>
<td>83</td>
</tr>
</tbody>
</table>

Table 4.4: Average response time $T_R$ required to perform five different eye gestures over all participants without initial test run in comparison to a video-based system. The third column gives estimated times correcting a difference in the experimental setups (see text for details).

<table>
<thead>
<tr>
<th>Gesture</th>
<th>$T_R$ [ms]</th>
</tr>
</thead>
<tbody>
<tr>
<td>RDLU</td>
<td>3740</td>
</tr>
<tr>
<td>DRUL</td>
<td>4130</td>
</tr>
<tr>
<td>RLLRLR</td>
<td>6680</td>
</tr>
<tr>
<td>3U1U</td>
<td>4300</td>
</tr>
<tr>
<td>DR7RD7</td>
<td>12960</td>
</tr>
</tbody>
</table>

took about half a second. We recognised that the participants introduced a delay in waiting for the arrows to disappear. This obviously affected their overall response time. To take this into account, we assumed all movements had been performed without hiding the arrows. We estimated the response time without delay ($T_R$ EOG w/o delay) as the average number of eye movements multiplied with the delay and subtracted from the average total time $T_T$. 


<table>
<thead>
<tr>
<th>Run</th>
<th>Visual task</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1. large horizontal eye movements standing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. large vertical eye movements standing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. fixation</td>
<td>walking</td>
</tr>
<tr>
<td></td>
<td>4. fixation</td>
<td>walking</td>
</tr>
<tr>
<td>2</td>
<td>1. simple eye movements standing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. simple eye movements walking</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. simple eye movements walking</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1. complex eye movements standing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. complex eye movements walking</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. complex eye movements walking</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.5: Experimental procedure consisting of three runs with performing different eye movements: Baseline measurement with large eye movements without using the head-up display and fixations, simple (run 2) and complex (run 3) eye movements each while standing and walking down a corridor.

4.6. Experiment II - Eye Movements for Mobile HCI

In this experiment, we target a mobile scenario and investigate how artefacts induced by physical activity can be detected and compensated in EOG signals. The experimental scenario involved participants to perform different eye movements on a head-up display (HUD) while standing and walking down a corridor. A custom software showed the expected movements with a defined order and timing.

4.6.1. Setup

The experiments were done using the wearable eye tracker running at 100Hz sampling rate, a standard laptop, a SV-6 head-up display from MicroOptical with a resolution of 640x480 pixels mounted to the Goggles frame and a wearable keyboard Twiddler2 (see Figure 4.7). The laptop was used to run the experimental software. Similar to the first experiment, eye movements were indicated on the HUD as arrows with a red dot denoting the start and end point of each movement. During the experiments, the laptop was worn in a backpack in order not to constrain the participants during walking. As the experimental assistant did not have control over the system, once the experiment was started, the Twiddler2 was needed to allow the participants to control the software and start the different recordings.
4.6. Experiment II - Eye Movements for Mobile HCI

Figure 4.7: Experimental setup consisting of a head-up display (1), the wearable eye tracker (2), a laptop (3) and a Twiddler2 (4). The screenshots on the right show the different eye movements performed in the three runs: Fixations on the centre of the screen (a), simple movements in vertical and horizontal direction L’RLRLRL’U’DUDUDU’ (b) and additional movements along the diagonals 7’R1U3ULD9D7R1’ (c) (cf. Figure 4.4, quotation marks indicate movements of only half the distance). The red dots in the centre denote the start; arrows indicate the directions of the movements.

4.6.2. Experimental Procedure

The participants were not the same as for the first experiment, thus unfamiliar with the recording system. Therefore, they were first trained on the game using the laptop screen. Once the game was finished, the HUD was attached and the laptop was put in the backpack to start the experiment. The participants performed three runs each consisting of different visual tasks while standing and walking down a corridor (see Table 4.5). A moving dot indicated the sequence and direction of the expected eye movements for each of these tasks. The participants were asked to concentrate on their movements and fixate this dot permanently. The timing was specified which resulted in exactly one eye movement about every 3s.

The first run was carried out as a baseline case with fixations on the centre of the screen and large saccades without using the HUD. In two subsequent
runs the participants were asked to perform different sequences of eye movements on the HUD while standing and walking: The second run only contained simple movements in vertical and horizontal direction. The third run also included additional movements along the diagonals. Starting in the centre of the screen, these two sequences encode to L’RLRLRL’U’DUDUDU’ and 7’R1U3ULD9D7R1’ (cf. Figure 4.4, quotation marks indicate movements of only half the distance).

4.6.3. Results

We recorded five male participants between the age of 21 and 27 totalling roughly 35 minutes of recording with walking activity accounting for about 22 minutes. As the mobile setting did not allow to record a ground truth, we decided to do a comparison to assess a relative performance measure. In a first step, the total number of detected saccades was calculated using the raw data of all participants. This was done separately for run 2 and 3, for both standing and walking, for each participant and for the horizontal and the vertical signal component. The thresholds for the saccade detection algorithm were fixed to $T_{saccH} = 700$, $T_{saccV} = 2000$ for all participants and runs. This analysis was then repeated twice with the same data: Once filtered by a median filter on a sliding window with a fixed size of 20 samples (0.2s) and once after applying the adaptive filter. As the stationary results were used as a reference the fixed window size was selected to show good results for these recordings.

Figure 4.8 shows boxplots for the total number of detected saccades in the horizontal EOG signal component of run 3. Each box summarises the statistical properties of the data of the five participants: The horizontal red lines in each box indicates the median and the upper and lower quartiles. The vertical dashed lines indicate the data range, points outside their ends are outliers. Boxes are plotted for the following cases: stationary and raw signal, stationary and fixed median filter, stationary and adaptive filter, walking and raw signal, walking and fixed median filter, walking and adaptive filter. The single solid horizontal line indicates the expected number of saccades defined by the experimental procedure.

What can be seen from the figure is that, in general, more artefacts are detected as saccades in the vertical EOG signal component. In the stationary case, both filters perform equally well but, compared to the expected number of saccades, improve the results only slightly. During walking, however, significant differences can be recognised: The raw recordings show about eight times more detected saccades than in the stationary case which renders eye movement detection impossible. While the median filter with a fixed window size fails in removing these artefacts, the adaptive filter still performs well, particularly for the horizontal EOG signal component.
4.7. Discussion

4.7.1. On Eye Tracking Using EOG

In this work, a novel wearable eye tracker was described and evaluated. In contrast to common solutions using video, which require rather bulky equipment, this compact device is based on EOG. This enables a light-weight and unobtrusive integration into goggles which makes the system suited for mobile recordings in daily-life. As EOG requires much less computational power, this allows for low-power design and on-board storage which are crucial points for autonomous long-term recordings.

Our results show that EOG performs equally well to video if gaze tracking is not required. In the same way as video-based systems, EOG requires a calibration procedure. Once the thresholds are set, EOG is robust to varying distances between the person and the screen. In mobile settings with different screen sizes, the algorithm for eye gesture recognition using video introduced in [12] may require the grid size and timeout parameters to be continuously adapted. This is difficult to achieve automatically without knowing the screens’ dimensions and the user’s relative distance. With EOG, the thresholds only need to be adapted if the screen size is reduced considerably. By implementing a procedure which automatically detects if a re-calibration is required, adaptation can be performed in the background without distracting the participant.

Wearability and comfort are important for long-term use. Nine participants from our study reported that they felt uncomfortable due to rather high electrode pressure especially below the eye. In the questionnaire, however, they did not report of having felt physically constrained or distracted during the game by wearing the goggles. For three out of 14 participants eye movements could not be detected. For the first participant, the goggles did not fit well which resulted in a lot of crosstalk in the vertical signal component. Suitable thresholds could not be found as the crosstalk was almost as strong as the vertical signal itself. For the second participant, probably due to dry skin, the vertical signal component was poor even though the goggles did fit well. The third participant declared afterwards having been very tired during the experiment. This could clearly be seen in the EOG signal by the presence of a lot of blinks and correctional saccades. These artefacts could not be removed completely and rendered eye gesture recognition impossible. We aim to solve these problems with the next prototype of the eye tracker which is currently under development. This includes revised goggles which can more easily be adapted to individual differences in head shape and size and which provide a mechanically improved electrode mounting.

4.7.2. On EOG-Based Eye Gestures

From the first experiment we found that EOG is a potentially robust input modality for HCI applications. EOG signals can be efficiently processed to recognise even complex gestures consisting of different consecutive eye movements. Table 4.4 shows that a set of gestures used to play a computer
game can be recognised with equal performance to a video-based system. It has to be noted, however, that we estimated part of the results due to a different experimental setup. Still, we are confident this estimation is reasonable on average and that the eye tracker would still perform comparably after changing this setup.

The concept of playing a computer game using eye gestures was quickly understood by all participants. We see in Table 4.2 that all of them were able to achieve a eye gesture accuracy of around 90% and often managed to perform the various gestures at the first try. Surprisingly, the accuracy for the easiest gesture in the first game level was similar to the last two (see Table 4.3). A possible explanation for this might be that the players were more inexperienced at the beginning and needed time to accustom to the game.

Although not shown here, from their performance we found that all participants quickly learned how to use their eyes as a control input. However, using explicit eye gestures remained odd and 30% of the participants reported of having had problems to stay concentrated during the experiment. They accounted this to the fact that controlling their eye movements consciously was a bit tiring. However, fatigue is an intrinsic problem not only for eye gestures but also for common input modalities such as speech or hand gestures. Eye gestures outperform these modalities if the hands can not be used (e.g. during driving, during a surgery or while working on the computer) or if speech input is not possible (e.g. for privacy reasons or in very silent or very noisy surroundings).

Six participants usually needed vision aids which they could not use during the experiment. Surprisingly, Table 4.2 shows that these participants performed equally well compared to those with normal sight. At least for the distance between the participant and the screen used in the experiment, the missing sight correction did not prevent to perform the different gestures successfully. However, it is clear that with a view to long-term use in ubiquitous settings with a variety of interfaces in different distances, goggles which still allow to use spectacles at the same time are desirable.

By analysing the overall performance we uncovered an interesting result: Gestures only consisting of large movements in the horizontal, vertical and diagonal directions worked well while those with smaller scale were more difficult to detect reliably. This is probably caused by the low amplitudes in the vertical EOG signal also recognised in the second experiment. This might be solved by simple means, e.g. by optimising the electrode placement and mounting or by designing HCI interfaces which rely on eye gestures performed mostly on the horizontal axis or with larger vertical eye movements. For mobile HCI with simultaneous physical activity, these aspects will become even more important.

### 4.7.3. On Artefact Compensation

From the experiments we found that for EOG recordings, particularly in mobile settings, efficient algorithms able to cope with signal artefacts caused by physical activity are required. Without compensation, artefacts may dominate the signal which renders eye movement detection impossible. Figure 4.8
4.7. Discussion

shows that the proposed adaptive filter tuned to the walking pace can remarkably reduce the number of signal artefacts caused by walking activity. For long-term eye movement recordings with a wide range of different activities which constantly change during the day, more complex algorithms are clearly needed.

4.7.4. Conclusion

In this work we have shown the feasibility of building an autonomous eye tracker based on EOG. The device can be worn on the body which makes it particularly suited for long-term eye movement analysis in daily-life. A major benefit of EOG lies in the minimal amount of power and computation that is required for signal processing. By connecting several eye trackers, concurrent eye movement recordings for several people and distributed activity recognition may become possible. We have also shown that recognition of explicit eye gestures from EOG can be implemented as efficiently and robustly across different participants as for video-based systems. EOG-based eye input allows for versatile human-computer interaction and may eventually provide new means of light-weight interaction in mobile settings by complementing current input modalities.

Our long-term objective is to investigate how much information eye motion can provide about the user’s activity and context. Eye gestures could be performed by the user to provide explicit contextual information. In addition to further HCI refinements, we also plan to investigate unconscious eye movements. Unconscious eye movements are the result of cognitive processes in the human brain. These processes are related to external aspects such as the user’s activity or his environment, but also to internal aspects of visual perception, memory and learning. The analysis of eye motion thus may allow to deduce these aspects which would give important input for future context-aware systems.
Figure 4.8: Boxplots for the total number of detected saccades in the horizontal (1) and vertical (2) EOG signal component of run 3 with fixed thresholds $T_{saccH} = 700, T_{saccV} = 2000$ over all participants: stationary/raw (a), stationary/fixed median filter (b), stationary/adaptive filter (c), walking/raw (d), walking/fixed median filter (e), walking/adaptive filter (f). Horizontal red lines in each box indicate the lower quartile, median and upper quartile; dashed vertical lines show the data range; outliers are given as red crosses; the single solid horizontal line indicates the expected number of saccades.
Bibliography


Eye Movements for Context-Aware Gaming

Andreas Bulling, Daniel Roggen and Gerhard Tröster

Full publication title: EyeMote - Towards Context-Aware Gaming Using Eye Movements Recorded From Wearable Electrooculography

Abstract

Physical activity has emerged as a novel input modality for so-called active video games. Input devices such as music instruments, dance mats or the Wii accessories allow for novel ways of interaction and a more immersive gaming experience. In this work we describe how eye movements recognised from electrooculographic (EOG) signals can be used for gaming purposes in three different scenarios. In contrast to common video-based systems, EOG can be implemented as a wearable and light-weight system which allows for long-term use with unconstrained simultaneous physical activity. In a stationary computer game we show that eye gestures of varying complexity can be recognised online with equal performance to a state-of-the-art video-based system. For pervasive gaming scenarios, we show how eye movements can be recognised in the presence of signal artefacts caused by physical activity such as walking. Finally, we describe possible future context-aware games which exploit unconscious eye movements and show which possibilities this new input modality may open up.

5.1. Introduction

The recognition of user activity has turned out to play an important role in the development of today’s video games. Getting the player physically involved in the game provides a more immersive experience and a feeling of taking direct part rather than just playing as an external beholder. Motion sensors have already been implemented to recognise physical activity: Game controllers such as music instruments, guns, dance mats or the Wii accessories make use of different sensors to open up a whole new field of interactive game applications. However, in pervasive settings, the use of physical activity may not be sufficient or not always be desired. Furthermore, cognitive aspects like user attention and intentionality remain mainly unexplored despite having a lot of potential for gaming scenarios. Therefore, alternative input modalities need to be developed which enable new gaming scenarios, are unobtrusive and can be used in public without affecting privacy.

A lot of information about the state of the user can be found in the movement of the eyes. Conscious eye movement patterns provide information which can be used to recognise user activity such as reading [1]. Explicit eye gestures performed by the player may be used for direct game input. Unconscious eye movements are related to cognitive processes such as attention [2], saliency determination [3], visual memory [4] and perceptual learning [5]. The analysis of these movements may eventually allow novel game interfaces to deduce information on user activity and context not available with current sensing modalities. In this paper we describe how Electrooculography (EOG) can be used for tracking eye movements in stationary and pervasive game scenarios. Additionally, we discuss which possibilities unconscious eye movements may eventually provide for future gaming applications.
5.2. Related Work

5.2.1. Eye-based Human-Computer Interaction

Eye tracking using vision for human-computer interaction has been investigated by several researchers. Most of their work has focused on direct manipulation of user interfaces using gaze (e.g. [6, 7]). Drewes et al. proposed to use eye gestures to implement new ways of human-computer interaction [8]. They showed that gestures are robust to different tracking accuracies and calibration shift and do not exhibit the “Midas touch” problem [9].

5.2.2. Eye Movements and Games

Smith et al. studied eye-based interaction for controlling video games across different genres, a first-person shooter, a role-playing game and an action/arcade game [10]. By comparing eye-based and mouse-based control they found that using an eye tracker can increase the immersion and leads to a stronger feeling of being part of the game. In a work by Charness et al., expert and intermediate chess players had to choose the best move in five different chess positions with their eyes [11]. Based on the analysis of eye motion they found that more experienced chess players showed eye movement patterns with a higher selectivity depending on chess piece saliency. Lin et al. developed a game interface using eye movements for rehabilitation [12]. They reported that the participants’ eyes became more agile which may allow for specific applications to help people with visual disabilities.

5.2.3. EOG-based Interfaces

Several studies show that EOG can be implemented as an easy to operate and reliable interface. Eye movement events detected in EOG signals such as saccades, fixations and blinks have been used to control robots [13] or a wearable system for medical caregivers [14]. Patmore et al. described a system that provides a pointing device for people with physical disabilities [15]. All of these systems use basic eye movements or eye-gaze direction but they do not implement movement sequences which provide a more versatile input modality for gaming applications.

5.3. Wearable Electrooculography

5.3.1. Eye Movement Characteristics

The eyes are the origin of an electric potential field which is usually described as a dipole with its positive pole at the cornea and its negative pole at the retina. This so-called corneoretinal potential (CRP) is the basis for a signal measured between two pairs of electrodes commonly placed above and below, and on the left and right side of the eye, the so-called Electrooculogram (EOG).
If the eyes move from the centre position towards the periphery, the retina approaches one of the electrodes while the cornea approaches the opposing one. This results in a change in the electric potential. Inversely, eye movements can be tracked by analysing these changes in the EOG signal.

In the human eye, only a small central region of the retina, the fovea, is sensitive enough for most visual tasks. This requires the eyes to move constantly as only small parts of a scene can be perceived with high resolution. Simultaneous movements of both eyes in the same direction are called saccades. Fixations are static states of the eyes during which gaze is held at a specific location.

### 5.3.2. EOG Data Recording

EOG, in contrast to well established vision-based eye tracking\(^1\), is measured with body-worn sensors, and can therefore be implemented as a wearable system. In earlier work we described how unobtrusive EOG recordings can be implemented with a light-weight and potentially cheap device, the wearable eye tracker [16]. The device consists of Goggles with integrated dry electrodes and a signal processing unit called Pocket with a Bluetooth and a MMC module. This unit can also be worn on the body, e.g. in a cloth bag fixed to one of the upper arms. Four EOG electrodes are arranged around the left eye and mounted in such a way as to achieve permanent skin contact. Finally, a 3-axis accelerometer and a light sensor are attached to the processing unit with the latter pointing forward in line of incident light (see Figure 5.1). The system weights 208g and allows for more than 7 hours of mobile eye movement recording.

### 5.3.3. EOG Signal Processing

To detect complex eye gestures consisting of several distinct movements from EOG signals the stream of saccades needs to be processed and analysed in a defined sequence [16]. This detection has several challenges with the most important one being to reliably detect the saccade events in the continuous vertical and horizontal EOG signal streams. Another challenge are the various types of signal artefacts which hamper the signal and can affect eye gesture recognition. This involves common signal noise, but also signal artefacts caused by physical activity which need to be removed from the signal. The characteristics of blinks are very similar to those of vertical eye movements, therefore they may need to be removed from the signal. However, for certain applications, blinks may also provide a useful input control, thus only reliable detection is required.

The output of the wearable eye tracker consists of the primitive controls left, right, up, down and diagonal movements (see Figure 5.2), blinks (conscious and unconscious), saccades and fixations. In addition, the system pro-

---

\(^1\)With “eye tracking” we understand the recording and analysis of eye motion in contrast to “gaze tracking” which deals with tracking eye-gaze direction.
Figure 5.1: Components of the EOG-based wearable eye tracker: armlet with cloth bag (1), the Pocket (2), the Goggles (3) and dry electrodes (4). The pictures to the right show the Goggles worn by a person with the positions of the two horizontal (h) and vertical (v) electrodes, the light sensor (l) and the accelerometer (a).

Figure 5.2: Eye movement event encoding from horizontal and vertical EOG signals for gesture 3U1U: Windows marked in grey with distinct eye movement events detected in the horizontal and vertical signal component and final mapping to basic (U) and diagonal (3, 1) movements. The top right corner shows the symbols representing the possible directions for eye movement encoding.

provides the following low-level signal characteristics and additional sensor inputs: EOG signal amplitudes (horizontal, vertical), timing of eye movement events, relative gaze angle, head movement (3-axis acceleration signal) and level of ambient light. Finally, the device can provide high-level contextual information, e.g. on user activity [1] or eventually the user’s cognitive load or attention.
Table 5.1: Eye gestures of increasing complexity and their string representations used in the eight levels of the computer game (cf. Figure 5.2). The grey dot denotes the start and the arrows the order and direction of each eye movement.

### 5.4. Application Scenarios

In this section we describe how eye movements recorded from wearable EOG can be used for different game scenarios. We first focus on stationary and pervasive.

#### 5.4.1. Stationary Games

The first scenario considers interactive games which are played in stationary settings with constrained body movements. These types of gaming applications are typically found at home, e.g. while sitting in front of a console in the living room or at the computer in the workroom. As the player does not perform major body movements, the weight and size of a game controller is not a critical issue. Instead, aspects such as natural and fast interaction are of greater importance.

To assess the feasibility of using the wearable eye tracker as an input device in stationary settings we investigated a simplified game consisting of eight different levels. In each game level, participants had to perform a defined eye gesture consisting of a changing number of consecutive eye movements (see Table 5.1). The gestures in the experiment were selected to be of increasing complexity. For future stationary games, eye gestures may for example be used for direct user feedback or ingame task control.

Each eye gesture was to be repeatedly performed as fast as possible until the first successful try. If a wrong eye movement was recognised, i.e. one which was not part of the expected gesture, the level was restarted and a penalty was rewarded on the game score. Once the whole eye gesture was successfully completed the next game level showing the next gesture was
5.4. Application Scenarios

Figure 5.3: Experimental setup consisting of a desktop computer running the game (1), the Goggles (2), a flat screen with red dots used for calibration (3) and the Pocket (4). The screenshots on the right show the sequence of eye movements and the generated event symbols for gesture 3U1U (from top to bottom). The red dot denotes the start of the gesture and the blue dot its end point. Blue arrows indicate the order and the direction of each expected eye movement.

started. For each level, the number of wrong and correct eye movements as well as the required time were recorded.

The participants had to perform three runs with all game levels being played in each run. The first was a test run to introduce the game and calibrate the system for robust gesture recognition. In two subsequent runs the participants played all levels of the game once again. At the end of the experiment, the participants were asked on their experiences on the procedure in a questionnaire.

The experiment was conducted using the wearable eye tracker, a standard desktop computer and a 17” flat screen with a resolution of 1024x768 pixels. The participants were sitting in front of the screen facing its centre with movements of the head and the upper body allowed at any time. The expected eye movement order and their directions were shown as blue arrows with grey dots denoting the start and end point of a movement (see Figure 5.3).
<table>
<thead>
<tr>
<th>Gesture</th>
<th>S1(m)</th>
<th>S2(f)</th>
<th>S3(f)</th>
<th>S4(m)</th>
<th>S5(m)</th>
<th>S6(m)</th>
<th>S7(m)</th>
<th>S8(m)</th>
<th>S9(m)</th>
<th>S10(m)</th>
<th>S11(m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1R</td>
<td>88</td>
<td>69</td>
<td>100</td>
<td>100</td>
<td>69</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>90</td>
<td>90</td>
<td>66</td>
</tr>
<tr>
<td>DRUL</td>
<td>100</td>
<td>71</td>
<td>100</td>
<td>100</td>
<td>86</td>
<td>100</td>
<td>100</td>
<td>90</td>
<td>100</td>
<td>100</td>
<td>90</td>
</tr>
<tr>
<td>RDLU</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>88</td>
<td>100</td>
<td>90</td>
<td>100</td>
<td>89</td>
<td>100</td>
<td>88</td>
</tr>
<tr>
<td>RLRLRL</td>
<td>100</td>
<td>100</td>
<td>95</td>
<td>93</td>
<td>100</td>
<td>100</td>
<td>92</td>
<td>82</td>
<td>100</td>
<td>89</td>
<td>88</td>
</tr>
<tr>
<td>3U1U</td>
<td>79</td>
<td>100</td>
<td>90</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>75</td>
<td>100</td>
<td>90</td>
<td>90</td>
<td>92</td>
</tr>
<tr>
<td>DR7RD7</td>
<td>90</td>
<td>95</td>
<td>78</td>
<td>71</td>
<td>90</td>
<td>93</td>
<td>88</td>
<td>75</td>
<td>96</td>
<td>80</td>
<td>100</td>
</tr>
<tr>
<td>1379</td>
<td>73</td>
<td>90</td>
<td>100</td>
<td>83</td>
<td>88</td>
<td>81</td>
<td>100</td>
<td>76</td>
<td>85</td>
<td>89</td>
<td>100</td>
</tr>
<tr>
<td>DDR7L9</td>
<td>95</td>
<td>91</td>
<td>93</td>
<td>71</td>
<td>89</td>
<td>77</td>
<td>100</td>
<td>73</td>
<td>76</td>
<td>76</td>
<td>100</td>
</tr>
<tr>
<td>Average</td>
<td>91</td>
<td>89</td>
<td>95</td>
<td>90</td>
<td>90</td>
<td>92</td>
<td>94</td>
<td>86</td>
<td>93</td>
<td>89</td>
<td>92</td>
</tr>
</tbody>
</table>

Table 5.2: Accuracy for the different gestures for each individual participant without test run. The accuracy gives the ratio of eye movements resulting in a correct gesture to the total number of eye movements performed. The table also shows the participants’ gender (f: female, m: male).
5.4. Application Scenarios

<table>
<thead>
<tr>
<th>Gesture</th>
<th>$T_T$ [ms]</th>
<th>$T_S$ [ms]</th>
<th>Accuracy [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1R</td>
<td>3370</td>
<td>2890</td>
<td>85</td>
</tr>
<tr>
<td>DRUL</td>
<td>4130</td>
<td>3490</td>
<td>90</td>
</tr>
<tr>
<td>RDLU</td>
<td>3740</td>
<td>3600</td>
<td>93</td>
</tr>
<tr>
<td>RLRLRL</td>
<td>6680</td>
<td>5390</td>
<td>90</td>
</tr>
<tr>
<td>3U1U</td>
<td>4300</td>
<td>3880</td>
<td>89</td>
</tr>
<tr>
<td>DR7RD7</td>
<td>12960</td>
<td>5650</td>
<td>83</td>
</tr>
<tr>
<td>1379</td>
<td>6360</td>
<td>3720</td>
<td>84</td>
</tr>
<tr>
<td>DDR7L9</td>
<td>25400</td>
<td>5820</td>
<td>83</td>
</tr>
</tbody>
</table>

Table 5.3: Average performance and accuracy for the different gestures over all participants without test run. The accuracy is the ratio of eye movements resulting in a correct gesture to the total number of eye movements performed until success. $T_T$ is the total time spent to complete the gesture and $T_S$ the success time spent only on successful attempts.

Results

We collected data from 11 participants - 2 female and 9 male - aged 24 to 64. The results for each individual participant only show a small range of different accuracies (see Table 5.2). The results were calculated using data from the second and the third run as the first one was for testing. The accuracy was calculated as the ratio of eye movements resulting in a correct gesture to the total number of eye movements performed in the level. The highest result is 95% (participant 3) while the worst result is for participant 8, with an accuracy of 86%. It can be seen from the table that performance does not correlate to the gender of the participant.

Table 5.3 shows the average performance over all participants, i.e. the time and the accuracy to perform each of the eight gestures. $T_T$ denotes the total time the participants spent trying to complete each of the gestures; the success time $T_S$ only measures the time spent on all successful attempts.

Table 5.4 shows the average response time $T_R$ required to perform five gestures in comparison to a video-based system [8]. $T_R$ was calculated from $T_T$ to take the different experimental setups into account (see [16] for details).

Figure 5.4 shows the average accuracy for different eye movements and its increase during the three experimental runs.

5.4.2. Pervasive Games

The second scenario considers pervasive games which are not constrained in terms of the players’ body movements and/or not restricted to a certain location [17]. These games may therefore either be played indoors in front of a console, in combined virtual and physical environments or in daily life set-
Chapter 5: Eye Movements for Context-Aware Gaming

<table>
<thead>
<tr>
<th>Gesture</th>
<th>$T_R[ms]$</th>
<th>EOG</th>
<th>Video</th>
</tr>
</thead>
<tbody>
<tr>
<td>DRUL</td>
<td>1520</td>
<td>1818</td>
<td></td>
</tr>
<tr>
<td>RDLU</td>
<td>1630</td>
<td>1905</td>
<td></td>
</tr>
<tr>
<td>RLRLRL</td>
<td>2860</td>
<td>3113</td>
<td></td>
</tr>
<tr>
<td>3U1U</td>
<td>1940</td>
<td>2222</td>
<td></td>
</tr>
<tr>
<td>DR7RD7</td>
<td>5890</td>
<td>3163</td>
<td></td>
</tr>
</tbody>
</table>

**Table 5.4:** Average response time $T_R$ required to perform five different eye gestures over all participants without initial test run in comparison to a video-based system.

**Figure 5.4:** Plot of distinct eye movement performance with standard deviation over all participants for each experimental run. The red line shows the accuracy for movements in the basic directions (U,D,R,L), the blue one for diagonal movements (1,3,7,8) and the black plot the average over all movements.

...tings (e.g. role plays in natural environments). They require wearable equipment which needs to be light-weight and low-power to allow for unobtrusive and autonomous (long-term) use. Furthermore, pervasive games allow potentially more complex multi-modal and ubiquitous interaction with(in) the environment, for example with combined hand and eye gestures. Eye gestures in pervasive games may provide two functions: (1) they allow the player to be immersed in the environment, especially when/if combined with head up displays and (2) at the same time allow for privacy, since eye gestures are not likely to be noticed as it is the case for body gestures.

As EOG is measured with body-worn sensors, body motion causes artefacts in the signals and affects eye movement detection. However, EOG can still be used in mobile settings. To show the feasibility of using the *wearable*
eye tracker with simultaneous physical activity we carried out an experiment which involved participants to perform different eye movements on a head-up display (HUD) while standing and walking down a corridor. Walking is a common activity, thus serves well as a test bench for investigating how artefacts induced by body motion can be automatically compensated in EOG signals. We evaluated an adaptive median filter which first detects walking activity using the data from the acceleration sensor attached to the Goggles. If walking is detected, the filter then continuously optimises its parameters to the walking pace to reduce signal artefacts.

The experiment was done using the wearable eye tracker, a standard laptop, a SV-6 head-up display from MicroOptical with a resolution of 640x480 pixels mounted to the Goggles frame and a wearable keyboard Twiddler2 (see Figure 5.5). The laptop was used to run the experiment software. During the experiments, the laptop was worn in a backpack in order not to constrain the participants during walking. As the experimental assistant did not have control over the system, once the experiment was started, the Twiddler2 was needed to allow the participants to control the software and start the different recordings.

The participants were first trained on the game from the first experiment using the laptop screen. Once the game was finished, the HUD was attached to start the second experiment. The participants performed three runs each consisting of different visual tasks while standing and walking down a corridor. Similar to the first experiment, for each of these tasks, the sequence and direction of the expected eye movements were indicated on the HUD as arrows and a moving red dot. The participants were asked to concentrate on their movements and fixate this dot permanently.

The first run was carried out as a baseline case with fixations on the centre of the screen and large saccades without using the HUD. In two subsequent runs the participants were asked to perform different sequences of eye movements on the HUD while standing and walking: The second run only contained simple movements in vertical and horizontal direction. The third run also included additional movements along the diagonals (cf. Figure 5.2).

**Results**

We recorded 5 male participants between the age of 21 and 27 totalling roughly 35 minutes of recording with walking activity accounting for about 22 minutes. To assess a relative performance measure, we did a comparison to a standard median filter with fixed window size.

Figure 5.6 shows a boxplot for the total number of detected saccades in the horizontal EOG signal component of run 3. Each box summarises the statistical properties of the data of the 5 participants: The horizontal red lines in each box indicates the median and the upper and lower quartiles. The vertical dashed lines indicate the data range, points outside their ends are outliers. Boxes are plotted for the following cases: stationary and raw signal, stationary and fixed median filter, stationary and adaptive filter, walking and raw signal, walking and fixed median filter, walking and adaptive filter. The sin-
Figure 5.5: Experimental setup consisting of a head-up display (1), the wearable eye tracker (2), a laptop (3) and a Twiddler2 (4). The screen-shots on the right show the different eye movements performed in the three runs: Fixations on the centre of the screen (a), simple movements in vertical and horizontal direction (b) and additional movements along the diagonals (c) (cf. Figure 5.2). The red dots in the centre denote the start; arrows indicate the directions of the movements.

A single solid horizontal line indicates the expected number of saccades defined by the experimental procedure.

What can be seen from the figure is that in the stationary case, both filters perform equally well. During walking, however, significant differences can be recognised: The raw recordings show about eight times more detected saccades than in the stationary case. As the number of expected eye movements was constrained by the software these additionally detected saccades can be considered signal artefacts caused by walking. While the median filter with a fixed window size fails in removing these artefacts, the adaptive filter still performs well. This shows that signal artefacts caused by motion can be cancelled, thereby enabling the use of EOG-based game interfaces in pervasive gaming scenarios.
5.4. Application Scenarios

![Boxplot for the total number of detected saccades in the horizontal EOG signal component of run 3 with fixed thresholds over all participants: stationary/raw (a), stationary/fixed median filter (b), stationary/adaptive filter (c), walking/raw (d), walking/fixed median filter (e), walking/adaptive filter (f). Horizontal red lines in each box indicate the lower quartile, median and upper quartile; dashed vertical lines show the data range; outliers are given as red crosses; the single solid horizontal line indicates the expected number of saccades.](image)

**Figure 5.6:** Boxplot for the total number of detected saccades in the horizontal EOG signal component of run 3 with fixed thresholds over all participants: stationary/raw (a), stationary/fixed median filter (b), stationary/adaptive filter (c), walking/raw (d), walking/fixed median filter (e), walking/adaptive filter (f). Horizontal red lines in each box indicate the lower quartile, median and upper quartile; dashed vertical lines show the data range; outliers are given as red crosses; the single solid horizontal line indicates the expected number of saccades.

### 5.4.3. Future Context-Aware Games

Given the success of the new input controllers of today’s active video games, future games will probably see more natural and more sophisticated interaction. These may increasingly take place in everyday scenarios with multimodal input, several people being involved in the same game and a high level of collaboration. In terms of eye movements as an input modality, game interfaces based on direct input will probably remain an important focus of research [18]. However, additional information related to the underlying cognitive processes and the user’s context may open up new possibilities for game developers and players.

### Inducing Flow and Optimal Game Experience

Based on eye movement analysis, future games may be aware of the user’s cognitive load, and adapt the individual gaming experience accordingly [19]. In particular, such games may increase the demand on the user when his cog-
nitive load is assessed as being too weak, whereas demand may be decreased if cognitive load is recognised as being too high. This may enable the player to keep experiencing the feeling of full involvement and energised focus characteristic of the optimal experience, also known as flow [20]. In a collaborative game scenario this would allow to distinguish players with different game experience and adapt the game difficulty for a more balanced game experience.

Rehabilitation and Therapy Games

Designers may also develop special games which require eye movements to be performed as exercises for medical purposes in rehabilitation or visual training. By using wearable EOG, these games could be brought to daily-life settings which would allow for permanent training independently from a special therapy at the doctor. The game exercises may be automatically adapted to the visual learning process derived from eye movement characteristics to optimise the training [5]. These games could be specifically optimised to fit the special requirements of children, elderly or even disabled people who still retain control of eye motion.

Context-Aware Gaming

In a more general sense, future games may also provide new levels of context-awareness by taking into account different contextual aspects of the player. This context may comprise the player’s physical activity, his location or mental state. Specific activities expressed by the eyes such as reading [1] could for example be used in games to adaptively scroll or zoom textual information. Context-aware games may also incorporate additional information derived from eye movements such as attention [21], task engagement [22] or drowsiness [23] to adapt to individual players. Other aspects of visual perception such as attention [2], saliency determination [3] and visual memory [4] may also enable new types of context-aware game interfaces not possible today. For collaborative games, this knowledge could be exchanged and combined into a common game context to integrate several players over potentially large geographical distances.

5.5. Discussion and Conclusion

In this work we have shown how a wearable electrooculographic system can be used for tracking eye movements in stationary and pervasive gaming scenarios. In the experiments, several participants reported that the electrode placed below the eye was rather uncomfortable. In general, however, they did not feel constrained or distracted by wearing the goggles while gaming. To solve this issue, we are currently developing a new prototype of the system with improved electrode mounting. From the questionnaire we found that all participants easily learned to use their eyes for direct game control. However, using explicit eye gestures was tiring and about 30% of the participants had problems to stay concentrated during the game. Fatigue is an inherent
problem also for input modalities such as hand gestures or speech. In pervasive settings, eye gestures outperform these modalities if the hands can not be used or if speech input is not possible.

Therefore, we believe EOG has a lot of potential for interactive gaming applications, in particular for those with unconstrained body movements. In contrast to video-based systems, EOG only requires light-weight equipment and allows for long-term use due to low power implementation. Unlike body movements, eye-based input allows for privacy which may prove extremely relevant in pervasive scenarios. Combined with a head-up display, EOG-based wearable eye tracking may eventually allow a more immersive game experience in outdoor environments. Information derived from unconscious eye movements may provide a more natural input modality for game control and future context-aware games. Thus, wearable EOG may become a powerful measurement technique for game designers and may pave the way for novel video games not possible so far.
Bibliography


Chapter 5: Eye Movements for Context-Aware Gaming


Recognition of Reading Activity in Transit

Andreas Bulling, Jamie A. Ward, Hans Gellersen and Gerhard Tröster

Full publication title: Robust Recognition of Reading Activity in Transit Using Wearable Electrooculography

DOI: 10.1007/978-3-540-79576-6_2
Chapter 6: Recognition of Reading Activity in Transit

Abstract

In this work we analyse the eye movements of people in transit in an everyday environment using a wearable electrooculographic (EOG) system. We compare three approaches for continuous recognition of reading activities: a string matching algorithm which exploits typical characteristics of reading signals, such as saccades and fixations; and two variants of Hidden Markov Models (HMMs) - mixed Gaussian and discrete. The recognition algorithms are evaluated in an experiment performed with eight participants reading freely chosen text without pictures while sitting at a desk, standing, walking indoors and outdoors, and riding a tram. A total dataset of roughly 6 hours was collected with reading activity accounting for about half of the time. We were able to detect reading activities over all participants with a top recognition rate of 80.2% (71.0% recall, 11.6% false positives) using string matching. We show that EOG is a potentially robust technique for reading recognition across a number of typical daily situations.

6.1. Introduction

Activity recognition has recently emerged as a key area of research in building context-aware interfaces for mobile and pervasive computing. The problem of recognising physical activity in mobile situations, for example using body worn sensors, has been investigated by several researchers [1, 2]. However, recognition of activities based on more subtle cues, such as user attention and intention - a far more difficult problem - remains relatively unexplored.

A rich source of information on user activity is in the movement of the eyes. The paths that our eyes follow as we carry out specific activities also reveal much about the activities themselves. This is particularly true for activities with very specific eye movements, such as reading. Reading is a pervasive activity, e.g. on computer screens at work, advertisements and signs in public, and books read at home or while travelling. Thus information on a person’s reading activities can be a useful indicator of his daily situation as well as a gauge of task engagement and attention. Attentive user interfaces could comprise the current level of user interruptability or provide assistance to people with reading disabilities by automatically magnifying or explaining words or context in the text (for example see [3, 4]).

We propose Electrooculography (EOG) as a novel measurement technique for recognition of user activity and attention in wearable settings. EOG, in contrast to well established vision-based eye tracking, is measured with body-worn sensors, and can be implemented as a wearable system. Although requiring facial skin contact, we believe EOG electrodes can be designed to be relatively unobtrusive, such as through integration into spectacles. A compact on-body device can then be used to process the incoming EOG signals.

The primary aim of this research is to assess the feasibility of recognising reading activity in different daily situations using wearable EOG. The wider
goal of this is to gain insight into the potential of EOG for activity recognition. The specific contributions of the work are (1) an experiment involving data collection of participants reading text while travelling to and from work, (2) a new method for saccade detection as a basis for reading recognition, and (3) an analysis of reading classification using string matching and Hidden Markov Models (HMMs).

The aim of our experiment is to capture reading in transit during different mobile situations. Despite the unavoidable fact that participants wore sensing equipment on their faces, we took particular care to ensure that the chosen scenario - reading while travelling to and from work - was as realistic as possible. This scenario involved a continuous sequence of daily activities such as sitting at a desk, walking along a corridor, walking along a street, waiting at a tram stop and riding a tram. We recorded an 8 participant, ground truth annotated dataset, totalling nearly 6 hours of recordings - half of which involved reading.

Our work makes use of a new algorithm for detecting saccade features in EOG signals using Wavelet decomposition. Inspired by the typical characteristics of EOG signals during reading, we carry out a preliminary investigation into three different classification algorithms: a string matching algorithm on the horizontal saccade features; a discrete HMM also using the horizontal features; and a mixture of Gaussian HMM using the denoised signals from both horizontal and vertical EOGs. Our best result over all datasets was obtained using the string matching algorithm. Our main finding is that reading can be detected regardless of whether the participant is sitting, standing or walking, and in a variety of indoor and outdoor situations.

6.1.1. Related Work

In a recent work, Logan et al. aimed at recognising common activities in a “real world” setting using a large variety and number of common sensors such as wired reed switches, RFID tags and infra-red motion detectors in the environment [5]. They discovered that reading was one of the most difficult activities to detect and concluded that for covering all types of physical activity in daily life, additional sensors and improved algorithms need to be found.

All previous attempts to recognise reading have been based on vision to record eye movements. With the goal of building a more natural computer interface based on user activity, Campbell et al. investigated on-screen reading recognition using infra-red cameras to track eye movements [6]. The approach used was participant independent, robust to noise and had a reported accuracy of 100%. However, the system required that each participant’s head was kept still using a chin rest.

In a later work, Keat et al. proposed an improved algorithm to determine whether a user is engaged in reading activity on a computer monitor [7]. Using an ordinary video camera placed between the participant and monitor, 10 participants were asked to read an interesting text from a list of preselected articles. The participants were explicitly asked to undertake other types of common computer-related activities such as playing computer games or watching video clips during the course of the experiment. Using user-dependent train-
ing, they achieved an average reading detection accuracy of 85.0% with a false alarm rate of 14.2%. However, to ensure correct detection of gaze direction, participants were required to face the screen throughout the experiments.

Motivated by the goal of improving reading skills for people with reading disabilities, Sibert et al. developed a system for remedial reading instruction [3]. Based on visual scanning patterns, the system used visually controlled auditory prompting to help the user with recognition and pronunciation of words. Following the study, participants reported that the most obtrusive part of the system was the video camera used to track eye movements.

Eye tracking using vision is a well studied field with a growing number of researchers looking at the movements of the eyes during daily activities in natural environments. Important advances are being made to the understanding of how our brains process tasks, and of the role that our visual system plays in this [8]. To assist with this work, a number of commercial trackers are available, some of which are targeted at mobile use. The most wearable of these - the Mobile Eye from Applied Science Laboratories (ASL) and the iView X HED from SensoMotoric Instruments (SMI) - both require bulky headgear and additional, cumbersome equipment to process the incoming video streams. To-date no solution for portable eye tracking exists that is convenient and unobtrusive enough to allow for unaffected physical activity.

Eye movement characteristics such as saccades, fixations and blinks, as well as deliberate movement patterns detected in EOG signals, have already been used for hands-free operation of static human-computer [9] and human-robot [10] command interfaces. All of these studies show that EOG is a promising measurement technique that can be remarkably accurate, easy to operate, reliable and can also be made cosmetically acceptable. Another interesting application is the use of EOG-based switches in a hospital alarm system which provide immobile patients with a safe and reliable way of signalling an alarm [11].

EOG-based interfaces have also been developed for assistive robots [12] and particularly as a control for an electric wheelchair [13]. These systems are intended to be used by physically disabled people who have extremely limited peripheral mobility but still retain eye motor coordination. Although both applications target mobile settings the peoples’ movements are constrained and the situation therefore differs from the one investigated in this work.

For EOG to be truly unobtrusive - particularly for mobile settings - the design of novel electrodes and electrode configurations is a critical topic and still subject to research. Manabe et al. propose the idea of an EOG gaze detector using an electrode array mounted on ordinary headphones [14]. While this placement might reduce the problem of obtrusiveness, it raises two other issues - namely, low signal-noise ratio (SNR) and poor separation of horizontal and vertical components. In another work, Vehkaoja et al. made electrodes from conducting fibres and sewed them into a head cap [15]. As yet the device is still to be evaluated in operation.
6.2. Eye Movement Analysis

6.2.1. Wearable Electrooculography

The eyes are the origin of a steady electric potential field. This can be detected in total darkness and even while the eyes are closed. It can be described as a fixed dipole with its positive pole at the cornea and its negative pole at the retina. The magnitude of this corneoretinal potential (CRP) is in the range of 0.4-1.0 mV. It is not generated by excitable tissue but is rather attributed to a higher metabolic rate in the retina. This potential difference is the basis for a signal measured between two pairs of surface electrodes placed in periorbital positions around one eye, the so-called Electrooculogram.

If the eyes move from the centre position towards one of these electrodes, the retina approaches this electrode while the cornea approaches the opposing one. This results in a change in the potential - the EOG signal - which can be used to track eye movements. The movement is split into horizontal and vertical signal components reflecting the discretisation given by the electrode setup.

6.2.2. Eye Movement Characteristics

To be able to take advantage of the typical characteristics of eye movements during reading, it is important to understand its two main types, namely saccades and fixations.

Saccades: Humans do not look at a scene in a steady way. Instead, their eyes move around and locate interesting parts of the scene to build up a mental
“map” representing it. The main reason for this is that only a very small central part of the retina, the fovea, can sense in high resolution. The fovea has a very narrow field of view and in order to be able to see a wider scene, the eye must move constantly. The simultaneous movement of both eyes in the same direction is called a saccade. This is the fastest movement of any external part of the human body. The peak angular speed of the eyes during a saccade reaches up to 1000 degrees per second while lasting from about 20 to 200ms. The amplitude of a saccade is the angular distance that the eye needs to travel during a movement. For amplitudes of up to about 60 degrees, the duration of a saccade linearly depends on the amplitude. But beyond this, the velocity of the saccade remains constant and the duration of the larger saccades is no longer linearly dependent on the amplitude.

Eye movements during reading are characterised by a typical sequence of small and large saccades: First, sequences of small saccades occur while the eyes move over the words in a line of text. A large saccade is observed when the eyes move back to the beginning of the next line of text. Figure 6.1 shows how both of these types of saccade look in the horizontal part of an EOG reading signal.

**Fixations:** Fixation is the static state of the eye during which gaze is held upon a specific location. Humans typically alternate saccadic eye movements and fixations. However, visual fixation is never perfectly steady and fixational eye movements can also occur involuntarily. The term “fixation” can also be referred to as the time between two saccades during which the eyes are relatively stationary. Reading involves fixating on successive locations within a line but also across a page to reach different sections of the text.

### 6.2.3. Baseline Drift

Baseline drift is a slow signal change mostly unrelated to the actual eye movements but superposing the EOG signal. Baseline drift has many possible sources as for instance interfering background signals, electrode polarisation [16] or physical influences such as varying contact pressure of the electrodes. In a four electrode setup as used in this study, baseline drift can also be different for the horizontal and vertical EOG signal components.

Baseline drift poses problems on the various types of eye movement: For saccades, the difference between start and end can be assumed to be drift-free, as saccades are performed in a very short period of time. Only signal segments before and after a saccade can become subject to changes caused by baseline drift. During periods of smooth pursuit movements it is difficult to distinguish baseline drift from actual eye movement; similarly for fixations drift alters the EOG signal in a way that can be indistinguishable from that of a slow eye movement.

Several approaches to remove baseline drift from electrocardiogram signals (ECG) have been proposed in recent literature (for example see [17, 18]). As ECG shows repetitive characteristics, some of the algorithms perform sufficiently well at removing baseline drift from these signals. However, they perform worse for signals with non-repetitive characteristics such as EOG.
Thus the development of robust algorithms for baseline drift removal is still an active topic of research.

6.3. Data Collection

The experimental setup devised in this work was designed with two main objectives in mind: (1) to record eye movements using EOG in an unobtrusive manner in a real-world setting, and (2) to evaluate how well reading can be recognised using EOG for persons in transit. We defined a scenario of travelling to and from work containing a semi-naturalistic set of reading activities. It involved participants reading text while engaged in a sequence of activities such as sitting at a desk, walking along a corridor, walking along a street, waiting at a tram stop and riding a tram.

6.3.1. Experimental Procedure

Participants were asked to follow two different sequences: A first calibration step involved walking around a circular corridor for approximately 2 minutes while reading continuously. The second sequence involved a walk and tram ride to and from work (see Table 6.1). This sequence was repeated in three runs: The first run was carried out as a baseline case without any reading task. This was both to accustom the participants to the route, but also to provide a reasonable amount of NULL data - which contributed to the objective of obtaining a realistic dataset. In two subsequent runs the participants were asked to read a text throughout. Between each run the participants rested for about 5 minutes. The total experiment time for each participant was about one hour. At the end of each experiment, the participants were asked on their experiences on the procedure in a questionnaire.

In contrast to a previous study [6] we opted to allow a free choice on reading material. Only two conditions were made: (1) that the material was text-only, i.e. no pictures and (2) that participants only chose material they found interesting and long enough to provide up to an hour’s worth of reading. Thus the type of text, its style as well as page and font size could be chosen to each participant’s personal preference. Our objective was to induce a state where readers were engrossed in the task for the relatively long recording time, thus allowing us to gather realistic data without having to coerce participants. A further benefit was that if participants were engrossed in the task, they would be less likely to be distracted by other people.

We were able to collect data from 8 participants - 4 male and 4 female - between the ages of 23 and 35. (Originally there were 10 participants, but 2 had to be withdrawn due to recording problems resulting in incomplete data.) Most of the experiments were carried out in well lit, fair to cloudy conditions with two exceptions: One of the male participants was recorded at night where we had to rely on street lights while walking around outdoors. Another male was recorded in rain where an assistant had to hold an umbrella over the participant to protect the sensors and reading material. However, as
Figure 6.2: Experimental procedure involving a sequence of semi-naturalistic reading activities including sitting at a desk, walking along a corridor, walking along a street, waiting at a tram stop and riding a tram. The figure also shows the corresponding horizontal and vertical EOG signals as well as acceleration data from the head and the right wrist of one complete dataset.

neither of the datasets showed a decrease in signal quality, both were used for the analysis.

Annotation of Ground Truth

Each participant was tailed by an assistant who annotated both the participant’s current activity (sitting, standing, walking) and whether he was reading. For this level of detail (is the participant’s eyes on the page or not) the assistant had to monitor the participant from a close proximity - but without being so close as to cause a distraction. For this purpose we used a wireless controller from Nintendo, the Wii Remote (see Figure 6.3). Using the Wii’s thumb control buttons “up”, “down” and “right”, the assistant could annotate the basic activities of standing, sitting and walking. In parallel, the trigger button was held down whenever the participant appeared to be reading and released when he stopped. A fifth button was used to annotate special events of interest, such as when the participant passed through a door and while entering or leaving the tram. All other buttons were not used and disabled to not interfere with the labelling.
6.3. Data Collection

<table>
<thead>
<tr>
<th>Average Time (min:sec)</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>00:00</td>
<td>Start synchronisation gesture</td>
</tr>
<tr>
<td>00:10</td>
<td>Sit at desk (on 3rd floor, indoors)</td>
</tr>
<tr>
<td>01:00</td>
<td>Walk to lift</td>
</tr>
<tr>
<td>01:10</td>
<td>Stand and wait for lift</td>
</tr>
<tr>
<td>01:40</td>
<td>Take lift to ground floor</td>
</tr>
<tr>
<td>02:10</td>
<td>Walk to exit</td>
</tr>
<tr>
<td>02:40</td>
<td>Walk to tram stop (outdoors)</td>
</tr>
<tr>
<td>04:40</td>
<td>Wait at tram stop</td>
</tr>
<tr>
<td>07:00</td>
<td>Ride the tram down a stop</td>
</tr>
<tr>
<td>08:00</td>
<td>Wait at tram stop (outdoors)</td>
</tr>
<tr>
<td>11:00</td>
<td>Ride the tram up a stop</td>
</tr>
<tr>
<td>12:00</td>
<td>Walk to entrance (outdoors)</td>
</tr>
<tr>
<td>14:00</td>
<td>Walk to lift (indoors)</td>
</tr>
<tr>
<td>14:30</td>
<td>Stand and wait for lift</td>
</tr>
<tr>
<td>15:00</td>
<td>Take lift up</td>
</tr>
<tr>
<td>15:30</td>
<td>Walk to office</td>
</tr>
<tr>
<td>16:00</td>
<td>Sit in comfortable chair</td>
</tr>
<tr>
<td>17:00</td>
<td>End synchronisation gesture</td>
</tr>
</tbody>
</table>

Table 6.1: Scenario of travelling to and from work. This was repeated 3 times for each participant: Once as a baseline measurement without reading and twice with participants reading interesting texts. The total time per scenario averaged around 17 minutes: 6 min. indoors, 10 min. outdoors and 4 min. on tram).

Eye Movements

For EOG data collection we used a commercial system, the Mobi from Twente Medical Systems International (TMSI), which was worn on a belt around each participant’s waist (see Figure 6.3). The device is capable of recording a four-channel EOG with a joint sampling rate of 128Hz and transmitting aggregated data to a laptop carried by an assistant via Bluetooth.

The data was collected using an array of five electrodes positioned around the right eye (see Figure 6.3). The electrodes used were the 24mm Ag/AgCl wet ARBO type from Tyco Healthcare equipped with an adhesive brim to stick them to the skin. The horizontal signal was collected using one electrode on the nose and another directly across from this on the edge of the eye socket. The vertical signal was collected using one electrode above the eyebrow and another on the lower edge of the eye socket. The fifth electrode, the signal
Figure 6.3: Experimental setup consisting of five EOG electrodes (h: horizontal, v: vertical, r: reference) and the Mobi (1). The bottom figure shows the wireless Wii controller used for annotating the ground truth with three thumb control buttons (2), the reading trigger button (3) and a button for labelling special events (4). Also shown though not used in the current work are three Xsens motion sensors on the head and on the back of both hands (5a) and the XBus Master (5b).

reference, was placed away from the other electrodes in the middle of the forehead.

Physical Activity

One of our objectives for future work is to analyse the correlation of reading activities with physical movement and posture. Though beyond the focus of this current work, for completeness we include a short description of the additional recording setup. Briefly, this included MTx sensor nodes from XSens Technologies containing 3-axis accelerometers, magnetic field sensors and
gyroscopes positioned on each participant’s head as well as on the back of their hands (see Figure 6.3).

Unfortunately, the MTx system performed poorly using its bluetooth connection and so we were forced to use a wired connection. This was the only physical connection between the participant and assistant. Care was taken throughout by the assistant to ensure that the trailing wire did not interfere or distract the participant.

**Data Recording**

All recorded data was sent to a laptop in the backpack worn by the assistant. Data synchronisation was handled using the Context Recognition Network (CRN) Toolbox (see [19] for details). We made two extensions to the toolbox: the first was a reader to process and synchronise labelled data from the Wii Remote controller; the second extension was to implement a “heartbeat” component that provided audio feedback to the assistant on whether the toolbox was running and recording data, thus providing instant notification of device failures. The addition of the “heartbeat” was particularly useful as it allowed the assistant to concentrate on labelling and observing the participants rather than continually disturbing the procedure by checking the recording status.

### 6.4. Continuous Recognition Methods

The methods in this work were all implemented offline using MATLAB. However, with a view to a future online implementation on a wearable device, the algorithms were chosen to keep computation costs low. In this section we first describe the signal processing steps required for removing noise and baseline drift, and for extracting saccade information. We then describe three algorithms we use for classification: string matching (STR), discrete HMM (D-HMM) and mixed Gaussian HMM.

#### Noise and Baseline Drift Removal

In a parallel work we evaluated algorithms for baseline drift removal and integrated them into a general artefact compensation framework for EOG signals [20]. There, we adapted a wavelet packet approach originally proposed for ECG signals [17]. First, the data is stripped of high frequency noise using a simple median filter. The algorithm then performs an approximated multilevel 1-D wavelet decomposition at level nine using Daubechies wavelets on the horizontal signal component. The reconstructed decomposition coefficients then give a baseline drift estimation. Subtracting this estimation from the original signal finally yields the corrected signal with reduced drift offset.

#### Saccade Detection and String Encoding

Both the string matching and the D-HMM algorithms rely on the detection of saccades in the horizontal component of the reading signal. For this pur-
pose we developed the so-called Continuous Wavelet Transform - Saccade Detection (CWT-SD) algorithm, and applied it to the de-noised, baseline drift removed, horizontal EOG signal [20]. The CWT-SD first computes the continuous 1-D wavelet coefficients from the signal at scale 20 using Haar wavelets. A saccade is detected for all samples where the absolute value of the corresponding coefficient vector exceeds a threshold.

To keep the algorithm simple, the string encoding for this initial work only uses the horizontal component of the signal\(^1\). The saccade detection is run twice using two different threshold values: \(T_{\text{sac}}^L\) represents small saccades, such as jumps between words during reading; and \(T_{\text{sac}}^H\) represents large saccades, such as those observed during an end of line “carriage return”. The resulting sequence of large and small saccades together with their direction are then encoded into a string using characters according to the following scheme:

- “L”: large saccade to the left
- “R”: large saccade to the right
- “l”: small saccade to the left
- “r”: small saccade to the right

An example of an encoded reading segment is shown in Figure 6.4.

6.4.1. String Matching

Of all the algorithms presented here, string matching is computationally the most simple. It can be considered a light-weight approach, using only simple

---

\(^1\)Our future work will be based on 2D encoding using both horizontal and vertical information.
6.4. Continuous Recognition Methods

arithmetic, and can be easily adapted to a future online implementation, for example on a wearable device.

The matching is performed by moving a prototype string, representing a typical reading segment, over the signal string encoding, character by character. In each of these steps, the Levenshtein distance between the string template and the current signal string is calculated. The Levenshtein distance between two strings is given by the minimum number of operations needed to transform one string into the other, where an operation is an insertion, deletion, or substitution of a single character.

The algorithm then applies a threshold $T_{ed}$ on the Levenshtein distance vector to separate the two classes “reading” and “not reading”. This threshold defines how tolerant the classification is towards relative error in the edit distances. As this method does not yet adapt the string template to the signal while calculating the distances, it is sensitive to fluctuations in the number of small saccades. This results in a high number of false insertions. To counter this we slide a majority vote window $W_{str}$ across the event-based output classification to “smooth” the final result.

6.4.2. Hidden Markov Models

Hidden Markov Models are probabilistic models used to represent non-deterministic processes in partially observable domains and are defined over a set of states, transitions and observations. In this work we evaluate two different implementations of HMM: discrete and mixture of Gaussian. In each we use a 2-state model. This helps to keep the computation costs low. Intuitively, the two states represent the large left saccades when the eyes move back to the beginning of a line and the saccades during word reading. The parameters - state transition probabilities, observation likelihoods and Gaussian mixture settings - are set in a training step using the Baum-Welch (forwards-backwards) algorithm. For classification, we recursively apply the forward algorithm on incoming data samples. See [21] for details on these algorithms, and [22] for the Matlab implementation we used.

HMMs with Discrete Observations

For the discrete case, we use the same character features as for string matching. First we segment the incoming data using the large left saccades as boundaries (with each saccade beginning with an “L”). This leaves us with a sub-string of saccades (“r”, “l” and “R”) for each segment, which we then feed into the discrete HMM. Note that in this simplified model, the influence of a large right saccade, “R”, which might be encountered during page turns, is not modelled explicitly. For each segment, the forward algorithm returns a single log likelihood value. A threshold $T_{dhmm}$ is then applied and if the threshold is passed, the classifier returns reading, if not NULL is returned.
Chapter 6: Recognition of Reading Activity in Transit

HMMs with Gaussian Observations

For the Gaussian case, we use the horizontal and the vertical EOG signal component as the observation feature space (de-noised and baseline drift removed). To avoid singularities with the implementation of the forward-backward algorithm, we standardise the signals by setting the variance of the entire dataset to 1. Our model is again 2 state, but with a mixture of two Gaussian probability density functions to model the observations.

At each of its steps, the forward algorithm outputs a log likelihood value. We then smooth the sequence of these likelihoods using a sliding mean window \( W_{hmm} \). Finally, a threshold \( T_{hmm} \) is applied and if the threshold is passed, the classifier returns reading, if not NULL is returned.

6.5. Results

6.5.1. Parameter Selection and Training

Saccade Detection and String Matching

To determine the threshold parameters we applied the saccade detection algorithm using a threshold sweep on a manually cut subset of the data. On average for all participants, this subset of data contained 15 large reading saccades plus noise and artifacts caused by interrupting eye movements. For each threshold, we counted the number of large saccades that were detected and calculated the relative error \( \frac{\text{Total} - \text{Detected}}{\text{Total}} \). Based on this sweep we chose the large saccade threshold at \( T_{L_{sac}} = 7000 \). Due to the difficulties in manually segmenting samples of small saccades, we approximated the small threshold \( T_{s_{sac}} = 1500 \).

The string matching parameters were evaluated across a sweep of the majority vote window length \( W_{str} \), the distance threshold \( T_{ed} \) and a selection of different templates. In the analysis presented below, we chose to fix \( W_{str} = 30 \) and template “Lrrrrrrrrr” for all participants. Figure 6.5 shows an example output of the basic algorithm compared against the smoothed result and the ground truth labelling. Note the typical errors encountered: “merge” (m) when the output detects a single reading sequence where there should be two; “overfill” (o) when the output correctly detects reading just outside the labelling; and “underfill” (u) where the detected reading signal falls short of the labelling. These error types are explored in more detail later in this section.

HMM Training

Both HMM and D-HMM models were trained using data from the recordings of reading while walking around a corridor. Two leave-one-out training schemes were used: participant dependent only using the calibration data from the participant being tested and participant independent using calibration data only from other participants.
6.5. Results

Figure 6.5: Example string matching result showing the ground truth labelling, the classification result returned by the algorithm and the result after applying the smoothing filter. Also shown is an example merge (m) and an overfill (o) and underfill (u) error.

We evaluated the HMMs over a sweep of the main parameters. For the mixed Gaussian HMM, we discovered that sliding window size had limited influence, and so could be fixed at $W_{hmm} = 5$ seconds.

6.5.2. Continuous Classification

The different methods were compared across a sweep of their main parameters: $T_{ed} = 1...10$ (in 10 steps), $T_{hmm} = T_{dhmm} = -1...-9$ (in 32 steps). The results from all participants were then summed together. The resulting Receiver Operating Characteristics (ROC) curves are shown on the left of Figure 6.6. These plot true positive rate (recall) ($\frac{TP}{TP+FN}$) against false positive rate (FPR) ($\frac{FP}{FP+TN}$), where $TP, FP, TN$ and $FN$ represent true positive, false positive, true negative and false positive counts respectively. Best case results approach the top left corner while worst case (which means random) follow the diagonal.

The ROC clearly shows that string matching outperforms the HMMs. At its “best” ($T_{ed} = 5$), string matching returns a recall of 71.0% and FPR of 11.6% (total accuracy 80.2%). The mixed Gaussian returns a lower best-case at recall 62.0%, FPR 24.0% and accuracy 68.9% for $T_{hmm} = -3.5$ and participant independent training. The worst performing algorithm is the discrete HMM. However, both types of HMMs perform similar for the participant-dependent and the participant-independent case.

Further Analysis of String Matching

Based on these results, we chose string matching (with $T_{ed} = 5$) for further analysis. The error division diagram (EDD) of Figure 6.6 shows a detailed...
Figure 6.6: Left: ROC curves showing a performance comparison between string matching (STR) over a sweep of edit distance $T_{ed}$, Gaussian HMMs over a sweep of log threshold $T_{hmm}$ and discrete HMM over a sweep of $T_{dhmm}$. For the HMMs, both the participant-dependent and the participant-independent results are shown. Right: Detailed result for string matching with $T_{ed} = 5$ (corresponding to the point circled on the ROC curve). The EDD shows the proportion of the total dataset comprising true positives (TP), true negatives (TN), overfill, underfill, merge, insertion and fragmentation errors. Note that the proportion of negatives (NULL) in the dataset is roughly half (52.7%). Though overall accuracy is 80.2% ($TP + TN = 33.6% + 46.6%$), a large part of the errors are actually underfill and overfill timing errors.

timewise breakdown of the errors. The EDD highlights typical errors that occur in continuous recognition systems - beyond the basic FP and FN categorisation. Specifically there are three classes of error that we consider. Details on how these error classes are derived is outwith the scope of the current work, but can be found in [23]:

1. The “classical” errors, such as insertion (a reading event is detected where there is none in the ground truth) and deletion (failure to detect a reading event).

2. Fragmentation and merge: Fragmentation errors describe when a reading event in the ground truth corresponds to several events in the recognition system output. Merge is the opposite: several ground truth reading events are combined into one event - see (m) in Figure 6.5.

3. Timing errors: Overfill errors are where an event in the output of the system extends into regions of NULL, such as in the event (o) in Figure 6.5. The opposite of overfill is underfill (u), in this case the event...
6.5. Results

Figure 6.7: Performance evaluation for reading recognition during distinct activities - stand, sit and walk - using the string matching algorithm. The EDD shows the proportion of the total dataset comprising true positives (TP), true negatives (TN), overfill, underfill, merge, insertion and fragmentation errors. Note the different distributions of NULL in each of these cases: 65.1%, 40.3% and 50.7% for stand, sit and walk respectively. As expected, the results for walking are slightly worse than the others, with 12.5% serious error level (SEL). Interestingly, the sit case is slightly worse than stand. This is due in part to the greater number of fragmentation errors from sit.

In addition, EDDs also show the total true negative (TN) and true positive (TP) times. Using this breakdown, Figure 6.6 shows that 9.0% of the total time is underfill and overfill “error”. These are cases where the fault may be slightly offset labelling, or delays in the recognition system. The errors that might be regarded as more serious - insertion, deletion, merge and fragmenting - account for 10.8% of the total experiment time, this is the so-called serious error level (SEL).

Reading in Different Situations

To analyse the performance of string matching for the different activity situations, we divided the data into three sets, separating sitting, standing or walking activities. Each activity represented roughly equal sized portions of the dataset (88, 98 and 95 minutes respectively). The EDDs in Figure 6.7 show the results of this evaluation over all participants.
For standing, the best case result shows 72.8% recall and 9.2% FPR; for sitting the result is 73.9% recall, 13.2% FPR; and for walking it is 64.9% recall and 11.0% FPR. Looking further at the type of errors encountered, we see that both sit and walk contain large periods of underfill timing errors. Also of interest is the fact that 3.3% of the time during sit is classed as a fragmentation error.

**Results for Each Participant**

The results for each individual participant show a range of differences in recognition performance using the string matching algorithm (see Table 6.2). The highest recall result is 92.9% (participant 3) but with a FPR of 15.5%. The worst result was for participant 1, with 25.9% recall and 3.8% FPR. This was to be expected as the raw EOG signal quality for that participant was extremely low. During the experiment recordings, the overall signal level for this participant was known to be weak and reading saccades could hardly be seen. This was also a problem for participant 5 and was most likely caused by poor electrode placement and dry skin.

What can be seen from the table, however, is that the differences do not seem to correlate to the gender of the person. Also the two special datasets recorded at night and in rain do not show marked differences, but have comparable performance to the other sets.

### 6.6. Discussion

**On the recognition performance**

Three different approaches of recognising reading based on EOG in a wearable setting have been described and investigated in this work. Among the algorithms evaluated, string matching performs best, with an overall recognition rate of 80.2% (71.0% recall, 11.6% FPR). It is interesting to note that almost half of all errors are overfill and underfill. The most probable causes of these errors are the inaccuracies of the sliding majority filter and the labelling process.

The D-HMM method does not perform well. This is probably due the lack of sufficiently descriptive features. We see in Figure 6.6 that it is robust to different training configurations. This indicates that the novel method of saccade detection CWT-SD, upon which both string matching and D-HMM methods are based, is fairly robust to variations in EOG signals. This is a particularly useful trait for applications where a “training” step would be inconvenient or impractical. A further advantage of the CWT-SD is that it only requires the horizontal EOG signal component and therefore only 2 electrodes. Although we strongly believe that most of the useful information lies in the horizontal component, the (as yet unused) vertical component might also contain information that in a future study could improve the performance of both methods.
Table 6.2: Recall and FPR for each individual participant, and the median over all, using string matching. The table also shows the participants’ gender (f: female, m: male); the dataset recorded at night is marked with * while the one recorded in heavy rain with **.

<table>
<thead>
<tr>
<th></th>
<th>S1 (f)</th>
<th>S2 (m*)</th>
<th>S3 (f)</th>
<th>S4 (m**)</th>
<th>S5 (f)</th>
<th>S6 (m)</th>
<th>S7 (f)</th>
<th>S8 (m)</th>
<th>median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recall [%]</td>
<td>25.9</td>
<td>65.2</td>
<td>92.9</td>
<td>83.1</td>
<td>47.9</td>
<td>89.5</td>
<td>71.3</td>
<td>85.7</td>
<td>77.2</td>
</tr>
<tr>
<td>FPR [%]</td>
<td>3.8</td>
<td>12.1</td>
<td>15.6</td>
<td>17.9</td>
<td>4.3</td>
<td>18.3</td>
<td>12.9</td>
<td>8.9</td>
<td>12.5</td>
</tr>
</tbody>
</table>
Chapter 6: Recognition of Reading Activity in Transit

Despite using both horizontal and vertical signal components and the perceived aptitude of HMMs for this type of problem, the Gaussian HMM did not produce the results that have been expected. Part of the reason for this relatively poor performance (compared to the simple string matching) might be due to the specifics of our initial implementation. It is the authors’ belief that with careful selection of more descriptive features and perhaps a larger number of states, this method can be improved.

By analysing the different activity situations - reading while standing, sitting and walking - we uncover some interesting results. As expected, walking produces the worst results but only by a small margin - its serious error level is only 2.2% worse than sitting (Figure 6.7). Surprisingly, recognition is better while participants stand than when they sit. We would have expected the “sit” class to perform best as this is the one with potentially the least external influences. So far, we have no satisfactory explanation for this behaviour but we plan to investigate this further in detail in future experiments.

On EOG

From this initial study we found that EOG is a robust technique for recording eye movements in wearable settings. The main advantage of the EOG-based measurement technique is the fact that, in contrast to common video-based eye trackers, the participant only has to wear relatively unobtrusive and lightweight equipment. This contributes to the participants feeling unconstrained during the experiments and therefore allows for natural reading behaviour and unaffected physical activity.

One drawback is that EOG electrodes require good skin contact. Poor placement of electrodes was the reason for many of the problems in our work, particularly with participants 1 and 5 (see Table 6.2). This problem was usually solved by removal and reattachment of fresh electrodes. The fact that these electrodes are stuck to the face can also be regarded as inconvenient. In the post-experiment questionnaire, the participants from our study reported that they did not feel physically constrained by the sensors and wires - not even by the electrodes. However, it is clear that for long-term use a more comfortable and robust solution would be desirable.

Baseline drift is perhaps an unavoidable problem for wearable EOG recording. It is for this reason that accurate gaze tracking, for purposes such as target detection, might be difficult to achieve using mobile EOG. By analysing only the rough patterns created by eye movement, however, we can detect activities (such as reading) without the need for such pinpoint tracking.

On the experiment

The Wii-remote proved to be a useful and unobtrusive annotation tool - and was certainly preferable to the chore of video-based offline annotation. This method is certainly subject to inaccuracies when, for example, the assistant is distracted, or when buttons are pressed and released too early or too late. However, labelling errors are an intrinsic problem especially in wearable settings and a satisfying solution has not been found yet. One possible approach for a
future study could be to introduce redundancy to the labelling process: either a second assistant, or some semi-automatic solution, perhaps using additional sensors (e.g. using body movements or video-based eye tracking).

In the questionnaire, all participants declared that they did not feel distracted by people on the street and were only partially conscious about the experiment assistant. Half of the participants did report a feeling of unease while reading and walking. This unease could clearly be seen in the EOG signal by the occasional presence of small vertical saccades during reading - indicating whenever a participant looked up to check the way ahead.

Obviously reading while walking can be a dangerous activity, however this does not detract from the fact that many people actually do it. The other half of our participants found no problem with reading and walking - they all claimed to have occasionally read long texts (a book, newspaper or scientific paper) while in transit. It is more common for people to read shorter texts while walking, for example advertisements, timetables, etc., and this would be an interesting scenario for future study. Because such reading sequences are usually very short, the recognition methods and also the labelling scheme would probably have to be adapted.

Ideally, the most natural scenario would have involved recordings over a period of weeks or months. This would allow us to better study the general reading behaviour of our participants - and to open up interesting questions regarding daily reading habits. Unfortunately, the battery life and reliability of our recording equipment limited recordings to a few hours. Therefore, the main improvement concerning the experimental setup is to develop a wearable EOG device which does not impose these restrictions but allows for robust long-term eye movement recordings. This would also include an investigation of how to apply dry electrodes, e.g. by integrating them into spectacle frames, as they are more convenient for everyday use than wet electrodes stuck to the skin. Another interesting question is how dry electrodes would influence the quality of the signals and recognition performance.

6.6.1. Conclusion

Our work has shown that wearable EOG is a feasible approach for recognising reading in daily-life scenarios and is robust across an example set of activities for different participants. This raises the question of whether different reading behaviours and attention levels to written text can be detected automatically. A “reading detector” could enable novel attentive user interfaces which take into account aspects such as user interruptability and level of task engagement.

Given appropriate hardware, EOG offers the potential of long-term eye movement recordings. The movement patterns the eyes follow in daily routine reveal much about what people are doing - as well as what they intend to do. With growing interest in activity recognition as a topic within pervasive computing, this information may prove extremely relevant. In the future, eye movements may be used as a new sensing modality, providing access to the underlying cognitive processes not available with current sensing modalities.
Chapter 6: Recognition of Reading Activity in Transit
Bibliography


6.6. Discussion


Recognition of Office Activities

Andreas Bulling, Jamie A. Ward, Hans Gellersen and Gerhard Tröster

Full publication title: Eye Movement Analysis for Activity Recognition Using Electrooculography

DOI: 10.1109/TPAMI.2010.86

An earlier version of this work was published as

Eye Movement Analysis for Activity Recognition
Andreas Bulling, Jamie A. Ward, Hans Gellersen and Gerhard Tröster
DOI: 10.1145/1620545.1620552
Abstract

In this work we investigate eye movement analysis as a new sensing modality for activity recognition. Eye movement data was recorded using an electrooculography (EOG) system. We first describe and evaluate algorithms for detecting three eye movement characteristics from EOG signals - saccades, fixations, and blinks - and propose a method for assessing repetitive patterns of eye movements. We then devise 90 different features based on these characteristics and select a subset of them using minimum redundancy maximum relevance feature selection (mRMR). We validate the method using an eight participant study in an office environment using an example set of five activity classes: copying a text, reading a printed paper, taking hand-written notes, watching a video, and browsing the web. We also include periods with no specific activity (the NULL class). Using a support vector machine (SVM) classifier and person-independent (leave-one-person-out) training, we obtain an average precision of 76.1% and recall of 70.5% over all classes and participants. The work demonstrates the promise of eye-based activity recognition (EAR) and opens up discussion on the wider applicability of EAR to other activities that are difficult, or even impossible, to detect using common sensing modalities.

7.1. Introduction

Human activity recognition has become an important application area for pattern recognition. Research in computer vision has traditionally been at the forefront of this work [1, 2]. The growing use of ambient and body-worn sensors has paved the way for other sensing modalities, particularly in the domain of ubiquitous computing. Important advances in activity recognition were achieved using modalities such as body movement and posture [3], sound [4], or interactions between people [5].

There are, however, limitations to current sensor configurations. Accelerometers or gyroscopes, for example, are limited to sensing physical activity; they cannot easily be used for detecting predominantly visual tasks, such as reading, browsing the web, or watching a video. Common ambient sensors, such as reed switches or light sensors, are limited in that they only detect basic activity events, e.g. entering or leaving a room, or switching an appliance. Further to these limitations, activity sensing using subtle cues, such as user attention or intention, remains largely unexplored.

A rich source of information, as yet unused for activity recognition, is the movement of the eyes. The movement patterns our eyes perform as we carry out specific activities have the potential to reveal much about the activities themselves - independently of what we are looking at. This includes information on visual tasks, such as reading [6], information on predominantly physical activities, such as driving a car, but also on cognitive processes of
7.1. Introduction

visual perception, such as attention [7] or saliency determination [8]. In a similar manner, location or a particular environment may influence our eye movements. Because we use our eyes in almost everything that we do, it is conceivable that eye movements provide useful information for activity recognition.

Developing sensors to record eye movements in daily life is still an active topic of research. Mobile settings call for highly miniaturised, low-power eye trackers with real-time processing capabilities. These requirements are increasingly addressed by commonly used video-based systems of which some can now be worn as relatively light headgear. However, these remain expensive, with demanding video processing tasks requiring bulky auxiliary equipment. Electrooculography (EOG) - the measurement technique used in this work - is an inexpensive method for mobile eye movement recordings; it is computationally light-weight and can be implemented using wearable sensors [9]. This is crucial with a view to long-term recordings in mobile real-world settings.

7.1.1. Paper Scope and Contributions

The aim of this work is to assess the feasibility of recognising human activity using eye movement analysis, so-called eye-based activity recognition (EAR)\(^1\). The specific contributions are: (1) the introduction of eye movement analysis as a new sensing modality for activity recognition; (2) the development and characterisation of new algorithms for detecting three basic eye movement types from EOG signals (saccades, fixations, and blinks) and a method to assess repetitive eye movement patterns; (3) the development and evaluation of 90 features derived from these eye movement types; and (4) the implementation of a method for continuous EAR, and its evaluation using a multi-participant EOG dataset involving a study of five real-world office activities.

7.1.2. Paper Organisation

We first survey related work, introduce EOG, and describe the main eye movement characteristics that we identify as useful for EAR. We then detail and characterise the recognition methodology: the methods used for removing drift and noise from EOG signals, and the algorithms developed for detecting saccades, fixations, blinks, and for analysing repetitive eye movement patterns. Based on these eye movement characteristics, we develop 90 features; some directly derived from a particular characteristic, others devised to capture additional aspects of eye movement dynamics.

We rank these features using minimum redundancy maximum relevance feature selection (mRMR) and a support vector machine (SVM) classifier. To evaluate both algorithms on a real-world example, we devise an experiment involving a continuous sequence of five office activities, plus a period without

\(^{1}\) An earlier version of this paper was published in [10].
any specific activity (the NULL class). Finally, we discuss the findings gained from this experiment and give an outlook to future work.

7.2. Related Work

7.2.1. Electrooculography Applications

Eye movement characteristics such as saccades, fixations, and blinks, as well as deliberate movement patterns detected in EOG signals, have already been used for hands-free operation of static human-computer [11] and human-robot [12] interfaces. EOG-based interfaces have also been developed for assistive robots [13] or as a control for an electric wheelchair [14]. Such systems are intended to be used by physically disabled people who have extremely limited peripheral mobility but still retain eye-motor coordination. These studies showed that EOG is a measurement technique that is inexpensive, easy to use, reliable, and relatively unobtrusive when compared to head-worn cameras used in video-based eye trackers. While these applications all used EOG as a direct control interface, our approach is to use EOG as a source of information on a person’s activity.

7.2.2. Eye Movement Analysis

A growing number of researchers use video-based eye tracking to study eye movements in natural environments. This has led to important advances on our understanding of how the brain processes tasks, and of the role that the visual system plays in this [15]. Eye movement analysis has a long history as a tool to investigate visual behaviour. In an early study, Hacisalihzade et al. used Markov processes to model visual fixations of observers recognising an object [16]. They transformed fixation sequences into character strings and used the string edit distance to quantify the similarity of eye movements. Elhelw et al. used discrete time Markov chains on sequences of temporal fixations to identify salient image features that affect the perception of visual realism [17]. They found that fixation clusters were able to uncover the features that most attract an observer’s attention. Dempere-Marco et al. presented a method for training novices in assessing tomography images [18]. They modelled the assessment behaviour of domain experts based on the dynamics of their saccadic eye movements. Salvucci et al. evaluated means for automated analysis of eye movements [19]. They described three methods based on sequence-matching and hidden Markov models that interpreted eye movements as accurately as human experts but in significantly less time.

All of these studies aimed to model visual behaviour during specific tasks using a small number of well-known eye movement characteristics. They explored the link between the task and eye movements, but did not recognise the task or activity using this information.
7.2.3. Activity Recognition

In ubiquitous computing, one goal of activity recognition is to provide information that allows a system to best assist the user with his or her task [20]. Traditionally, activity recognition research has focused on gait, posture, and gesture. Bao et al. used body-worn accelerometers to detect 20 physical activities, such as cycling, walking and scrubbing the floor, under real-world conditions [21]. Logan et al. studied a wide range of daily activities, such as using a dishwasher, or watching television, using a large variety and number of ambient sensors, including RFID tags and infra-red motion detectors [22]. Ward et al. investigated the use of wrist worn accelerometers and microphones in a wood workshop to detect activities such as hammering, or cutting wood [4]. Several researchers investigated the recognition of reading activity in stationary and mobile settings using different eye tracking techniques [6, 23]. Our work, however, is the first to describe and apply a general-purpose architecture for EAR to the problem of recognising everyday activities.

7.3. Background

7.3.1. Electrooculography

The eye can be modelled as a dipole with its positive pole at the cornea and its negative pole at the retina. Assuming a stable corneo-retinal potential difference, the eye is the origin of a steady electric potential field. The electrical signal that can be measured from this field is called the electrooculogram (EOG).

If the eye moves from the centre position towards the periphery, the retina approaches one electrode while the cornea approaches the opposing one. This change in dipole orientation causes a change in the electric potential field and thus the measured EOG signal amplitude. By analysing these changes, eye movements can be tracked. Using two pairs of skin electrodes placed at opposite sides of the eye and an additional reference electrode on the forehead, two signal components (EOG$_h$ and EOG$_v$), corresponding to two movement components - a horizontal and a vertical - can be identified. EOG typically shows signal amplitudes ranging from 5 $\mu$V/degree to 20 $\mu$V/degree and an essential frequency content between 0 Hz and 30 Hz [24].

7.3.2. Eye Movement Types

To be able to use eye movement analysis for activity recognition, it is important to understand the different types of eye movement. We identified three basic eye movement types that can be easily detected using EOG: saccades, fixations, and blinks (see Fig. 7.1).
Figure 7.1: Denoised and baseline drift removed horizontal (EOG$_h$) and vertical (EOG$_v$) signal components. Examples of the three main eye movement types are marked in grey: saccades (S), fixations (F), and blinks (B).

Saccades

The eyes do not remain still when viewing a visual scene. Instead, they have to move constantly to build up a mental “map” from interesting parts of that scene. The main reason for this is that only a small central region of the retina, the fovea, is able to perceive with high acuity. The simultaneous movement of both eyes is called a saccade. The duration of a saccade depends on the angular distance the eyes travel during this movement: the so-called saccade amplitude. Typical characteristics of saccadic eye movements are 20 degrees for the amplitude, and 10 ms to 100 ms for the duration [25].

Fixations

Fixations are the stationary states of the eyes during which gaze is held upon a specific location in the visual scene. Fixations are usually defined as the time between each two saccades. The average fixation duration lies between 100 ms and 200 ms [26].

Blinks

The frontal part of the cornea is coated with a thin liquid film, the so-called “precornial tear film”. To spread this fluid across the corneal surface, regular opening and closing of the eyelids, or blinking, is required. The average blink rate varies between 12 and 19 blinks per minute while at rest [27]; it is influenced by environmental factors such as relative humidity, temperature or brightness, but also by physical activity, cognitive workload, or fatigue [28]. The average blink duration lies between 100 ms and 400 ms [29].
Figure 7.2: Architecture for eye-based activity recognition on the example of EOG. Light grey indicates EOG signal processing; dark grey indicates use of a sliding window.

7.4. Methodology

We first provide an overview of the architecture for EAR used in this work. We then detail our algorithms for removing baseline drift and noise from EOG signals, for detecting the three basic eye movement types, and for analysing repetitive patterns of eye movements. Finally, we describe the features extracted from these basic eye movement types, and introduce the minimum redundancy maximum relevance feature selection, and the support vector machine classifier.

7.4.1. Recognition Architecture

Fig. 7.2 shows the overall architecture for EAR. The methods were all implemented offline using MATLAB and C. Input to the processing chain are the two EOG signals capturing the horizontal and the vertical eye movement components. In the first stage, these signals are processed to remove any artefacts that might hamper eye movement analysis. In the case of EOG signals, we apply algorithms for baseline drift and noise removal. Only this initial processing depends on the particular eye tracking technique used; all further stages are completely independent of the underlying type of eye movement data. In the next stage, three different eye movement types are detected from the processed eye movement data: saccades, fixations, and blinks. The corresponding eye movement events returned by the detection algorithms are the basis for extracting different eye movement features using a sliding window.
In the last stage, a hybrid method selects the most relevant of these features, and uses them for classification.

### 7.4.2. EOG Signal Processing

#### Baseline Drift Removal

Baseline drift is a slow signal change superposing the EOG signal but mostly unrelated to eye movements. It has many possible sources, e.g. interfering background signals or electrode polarisation [30]. Baseline drift only marginally influences the EOG signal during saccades, however, all other eye movements are subject to baseline drift. In a five electrode setup, as used in this work (see Fig. 7.8), baseline drift may also differ between the horizontal and vertical EOG signal component.

Several approaches to remove baseline drift from electrocardiography signals (ECG) have been proposed (for example see [31–33]). As ECG shows repetitive signal characteristics, these algorithms perform sufficiently well at removing baseline drift. However, for signals with non-repetitive characteristics such as EOG, developing algorithms for baseline drift removal is still an active area of research. We used an approach based on wavelet transform [34]. The algorithm first performed an approximated multilevel 1-D wavelet decomposition at level nine using Daubechies wavelets on each EOG signal component. The reconstructed decomposition coefficients gave a baseline drift estimation. Subtracting this estimation from each original signal component yielded the corrected signals with reduced drift offset.

#### Noise Removal

EOG signals may be corrupted with noise from different sources, such as the residential power line, the measurement circuitry, electrodes and wires, or other interfering physiological sources such as electromyographic (EMG) signals. In addition, simultaneous physical activity may cause the electrodes to lose contact or move on the skin. As mentioned before, EOG signals are typically non-repetitive. This prohibits the application of denoising algorithms that make use of structural and temporal knowledge about the signal.

Several EOG signal characteristics need to be preserved by the denoising. First, the steepness of signal edges needs to be retained to be able to detect blinks and saccades. Second, EOG signal amplitudes need to be preserved to be able to distinguish between different types and directions of saccadic eye movements. Finally, denoising filters must not introduce signal artefacts that may be misinterpreted as saccades or blinks in subsequent signal processing steps.

To identify suitable methods for noise removal we compared three different algorithms on real and synthetic EOG data: a low-pass filter, a filter based on wavelet shrinkage denoising [35] and a median filter. By visual inspection of the denoised signal we found that the median filter performed best; it preserved edge steepness of saccadic eye movements, retained EOG signal amplitudes, and did not introduce any artificial signal changes. It is crucial,
however, to choose a window size $W_{mf}$ that is small enough to retain short signal pulses, particularly those caused by blinks. A median filter removes pulses of a width smaller than about half of its window size. By taking into account the average blink duration reported earlier, we fixed $W_{mf}$ to 150 ms.

7.4.3. Detection of Basic Eye Movement Types

Different types of eye movements can be detected from the processed EOG signals. In this work, saccades, fixations, and blinks form the basis of all eye movement features used for classification. The robustness of the algorithms for detecting these is key to achieving good recognition performance. Saccade detection is particularly important because fixation detection, eye movement encoding, and the wordbook analysis are all reliant on it (see Fig. 7.2). In the following, we introduce our saccade and blink detection algorithms and characterise their performance on EOG signals recorded under constrained conditions.

Saccade and Fixation Detection

For saccade detection, we developed the so-called Continuous Wavelet Transform - Saccade Detection (CWT-SD) algorithm (see Fig. 7.3 for an example). Input to CWT-SD are the denoised and baseline drift removed EOG signal components EOG$_h$ and EOG$_v$. CWT-SD first computes the continuous 1-D wavelet coefficients at scale 20 using a Haar mother wavelet. Let $s$ be one of these signal components and $\psi$ the mother wavelet. The wavelet coefficient $C^a_b(s)$ of $s$ at scale $a$ and position $b$ is defined:

$$C^a_b(s) = \int_{\mathbb{R}} s(t) \frac{1}{\sqrt{a}} \psi \left( \frac{t - b}{a} \right) dt.$$ 

By applying an application-specific threshold $th_{sd}$ on the coefficients $C^i(s) = C^{20}_i(s)$, CWT-SD creates a vector $M$ with elements $M_i$:

$$M_i = \begin{cases} 
1, & \forall i: C_i(s) < -th_{sd}, \\
-1, & \forall i: C_i(s) > th_{sd}, \\
0, & \forall i: -th_{sd} \leq C_i(s) \leq th_{sd}.
\end{cases}$$

This step divides EOG$_h$ and EOG$_v$ in saccadic ($M = 1, -1$) and non-saccadic (fixational) ($M = 0$) segments.

Saccadic segments shorter than 20 ms and longer than 200 ms are removed. These boundaries approximate the typical physiological saccade characteristics described in literature [25]. CWT-SD then calculates the amplitude, and direction of each detected saccade. The saccade amplitude $S_A$ is the difference in EOG signal amplitude before and after the saccade (c.f. Fig. 7.3). The direction is derived from the sign of the corresponding elements in $M$. Finally, each saccade is encoded into a character representing the combination
Figure 7.3: Continuous Wavelet Transform - Saccade Detection (CWT-SD) algorithm. (a) Denoised and baseline drift removed horizontal EOG signal during reading with example saccade amplitude ($S_A$); (b) the transformed wavelet signal ($EOG_{wl}$), with application-specific small ($\pm th_{small}$) and large ($\pm th_{large}$) thresholds; (c) marker vectors for distinguishing between small ($M_{small}$) and large ($M_{large}$) saccades, and (d) example character encoding for part of the EOG signal.

of amplitude and direction. For example, a small saccade in EOG$_h$ with negative direction gets encoded as “r” and a large saccade with positive direction as “L”.

Humans typically alternate between saccades and fixations. This allows us to also use CWT-SD for detecting fixations. The algorithm exploits the fact that gaze remains stable during a fixation. This results in the corresponding gaze points, i.e. the points in visual scene gaze is directed at, to cluster together closely in time. Therefore, fixations can be identified by thresholding on the dispersion of these gaze points [36]. For a segment $S$ of length $n$ comprised of a horizontal $S_h$ and a vertical $S_v$ EOG signal component, the dispersion is calculated as

$$Dispersion(S) = max(S_h) - min(S_h) + max(S_v) - min(S_v)$$

Initially all non-saccadic segments are assumed to contain a fixation. The algorithm then drops segments for which the dispersion is above a maximum
7.4. Methodology

threshold \( th_{fd} \) of 10,000, or if its duration is below a minimum threshold \( th_{fd} \) of 200 ms. The value of \( th_{fd} \) was derived as part of the CWT-SD evaluation; that of \( th_{fd} \) approximates the typical average fixation duration reported earlier.

A particular activity may require saccadic eye movements of different distance and direction. For example, reading involves a fast sequence of small saccades while scanning each line of text while large saccades are required to jump back to the beginning of the next line. We opted to detect saccades with two different amplitudes, “small” and “large”. This requires two thresholds \( th_{sd\_small} \) and \( th_{sd\_large} \) to divide the range of possible values of \( C \) into three bands (see Fig. 7.3): no saccade \((-th_{sd\_small} < C < th_{sd\_small})\), small saccade \((-th_{sd\_large} < C < -th_{sd\_small} \text{ or } th_{sd\_small} < C < th_{sd\_large})\), and large saccade \((C < -th_{sd\_large} \text{ or } C > th_{sd\_large})\). Depending on its peak value, each saccade is then assigned to one of these bands.

To evaluate the CWT-SD algorithm, we performed an experiment with five participants - one female and four male (age: 25 - 59 years, mean = 36.8, \( sd = 15.4 \)). To cover effects of differences in electrode placement and skin contact the experiment was performed on two different days; in between days the participants took off the EOG electrodes. A total of twenty recordings were made per participant, 10 per day. Each experiment involved tracking the participants’ eyes while they followed a sequence of flashing dots on a computer screen. We used a fixed sequence to simplify labelling of individual saccades. The sequence was comprised of 10 eye movements consisting of five horizontal and eight vertical saccades. This produced a total of 591 horizontal and 855 vertical saccades.

By matching saccade events with the annotated ground truth we calculated true positives \((TP)\), false positives \((FP)\) and false negatives \((FN)\), and from these, precision \((\frac{TP}{TP+FP})\), recall \((\frac{TP}{TP+FN})\), and the F1 score \((2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}})\). We then evaluated the F1 score across a sweep on the CWT-SD threshold \( th_{sd} = 1 \ldots 50 \) (in 50 steps) separately for the horizontal and vertical EOG signal components. Fig. 7.4 shows the mean F1 score over all five participants with vertical lines indicating the standard deviation for selected values of \( th_{sd} \). What can be seen from the figure is that similar thresholds were used to achieve the top F1 scores of about 0.94. It is interesting to note that the standard deviation across all participants reaches a minimum for a whole range of values around this maximum. This suggests that also thresholds close to this point can be selected that still achieve robust detection performance.

**Blink Detection**

For blink detection, we developed the *Continuous Wavelet Transform - Blink Detection* (CWT-BD) algorithm. Similar to CWT-SD, the algorithm uses a threshold \( th_{bd} \) on the wavelet coefficients to detect blinks in EOG\(_v\). In contrast to a saccade, a blink is characterised by a sequence of two large peaks in the coefficient vector directly following each other: one positive, the other negative. The time between these peaks is smaller than the minimum time be-
Chapter 7: Recognition of Office Activities

Figure 7.4: Evaluation of the CWT-SD algorithm for both EOG signal components using a sweep of its main parameter, the threshold $th_{sd}$. The figure plots the mean F1 score over all five participants; vertical lines show the standard deviation for selected $th_{sd}$. Maximum F1 score is indicated by a dashed line.

between two successive saccades rapidly performed in opposite direction. This is because typically, two saccades have at least a short fixation in between them. For this reason, blinks can be detected by applying a maximum threshold $th_{bd_t}$ on this time difference.

We evaluated our algorithm on EOG signals recorded in a stationary setting from five participants looking at different pictures (two female and three male, age: 25 - 29 years, mean = 26.4, sd = 1.7). We labelled a total of 706 blinks by visual inspection of the vertical EOG signal component. With an average blink rate of 12 blinks per minute, this corresponds to about 1 hour of eye movement data. We evaluated CWT-BD over sweeps of its two main parameters: $th_{bd} = 100 \ldots 50,000$ (in 500 steps), and $th_{bd_t} = 100 \ldots 1000$ ms (in 10 steps).

The F1 score was calculated by matching blink events with the annotated ground truth. Fig. 7.6 shows the F1 scores for five selected values of $th_{bd_t}$ over all participants. CWT-BD performs best with $th_{bd_t}$ between 400 ms and 600 ms while reaching top performance (F1 score: 0.94) using a $th_{bd_t}$ of 500 ms. Time differences outside this range, as exemplarily shown for 300 ms and 1000 ms, are already subject to a considerable drop in performance. This finding nicely reflects the values for the average blink duration cited earlier from literature.
Figure 7.5: (a) Characters used to encode eye movements of different direction and distance: dark grey indicates basic, light grey diagonal directions. (b) Saccades detected in both EOG signal components and mapped to the eye movement sequence of the jumping point stimulus. Simultaneous saccades in both components are combined according to their direction and amplitude (e.g. “l” and “u” become “n”, and “R” and “U” become “B”).

7.4.4. Analysis of Repetitive Eye Movement Patterns

Activities such as reading typically involve characteristic sequences of several consecutive eye movements [6]. We propose encoding eye movements by mapping saccades with different direction and amplitude to a discrete,
Figure 7.6: Evaluation of the CWT-BD algorithm over a sweep of the blink threshold $th_{bd}$, for five different maximum time differences $th_{bd,t}$. The figure plots the mean F1 score over all participants; vertical lines show the standard deviation for selected $th_{bd}$. Maximum F1 score is indicated by a dashed line.

character-based representation. Strings of these characters are then collected in wordbooks that are analysed to extract sequence information on repetitive eye movement patterns.

Eye Movement Encoding

Our algorithm for eye movement encoding maps the individual saccade information from both EOG components onto a single representation comprising 24 discrete characters (see Fig. 7.5a). This produces a representation that can be more efficiently processed and analysed.

The algorithm takes the CWT-SD saccades from the horizontal and vertical EOG signal components as its input. It first checks for simultaneous saccades in both components as these represent diagonal eye movements. Simultaneous saccades are characterised by overlapping saccade segments in the time domain. If no simultaneous saccades are detected, the saccade’s character is directly used to denote the eye movement. If two saccades are detected, the algorithm combines both according to the following scheme (see Fig. 7.5b): the characters of two saccades with equally large EOG signal amplitudes are merged to the character exactly in between (e.g. “l” and “u” become “n”, “R” and “U” become “B”). If simultaneous saccades differ by more than 50% in EOG signal amplitude, their characters are merged to the closest neighbouring character (e.g. “l” and “U” become “O”). This procedure encodes each eye movement into a distinct character, thus mapping saccades of both EOG signal components into one eye movement sequence.
7.4. Methodology

Figure 7.7: Example wordbook analysis for eye movement patterns of length $l = 3$. A sliding window scans a sequence of eye movements encoded into characters for repetitive patterns. Newly found patterns are added to the wordbook; otherwise only the occurrence count (last column) is increased by one.

**Wordbook Analysis**

Based on the encoded eye movement sequence, we propose a wordbook analysis to assess repetitive eye movement patterns (see Fig. 7.7). An eye movement pattern is defined as a string of $l$ successive characters. As an example with $l = 4$, the pattern “LrBd” translates to large left (L) $\rightarrow$ small right (r) $\rightarrow$ large diagonal right (B) $\rightarrow$ small down (d). A sliding window of length $l$ and a step size of one is used to scan the eye movement sequence for these patterns. Each newly found eye movement pattern is added to the corresponding wordbook $W_b_l$. For a pattern that is already included in $W_b_l$, its occurrence count is increased by one.

**7.4.5. Feature Extraction**

We extract four groups of features based on the detected saccades, fixations, blinks, and the wordbooks of eye movement patterns. Table 7.1 details the naming scheme used for all of these features. The features are calculated using a sliding window (window size $W_{fe}$ and step size $S_{fe}$) on both EOG$_h$ and EOG$_v$. From a pilot study, we were able to fix $W_{fe}$ at 30 s and $S_{fe}$ at 0.25 s.

Features calculated from saccadic eye movements make up the largest proportion of extracted features. In total, there are 62 such features comprising the mean, variance and maximum EOG signal amplitudes of saccades, and the normalised saccade rates. These are calculated for both EOG$_h$ and EOG$_v$; for small and large saccades; for saccades in positive or negative direction; and for all possible combinations of these.

We calculate five different features using fixations: the mean and variance of the EOG signal amplitude within a fixation; the mean and the variance of fixation duration; and the fixation rate over window $W_{fe}$.
### Table 7.1: Naming scheme for the features used in this work.

For a particular feature, e.g. *S-rateSPHor*, the capital letter represents the group - saccadic (S), blink (B), fixation (F) or wordbook (W) - and the combination of abbreviations after the dash describes the particular type of feature and the characteristics it covers.

<table>
<thead>
<tr>
<th>Group</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>saccade (S-)</td>
<td>mean (mean), variance (var) or maximum (max) EOG signal amplitudes (Amp) or rate (rate) of small (S) or large (L), positive (P) or negative (N) saccades in horizontal (Hor) or vertical (Ver) direction</td>
</tr>
<tr>
<td>fixation (F-)</td>
<td>mean (mean) and/or variance (var) of the horizontal (Hor) or vertical (Ver) EOG signal amplitude (Amp) within or duration (Duration) of a fixation or rate of fixations</td>
</tr>
<tr>
<td>blink (B-)</td>
<td>mean (mean) or variance (var) of the blink duration or blink rate (rate)</td>
</tr>
<tr>
<td>wordbook (W-)</td>
<td>wordbook size (size) or maximum (max), difference (diff) between maximum and minimum, mean (mean) or variance (var) of all occurrence counts (Count) in the wordbook of length (-lx)</td>
</tr>
</tbody>
</table>

For blinks, we extract three features: blink rate, and the mean and variance of the blink duration.

We use four wordbooks. This allowed us to account for all possible eye movement patterns up to a length of four (*l* = 4), with each wordbook containing the type and occurrence count of all patterns found. For each wordbook we extract five features: the wordbook size, the maximum occurrence count, the difference between the maximum and minimum occurrence counts, and the variance and mean of all occurrence counts.
7.4.6. Feature Selection and Classification

For feature selection, we chose a filter scheme over the commonly used wrapper approaches because of the lower computational costs and thus shorter runtime given the large dataset. We use minimum redundancy maximum relevance feature selection (mRMR) for discrete variables [37, 38]. The mRMR algorithm selects a feature subset of arbitrary size $S$ that best characterises the statistical properties of the given target classes based on the ground truth labelling. In contrast to other methods such as the $F$-test, mRMR also considers relationship between features during the selection. Amongst the possible underlying statistical measures described in literature, mutual information was shown to yield the most promising results and was thus selected in this work. Our particular mRMR implementation combines the measures of redundancy and relevance among classes using the mutual information difference (MID).

For classification, we chose a linear support vector machine. Our SVM implementation uses a fast sequential dual method for dealing with multiple classes [39, 40]. This reduces training time considerably while retaining recognition performance.

These two algorithms are combined into a hybrid feature selection and classification method. In a first step, mRMR ranks all available features (with $S = 90$). During classification, the size of the feature set is then optimised with respect to recognition accuracy by sweeping $S$.

7.5. Experiment

We designed a study to establish the feasibility of EAR in a real-world setting. Our scenario involved five office-based activities - copying a text, reading a printed paper, taking hand-written notes, watching a video, and browsing the web - and periods during which participants took a rest (the NULL class). We chose these activities for three reasons. First, they are all commonly performed during a typical working day. Second, they exhibit interesting eye movement patterns that are both structurally diverse, and that have varying levels of complexity. We believe they represent the much broader range of activities observable in daily life. Finally, being able to detect these activities using on-body sensors such as EOG may enable novel attentive user interfaces that take into account cognitive aspects of interaction such as user interruptibility or level of task engagement.

Originally we recorded 10 participants, but two were withdrawn due to poor signal quality: One participant had strong pathologic nystagmus. Nystagmus is a form of involuntary eye movement that is characterised by alternating smooth pursuit in one direction and saccadic movement in the other direction. The horizontal EOG signal component turned out to be severely affected by the nystagmus and no reliable saccadic information could be extracted. For the second participant, most probably due to bad electrode placement, the EOG signal was completely distorted.

All of the remaining eight participants (two female and six male), aged between 23 and 31 years ($mean = 26.1$, $sd = 2.4$) were daily computer users, reporting 6 to 14 hours of use per day ($mean = 9.5$, $sd = 2.7$). They were
Figure 7.8: (top) Electrode placement for EOG data collection (h: horizontal, v: vertical, r: reference). (bottom) Continuous sequence of five typical office activities: copying a text, reading a printed paper, taking hand-written notes, watching a video, browsing the web, and periods of no specific activity (the NULL class).

asked to follow two continuous sequences, each composed of five different, randomly ordered activities, and a period of rest (see bottom of Fig. 7.8). For these, no activity was required of the participants but they were asked not to engage in any of the other activities. Each activity (including NULL) lasted about five minutes, resulting in a total dataset of about eight hours.

7.5.1. Apparatus

We used a commercial EOG device, the Mobi8, from Twente Medical Systems International (TMSI). It was worn on a belt around each participant’s waist and recorded a four-channel EOG at a sampling rate of 128 Hz. Participants were observed by an assistant who annotated activity changes with a wireless remote control. Data recording and synchronisation was handled by the Context Recognition Network Toolbox [41].

EOG signals were picked up using an array of five 24 mm Ag/AgCl wet electrodes from Tyco Healthcare placed around the right eye. The horizontal signal was collected using one electrode on the nose and another directly across from this on the edge of the right eye socket. The vertical signal was collected using one electrode above the right eyebrow and another on the lower edge of the right eye socket. The fifth electrode, the signal reference, was placed in the middle of the forehead. Five participants (two female, three male) wore spectacles during the experiment. For these participants, the nose electrode was moved to the side of the left eye to avoid interference with the spectacles (see top of Fig. 7.8).

The experiment was carried out in an office during regular working hours. Participants were seated in front of two adjacent 17 inch flat screens with a resolution of 1280x1024 pixels on which a browser, a video player, a word
processor and text for copying were on-screen and ready for use. Free movement of the head and upper body was possible throughout the experiment.

### 7.5.2. Procedure

For the text copying task, the original document was shown on the right screen with the word processor on the left screen. Participants could copy the text in different ways. Some touch typed and only checked for errors in the text from time to time; others continuously switched attention between the screens or the keyboard while typing. Because the screens were more than half a meter from the participants’ faces, the video was shown full screen to elicit more distinct eye movements. For the browsing task, no constraints were imposed concerning the type of website or the manner of interaction. For the reading and writing tasks, a book (12 pt, one column with pictures) and a pad with a pen were provided.

### 7.5.3. Parameter Selection and Evaluation

The same saccade and blink detection parameters were used throughout the evaluation: $th_{bd} = 23, 438$, $th_{bd_s} = 390$ ms, $th_{sdlarge} = 13, 750$, and $th_{sdsmall} = 2, 000$. The selection of $th_{sdsmall}$ was based on the typical length of a short scan saccade during reading, and $th_{sdlarge}$ on the length of a typical newline movement.

Classification and feature selection were evaluated using a leave-one-person-out scheme: we combined the datasets of all but one participant and used this for training; testing was done using both datasets of the remaining participant. This was repeated for each participant. The resulting train and test sets were standardised to have zero mean and a standard deviation of one. Feature selection was always performed solely on the training set. The two main parameters of the SVM algorithm, the cost $C$ and the tolerance of termination criterion $\epsilon$, were fixed to $C = 1$ and $\epsilon = 0.1$. For each leave-one-person-out iteration, the prediction vector returned by the SVM classifier was smoothed using a sliding majority window. Its main parameter, the window size $W_{sm}$, was obtained using a parameter sweep and fixed at 2.4 s.

### 7.6. Results

#### 7.6.1. Classification Performance

SVM classification was scored using a frame-by-frame comparison with the annotated ground truth. For specific results on each participant, or on each activity, class-relative precision and recall were used.

Table 7.2 shows the average precision and recall, and the corresponding number of features selected for each participant. The number of features used varied from only nine features (P8) up to 81 features (P1). The mean performance over all participants was 76.1% precision and 70.5% recall.
reported the worst result, with both precision and recall below 50%. In contrast, P7 achieved the best result, indicated by recognition performance in the 80s and 90s and using a moderate-sized feature set.

Fig. 7.9 plots the classification results in terms of precision and recall for each activity and participant. The best results approach the top right corner while worst results are close to the lower left. For most activities, precision and recall fall within the top right corner. Recognition of reading and copying, however, completely fails for P4, and browsing also shows noticeably lower precision. Similar but less strong characteristics apply for the reading, writing, and browsing task for P5.

The summed confusion matrix from all participants, normalised across ground truth rows, is given in Fig. 7.10. Correct recognition is shown on the diagonal; substitution errors are off-diagonal. The largest between-class substitution errors not involving NULL fall between 12% and 13% of their class times. Most of these errors involve browsing that is falsely returned during 13% each of read, write, and copy activities. A similar amount is substituted by read during browse time.

### 7.6.2. Eye Movement Features

We first analysed how mRMR ranked the features on each of the eight leave-one-person-out training sets. The rank of a feature is the position at which mRMR selected it within a set. The position corresponds to the importance with which mRMR assesses the feature’s ability to discriminate between classes in combination with the features ranked before it. Fig. 7.11 shows the top 15 features according to the median rank over all sets (see Table 7.1 for a description of the type and name of the features). Each vertical bar represents the spread of mRMR ranks: for each feature there is one rank per training set. The most useful features are those found with the highest rank (close to one) for most training sets, indicated by shorter bars. Some features are not always included in the final result (e.g. feature 63 only appears in five sets). Equally, a useful feature that is ranked lowly by mRMR might be the one that improves a classification (e.g. feature 68 is spread between rank five and 26, but is included in all eight sets).

This analysis reveals that the top three features, as judged by high ranks for all sets, are all based on horizontal saccades: 47 (S-rateSPHor), 56 (S-maxAmpPHor), and 10 (S-meanAmpSHor). Feature 68 (F-rate) is used in all sets, seven of which rank it highly. Feature 63 (B-rate) is selected for five out of the eight sets, only one of which gives it a high rank. Wordbook features 77 (W-maxCount-l2) and 85 (W-maxCount-l3) are not used in one of the sets, but they are highly ranked by the other seven.

We performed an additional study into the effect of optimising mRMR for each activity class. We combined all training sets and performed a one-versus-many mRMR for each non-NULL activity. The top five features selected during this evaluation are shown in Table 7.3. For example, the table reveals that reading and browsing can be described using wordbook features. Writing requires additional fixation features. Watching video is characterised
7.6. Results

<table>
<thead>
<tr>
<th></th>
<th>P1 (m)</th>
<th>P2 (m)</th>
<th>P3 (m)</th>
<th>P4 (m)</th>
<th>P5 (m)</th>
<th>P6 (f)</th>
<th>P7 (f)</th>
<th>P8 (m)</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>76.6</td>
<td>88.3</td>
<td>83.0</td>
<td><strong>46.6</strong></td>
<td>59.5</td>
<td>89.2</td>
<td><strong>93.0</strong></td>
<td>72.9</td>
<td>76.1</td>
</tr>
<tr>
<td>Recall</td>
<td>69.4</td>
<td>77.8</td>
<td>72.2</td>
<td>47.9</td>
<td><strong>46.0</strong></td>
<td><strong>86.9</strong></td>
<td>81.9</td>
<td>81.9</td>
<td>70.5</td>
</tr>
<tr>
<td># Features</td>
<td><strong>81</strong></td>
<td>46</td>
<td>64</td>
<td>59</td>
<td>50</td>
<td>69</td>
<td>21</td>
<td><strong>9</strong></td>
<td>50</td>
</tr>
</tbody>
</table>

Table 7.2: Precision, recall, and the corresponding number of features selected by the hybrid mRMR/SVM method for each participant. The participants’ gender is given in brackets; best and worst case results are indicated in bold.
Table 7.3: Top five features selected by mRMR for each activity over all training sets (see Table 7.1 for details on feature names).

<table>
<thead>
<tr>
<th>rank</th>
<th>read</th>
<th>browse</th>
<th>write</th>
<th>video</th>
<th>copy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>W-maxCount-l2</td>
<td>S-rateSHor</td>
<td>W-varCount-l4</td>
<td>F-meanVarVertAmp</td>
<td>S-varAmp</td>
</tr>
<tr>
<td>2</td>
<td>W-meanCount-l4</td>
<td>W-varCount-l4</td>
<td>F-meanVarVertAmp</td>
<td>F-meanVarHorAmp</td>
<td>S-meanAmpSNHor</td>
</tr>
<tr>
<td>3</td>
<td>W-varCount-l2</td>
<td>W-varCount-l3</td>
<td>F-varDuration</td>
<td>B-rate</td>
<td>S-meanAmpLPHor</td>
</tr>
<tr>
<td>4</td>
<td>F-varDuration</td>
<td>W-varCount-l2</td>
<td>F-meanDuration</td>
<td>S-varAmpNHor</td>
<td>S-rateS</td>
</tr>
<tr>
<td>5</td>
<td>B-rate</td>
<td>W-meanCount-l1</td>
<td>S-rateLPVer</td>
<td>S-meanAmpSPHor</td>
<td>F-meanVarHorAmp</td>
</tr>
</tbody>
</table>
by a mixture of fixation and saccade features for all directions and - as reading - the blink rate, while copying involves mainly horizontal saccade features.

7.7. Discussion

7.7.1. Robustness Across Participants

The developed algorithms for detecting saccades and blinks in EOG signals proved robust and achieved F1 scores of up to 0.94 across several people (see Fig. 7.4 and 7.6). For the experimental evaluation, the parameters of both algorithms were fixed to values common for all participants; the same applies to the parameters of the feature selection and classification algorithms. Under these conditions, despite person-independent training, six out of the eight participants returned best average precision and recall values of between 69% and 93%.

Two participants, however, returned results that were lower than 50%. On closer inspection of the raw eye movement data, it turned out that for both the EOG signal quality was poor. Changes in signal amplitude for saccades and blinks - upon which feature extraction and thus recognition performance directly depend - were not distinctive enough to be reliably detected. As was found in an earlier study [6], dry skin or poor electrode placement are the most likely culprits. Still, the achieved recognition performance is promising for eye movement analysis to be implemented in real-world applications, for example, as part of a reading assistant, or for monitoring workload to assess the risk of burnout syndrome. For such applications, recognition performance may be further increased by combining eye movement analysis with additional sensing modalities.

7.7.2. Results for Each Activity

As might have been expected, reading is detected with comparable accuracy to that reported earlier [6]. However, the methods used are quite different. The string matching approach applied in the earlier study makes use of a specific “reading pattern”. That approach is not suited for activities involving less homogeneous eye movement patterns. For example, one would not expect to find a similarly unique pattern for browsing or watching a video as there exists for reading. This is because eye movements show much more variability during these activities as they are driven by an ever-changing stimulus. As shown here, the feature-based approach is much more flexible and scales better with the number and type of activities that are to be recognised.

Accordingly, we are now able to recognise four additional activities - web browsing, writing on paper, watching video, and copying text - with almost, or above, 70% precision and 70% recall. Particularly impressive is video, with an average precision of 88% and recall of 80%. This is indicative of a task where the user might be concentrated on a relatively small field of view (like reading), but follows a typically unstructured path (unlike reading). Similar examples outside the current study might include interacting with a graphical
user interface or watching television at home. Writing is similar to reading in that the eyes follow a structured path, albeit at a slower rate. Writing involves more eye “distractions” - when the person looks up to think, for example. Browsing is recognised less well over all participants (average precision 79%, recall 63%) - but with a large spread between people. A likely reason for this is that it is not only unstructured, but that it involves a variety of sub-activities - including reading - that may need to be modelled. The copy activity, with an average precision of 76% and a recall of 66%, is representative of activities with a small field of view that include regular shifts in attention (in this case to another screen). A comparable activity outside the chosen office scenario might be driving, where the eyes are on the road ahead with occasional checks to the side mirrors. Finally, the NULL class returns a high recall of 81%. However, there are many false returns (activity false negatives) for half of the participants, resulting in a precision of only 66%.

Three of these activities - writing, copying, and browsing - all include sections of reading. From quick checks over what has been written or copied, to longer perusals of online text, reading is a pervasive sub-activity in this scenario. This is confirmed by the relatively high rate of confusion errors involving reading as shown in Fig. 7.10.

7.7.3. Feature Groups

The feature groups selected by mRMR provide a snapshot of the types of eye movement features useful for activity recognition.

Features from three of the four proposed groups - saccade, fixation, and wordbook - were all prominently represented in our study. The fact that each group covers complementary aspects of eye movement is promising for the general use of these features for other EAR problems. Note that no-one feature type performs well alone. The best results were obtained using a mixture of different features. Among these, the fixation rate was always selected. This result is akin to that of Canosa et al. who found that both fixation duration and saccade amplitude are strong indicators of certain activities [42].

Features derived from blinks are less represented in the top ranks. One explanation for this is that for the short activity duration of only five minutes the participants did not become fully engaged in the tasks, and were thus less likely to show the characteristic blink rate variations suggested by Palomba et al. [43]. These features may be found to be more discriminative for longer duration activities. Coupled with the ease by which they were extracted, we believe blink features are still promising for future work.

7.7.4. Features for Each Activity Class

The analysis of the most important features for each activity class is particularly revealing.

Reading is a regular pattern characterised by a specific sequence of saccades and short fixations of similar duration. Consequently, mRMR chose mostly wordbook features describing eye movement sequencing in its top ranks, as well as a feature describing the fixation duration variance. The fifth
feature, the blink rate, reflects that for reading as an activity of high visual engagement people tend to blink less [43].

Browsing is structurally diverse and - depending on the website being viewed - may be comprised of different activities, e.g. watching a video, typing or looking at a picture. In addition to the small, horizontal saccade rate, mRMR also selected several workbook features of varying lengths. This is probably due to our participants’ browsing activities containing mostly sequences of variable length reading such as scanning headlines or searching for a product in a list.

Writing is similar to reading, but requires greater fixation duration (it takes longer to write a word than to read it) and greater variance. mRMR correspondingly selected average fixation duration and its variance as well as a workbook feature. However, writing is also characterised by short thinking pauses, during which people invariably look up. This corresponds extremely well to the choice of the fixation feature that captures variance in vertical position.

Watching a video is a highly unstructured activity, but is carried out within a narrow field of view. The lack of wordbook features reflects this, as does the mixed selection of features based on all three types: variance of both horizontal and vertical fixation positions, small positive and negative saccadic movements, and blink rate. The use of blink rate likely reflects the tendency towards blink inhibition when performing an engaging yet sedentary task [43].

Finally, copying involves many back and forth saccades between screens. mRMR reflects this by choosing a mixture of small and large horizontal saccade features, as well as variance in horizontal fixation positions.

These results suggest that for tasks that involve a known set of specific activity classes, recognition can be optimised by only choosing features known to best describe these classes. It remains to be investigated how well such prototype features discriminate between activity classes with very similar characteristics.

### 7.7.5. Activity Segmentation Using Eye Movements

Segmentation - the task of spotting individual activity instances in continuous data - remains an open challenge in activity recognition. We found that eye movements can be used for activity segmentation on different levels depending on the timescale of the activities. The lowest level of segmentation is that of individual saccades that define eye movements in different directions - "left", "right", and so on. An example for this is the end-of-line “carriage return” eye movement performed during reading. The next level includes more complex activities that involve sequences composed of a small number of saccades. For these activities, the wordbook analysis proposed in this work may prove suitable. In earlier work, such short eye movement patterns, so-called eye gestures, were successfully used for eye-based human-computer interaction [44]. At the highest level, activities are characterised by complex combinations of eye movement sequences of potentially arbitrary length. Unless wordbooks are used that span these long sequences, dynamic modelling
of activities is required. For this it would be interesting to investigate methods such as hidden Markov models (HMM), Conditional Random Fields (CRF), or an approach based on eye movement grammars. These methods would allow us to model eye movement patterns at different hierarchical levels, and to spot composite activities from large streams of eye movement data more easily.

### 7.7.6. Limitations

One limitation of the current work is that the experimental scenario considered only a handful of activities. It is important to note, however, that the recognition architecture and feature set were developed independently of these activities. In addition, the method is not limited to EOG. All features can be extracted equally well from eye movement data recorded using a video-based eye tracker. This suggests that our approach is applicable to other activities, settings, and eye tracking techniques.

The study also reveals some of the complexity one might face in using the eyes as a source of information on a person’s activity. The ubiquity of the eyes’ involvement in everything a person does means that it is challenging to annotate precisely what is being “done” at any one time. It is also a challenge to define a single identifiable activity. Reading is perhaps one of the easiest to capture because of the intensity of eye focus that is required and the well defined paths that the eyes follow. A task such as web browsing is more difficult because of the wide variety of different eye movements involved. It is challenging, too, to separate relevant eye movements from momentary distractions.

These problems may be solved, in part, by using video and gaze tracking for annotation. Activities from the current scenario could be redefined at a smaller timescale, breaking browsing into smaller activities such as “use scrollbar”, “read”, “look at image”, or “type”. This would also allow us to investigate more complicated activities outside the office. An alternative route is to study activities at larger timescales, to perform situation analysis rather than recognition of specific activities. Long-term eye movement features, e.g. the average eye movement velocity and blink rate over one hour, might reveal whether a person is walking along an empty or busy street, whether they are at their desk working, or whether they are at home watching television. Annotation will still be an issue, but one that may be alleviated using unsupervised or self-labelling methods [21, 45].

### 7.7.7. Considerations for Future Work

Additional eye movement characteristics that are potentially useful for activity recognition - such as pupil dilation, microsaccades, vestibulo-ocular reflex, or smooth pursuit movements - were not used here because of the difficulty in measuring them with EOG. These characteristics are still worth investigating in the future as they may carry information that complements that available in the current work.
Eye movements also reveal information on cognitive processes of visual perception, such as visual memory, learning, or attention. If it were possible to infer these processes from eye movements, this may lead to cognitive-aware systems that are able to sense and adapt to a person’s cognitive state.

7.8. Conclusion

This work reveals two main findings for activity recognition using eye movement analysis. First, we show that eye movements alone, i.e. without any information on gaze, can be used to successfully recognise five office activities. We argue that the developed methodology can be extended to other activities. Second, good recognition results were achieved using a mixture of features based on the fundamentals of eye movements. Sequence information on eye movement patterns, in the form of a wordbook analysis, also proved useful and can be extended to capture additional statistical properties. Different recognition tasks will likely require different combinations of these features.

The importance of these findings lies in their significance for eye movement analysis to become a general tool for the automatic recognition of human activity.
Figure 7.9: Precision and recall for each activity and participant. Mean performance (P1 to P8) is marked by a star.

Figure 7.10: Summed confusion matrix from all participants, normalised across ground truth rows.
Figure 7.11: Top 15 features selected by mRMR for all eight training sets. X-axis shows feature number and group; the key on the right shows the corresponding feature names as described in Table 7.1; Y-axis shows the rank (top = 1). For each feature, the bars show: the total number of training sets for which the feature was chosen (bold number at the top), the rank of the feature within each set (dots, with a number representing the set count), and the median rank over all sets (black star). For example, a useful feature is 47 (S) - a saccadic feature selected for all sets, in 7 of which it is ranked 1 or 2; less useful is 63 (B) - a blink feature used in only 5 sets and ranked between 4 and 29.
Chapter 7: Recognition of Office Activities
Bibliography


Chapter 7: Recognition of Office Activities


What’s in the Eyes for Context-Awareness?

Andreas Bulling, Daniel Roggen and Gerhard Tröster

Publication title: What’s in the eyes for context-awareness?

DOI: 10.1109/MPRV.2010.49
Abstract

Research in pervasive computing has investigated a variety of modalities for developing context-aware systems. A rich source of information on context that has not yet been used is the movements of the eyes. Because we use our eyes in almost everything that we do - everywhere - eye movements are a promising modality that ought to be exploited for context-awareness. Moreover, the link between eye movements and cognition may allow us to develop pervasive computing systems that are able to derive the so-called cognitive context of a person. In this work, we first report on the state-of-the-art in eye-based activity recognition (EAR) and outline the potential but also the research challenges of inferring activity and the cognitive context from eye movements. We present results of two EAR studies and first results towards our vision of cognitive-awareness.

8.1. Introduction

Context-awareness has emerged as a key area of research in mobile and pervasive computing. Besides location, physical activity is widely considered to be one of the most important contextual cues [1]. Over the last decade a large body of research in activity recognition has addressed a variety of problem domains and applications. Physical activity in indoor environments was shown to be recognisable using ambient sensors such as video cameras, reed switches, or sound. In mobile settings body-worn sensors were successfully applied to detect physical activity. As body movements are directly related to a person’s physical activities, motion sensing is typically performed using accelerometers or gyroscopes.

A rich source of information on context that has not been used so far is the movements of the eyes. The dynamics of eye movements as we carry out specific activities have the potential to reveal much about the activities themselves (e.g. while reading). In the same manner, specific environments or locations influence our eye movements, e.g. while driving a car.

Finally, eye movements have the particularity that they can be consciously controlled but that they are also unconsciously generated by the brain. Unconscious eye movements are strongly related to cognitive processes of visual perception such as attention, visual memory or learning. Eye movement analysis is a promising modality to infer these processes in real-world settings. Eventually, this may allow us to extend the current notion of context with a cognitive dimension, leading to cognitive-aware systems that enable novel types of user interaction not possible today.

In this work we first report on two case studies that illustrate the current state-of-the-art in eye-based activity recognition (EAR). We then introduce our vision of cognitive-awareness and outline the potential of using eye movement analysis to infer the cognitive context of a person. As a first step towards this vision, we present results from a study on assessing visual memory recall processes from eye movements. Finally, we discuss the lessons learned and
8.2. Tracking Eye Movements

To infer context from eye movements, these movements first need to be tracked. There exist well-known tools to track gaze direction (the “what does one look at”), particularly in the field of human-computer interaction (HCI). Stationary video-based eye trackers - the most recent systems even integrated into computer screens - are widely available and extensively used in a variety of commercial applications. For example, in 2006, Toyota presented a driver monitoring system that analysed gaze direction to warn a car driver if he was not paying attention to the road anymore. Tobii Technology in Stockholm, Sweden, sells eye trackers that are used in market research and usability studies to analyse what attracts the customers’ attention and optimise product placement or improve website designs.

Research in eye-based HCI traditionally focused on direct manipulation of user interfaces using gaze tracking in stationary settings. For example, gaze has been successfully used as a computer input [2], and as a means for interactive trip planning for tourists [3]. A growing number of researchers investigate gaze direction in mobile daily life environments. Tracking eye movements in such settings, however, is a much more difficult problem. This is mainly due to the fact that the development of wearable eye trackers that are robust to physical activity and that allow for long-term recordings is still an active field of research (see Box 1).

8.3. Eye Movement Analysis

The complementary - but also less common - approach to using gaze direction is to analyse the dynamics of eye movements over time (the “how does one look at”). Eye movements can generally be categorised as conscious, unconscious, or as a combination of both [4]. Conscious eye movements are those we are most aware of, as we use them to deliberately direct gaze at certain points of interest. Unconscious eye movements are generated by the oculomotor plant, a visual neural system in the brain. For most natural tasks, an interaction between conscious and unconscious eye movements can be observed. For example, reading is a conscious visual activity but partially involves unconscious eye movements trained while acquiring reading skills.

Outside pervasive computing, eye movement analysis has a long history as a means to investigate visual perception. For example, researchers found that by analysing the sequence of gaze points, features that are most salient in a picture - and thus attract an observer’s attention - can be identified [5]. Others showed that it is possible to support the training of novice doctors in assessing tomography images by modelling the visual behaviour of domain experts based on the dynamics of their eye movements [6]. These studies are only two examples out of many that analysed and modelled eye movement
characteristics during specific tasks. In pervasive computing, however, eye movement analysis has so far not been used for context recognition.

8.4. Context Recognition Using Eye Movement Analysis

The approach taken in our work is to use machine learning techniques to map eye movement data to a defined set of context classes. In a training phase, eye movements are recorded while a person experiences a certain situation of interest. During operation, eye movements are then compared to those observed during training to decide for the most similar context class.

Figure 8.1 shows this architecture for eye-based context recognition. In the first stage of this architecture eye movement data is acquired. Depending on application requirements, different measurement techniques, such as video or Electrooculography (EOG), can be used. In the second stage, the data is preprocessed to remove any artefacts that might hamper eye movement analysis. This preprocessing directly depends on the particular recording technique. In the case of EOG signals we typically employ denoising and signal drift removal. From the preprocessed eye movement data different eye movement characteristics can be detected. These may include blinks, fixations, or saccades; or additional characteristics that cover specific aspects of eye movement dynamics. In practice, directly using these eye movement characteristics is challenging. Therefore, we rather rely on features calculated from these characteristics for classification (see Box 2).

Based on this methodology, in the following, we report on two case studies that address different activity recognition problems in mobile and stationary settings.

8.5. Case Study I: Recognition of Reading

In the first case study, we applied the methodology to the problem of recognizing reading of people in transit in everyday environments using a wearable EOG system (for details see [7]). Reading is a pervasive activity, e.g. on computer screens at work, advertisements and signs in public, or books read at home or while travelling. Thus, information on a person’s reading activities is a useful indicator of his daily situation as well as a gauge of task engagement and attention. Attentive user interfaces could comprise the current level of user interruptability or provide assistance to people with reading disabilities by automatically magnifying or explaining words or context in the text.

We defined a scenario of travelling to and from work. The experiment involved eight participants occasionally reading text during different modes of locomotion including sitting at a desk, walking along a street, waiting at a tram stop and riding a tram (see Figure 8.2a). Each participant was followed by an assistant who annotated both the current mode of locomotion, and whether the participant was reading. To be able to detect if the participant’s eyes were on the page or not the assistant had to monitor the participant
from close proximity. To avoid distractions, we used the wireless controller Wii Remote from Nintendo for labelling. In total, we recorded an EOG dataset of roughly six hours with reading occurring about half of the time. This required spotting reading in a dataset with more than 50% of other types of eye movements.

8.5.1. Lessons Learned

Reading is a regular pattern characterised by frequent, short scan saccades during reading, and less frequent, longer newline movements. We therefore chose to analyse the left and right saccadic movements of the eyes. We recognised reading whenever a sequence of left and right saccades occurred in proportions close to those measured during training. Using person-independent training, we achieved an accuracy of 80.2% over all participants. The resulting precision and recall values for each of the three modes of locomotion are shown in Figure 8.2b.

The main finding of this study was that EOG is a feasible measurement technique for recognising reading in daily life scenarios. The results also showed that EOG is robust for different participants across a set of typical modes of locomotion. We found the main advantage of EOG to be the fact that the participants only had to wear relatively light-weight equipment. This contributed to the participants not feeling constrained and allowed for natural reading behaviour. One drawback was that EOG electrodes needed to be stuck to the face. This may have been regarded as inconvenient. In a post-experiment questionnaire, however, the participants reported that they did not feel constrained neither by the electrodes nor by the connecting wires.

8.6. Case Study II: Recognition of Office Activities

In a second study, we investigated the recognition of a set of typical office activities from eye movements recorded using EOG (see [8, 9] for details). Eight participants took part in this study. The participants were involved in two continuous activity sequences each lasting for about 30 minutes. This resulted in a total dataset of about eight hours. Each sequence was comprised of five different activities performed in random order: copying a text between two screens, reading a printed paper, taking hand-written notes, watching a video, and browsing the web. In addition, we included a period of rest (the NULL class); for this period, no activity was requested from the participants but they were asked not to engage in any of the previous activities. We chose these activities for two reasons. First, they are all commonly performed during a typical working day. Second, they exhibit interesting eye movement patterns that are both structurally diverse, and that have varying levels of complexity. We believe that by their nature - some highly structured (such as reading), others less structured (such as watching a video) - these activities are a representative subset of the broad range of activities observable in daily life.

We carried out the experiment in a well-lit office during regular working hours. Participants were seated in front of two 17 inch flat screens with a res-
olution of 1280x1024 pixels on which a browser, a video player, a word processor and text for copying were on-screen and ready for use. Sheets of paper and a pen were available on the desk close to the participants. Free movement of the head and upper body were possible throughout the experiment.

For classification, we used a support vector machine (SVM) classifier. We developed 90 features based on three of the main eye movement types: saccades, fixations, and blinks. In addition, we devised features that capture information on repetitive patterns of eye movements (see Box 2). Figure 8.3 shows an example activity sequence, the corresponding horizontal and vertical EOG signals, four example eye movement features, and the final classifier output. As we can see, these features reflect characteristic differences in the eye movements performed during some of the activities. Copying - a combination of reading and jumping between screens - can be characterised by a high saccade amplitude variance (F22), and a high maximum horizontal saccade amplitude (F56). Reading involves a large number of small horizontal saccades (F47) and a low mean fixation duration (F66). In contrast, watching a video or browsing are less well-structured activities and can hardly be distinguished only from the features shown here.

8.6.1. Lessons Learned

Using person-independent training, we achieved an average precision of 76.1% and recall of 70.5% over all classes and participants. Reading was a pervasive activity also in this study - from quick checks over what has been written or copied, to reading longer text on a website or subtitles in the video. Consequently, this led to confusions with browsing that involves a variety of sub-activities including reading (see black squares in Figure 8.4).

This study provided useful insights for the general problem of activity recognition using eye movement analysis.

Without any information on gaze direction, eye movement analysis can already serve as an alternative sensing modality for recognising human activity. Good recognition performance required to use a combination of several eye movement features. Information on repetitive patterns of eye movements proved useful, and can probably be extended to capture additional statistical properties. As different recognition tasks likely require different combinations of features, we recommend that a mixture of feature types be considered for each new task.

8.7. Extending Context With a Cognitive Dimension

The findings gained from both studies underline the significance of eye movement analysis for context-awareness. The developed feature set and recognition methodology is person-independent and not limited to the chosen settings, activities, or eye tracking equipment. We therefore believe that eye movement analysis has the potential to be successfully applied to other con-
text recognition problems, in different settings, and for a broader range of visual and physical activities. Eventually, analysing the link between unconscious eye movements and cognition may even pave the way for a new genre of pervasive computing systems that are able to sense and react to what we call the cognitive context of a person.

The cognitive context includes all aspects related to mental information processing, such as engagement, memory, knowledge or learning. We define a computing system as cognitive-aware if it is able to sense and adapt to the cognitive context of a person. Let us consider the following application: A business reception is held in a room equipped with a cognitive-aware ubiquitous memory assistant system. The attendees wear eye tracking sensors that allow the system to analyse the eye movement patterns people perform while looking at the faces of others. From these patterns, the system can assess the result of memory recall processes, and can detect whether attendees know and still remember each other. It can then automatically collect contact information about new acquaintances or send mnemonics about people a particular person has met before but does not remember anymore.

This example application goes beyond matching faces to a database of previously encountered persons. It can only be implemented by extending the current notion of context with a cognitive dimension. Current context-aware systems have a hard time getting at the cognitive context in an unobtrusive manner. The cognitive context is encoded in complex neural dynamics, and few obvious cues are accessible by non-invasive measurement techniques.

8.8. Cognitive Context From Eye Movements

In experimental psychology, a large body of research has evidenced that unconscious eye movements are strongly related to the underlying cognitive and perceptive processes. For example, it has been shown that eye movements correlate with the type of memory access required to perform certain tasks, and are good measures of visual engagement [10] and drowsiness [11]. Differences in eye movement patterns were also found for persons looking at familiar and unfamiliar faces [12], and for doctors with different proficiencies in assessing tomography images. These findings show the rich information content available in eye movements related to cognition.

As a first step towards the example application sketched above and our vision of cognitive-awareness, we conducted an experiment to investigate the feasibility of assessing visual memory while looking at familiar and unfamiliar pictures. The experiment involved six participants (two female and four male). They looked at four continuous sequences of pictures showing four categories of photographs: abstract images, buildings, faces, and landscapes (see Figure 8.5a). We ensured that pictures in each category had similar visual features. For example, we selected landscape photographs that showed a lake as their main feature; faces and buildings were always centred in the picture. Within each sequence, 12 pictures were presented only once; five others were chosen and presented four times at regular intervals. We randomised this sequence across participants. The exposure time for each picture was 10
seconds; in between each exposure, a picture with Gaussian noise was shown for five seconds as a baseline measurement. The pictures were shown on a screen using a beamer resulting in a picture dimension of between 1x1 m and 1.5 x 1.5 m. Participants were seated 2 m in front of the screen facing its centre. Movements of the upper body were allowed at any time during the experiment. However, we encouraged the participants to sit still.

We recorded the participants’ eye movements using EOG. We then extracted a number of features including those known from psychology literature to be linked to visual perception and memory recall processes. Figure 8.5b shows one of these features, the fixation count, for looking at faces with 0, 1, 2, and 3 prior exposures, averaged over all participants. As we can see, the mean fixation count decreased significantly with the number of prior exposures (significance level $p < 0.05$). This finding is akin to that of Heisz et al. who reported a similar correlation using a stationary video-based eye tracker [12]. Compared to that work, the standard deviation across participants was larger in our study. We believe this is due to the limited dataset and may be improved by using a larger number of participants and an exposition to longer sequences of pictures.

Nevertheless, the study revealed two important findings regarding the link between eye movements and (visual) memory recall processes. First, it is feasible to capture eye movement characteristics that reflect these processes using on-body sensors such as EOG; data acquisition and analysis is not limited to stationary video-based eye tracking systems. This finding is important in that it supports the use of wearable sensors for recording eye movements in mobile settings. Second, depending on the particular visual stimulus, only one eye movement feature - in this case the mean fixation count - may already be enough to assess memory recall processes of a person. As a next step, we plan to analyse combinations of several eye movement features, and machine learning techniques to automatically detect and quantify such memory recall processes in controlled settings.

### 8.9. Challenges

While these initial results are promising, developing cognitive-aware systems for real-world applications certainly faces several challenges.

First, assessing the cognitive context requires to employ an appropriate experimental methodology. This methodology will bear more similarity to that used in experimental psychology rather than that used in pervasive computing. In particular, specific cognitive processes first need to be evoked reliably and measured in controlled settings before they can eventually be inferred in complex daily life situations.

Second, eye movement characteristics reflecting different cognitive processes need to be identified, extracted from eye movement data, and automatically analysed. This is likely to require domain-specific modelling and machine learning approaches. In the simplest case, this means combining and adapting existing recognition methods for this new problem domain as we showed here. However, research on cognitive-awareness will also require and
drive the development of new methods particularly geared towards cognitive context evaluation. This will probably require to include mechanisms to adapt to a person’s specific eye movement characteristics.

Third, new questions in terms of engineering pervasive cognitive-aware environments need to be addressed. For example, interaction with artefacts that adapt to a person’s cognitive context will open up new areas of research, particularly in HCI and design.

Finally, even if eye movements are only used to recognise activity, knowing that eye movements are influenced by cognitive processes requires to consider ethical and privacy issues. The benefits of a cognitive-aware system - such as the memory assistant presented here - need to be weighted against potential downsides. These issues are not unlike those raised by human activity recognition in pervasive computing environments. The wearable computing answer to these concerns may be to keep this information “on body” at first for a particular person.

Moreover, there are challenges associated with the co-influence of activity, situation, and cognitive processes on a person’s eye movements. In the above case studies, either the activity or the cognitive process were predominant. This allowed us to consider each aspect separately. In real-world applications, however, eye movements are subject to a joint influence of activity, situation, and cognitive context. It is important to identify and separate these sources of influence for robust eye-based context recognition. We believe future research will therefore require a multidisciplinary approach at the crossroads of cognitive sciences, psychology, machine learning, and engineering.

8.10. Conclusion

In this work we have shown that eye movement analysis is a rich modality for context-awareness. Findings from two activity recognition studies confirmed that the developed recognition methodology is robust to different people and settings using person-independent calibration and training. We introduced our vision of cognitive-awareness and presented the first results towards using eye movement analysis to infer the so-called cognitive context of a person. Eventually, eye movement analysis - along with other measurement techniques such as portable Electroencephalography (EEG) or functional near infrared spectroscopy (fNIRs) - may allow us to develop cognitive-aware pervasive computing systems - a new genre of systems that are able to sense and adapt to a person’s cognitive context.

8.11. Box One: Sensing Solutions for Wearable Eye Tracking

The acquisition of eye movement data in daily life situations calls for highly miniaturised, low-power eye trackers with real-time processing capabilities. These requirements are increasingly addressed by commercial video-based systems. Some of these are targeted at mobile use such as the Mobile Eye
Chapter 8: What’s in the Eyes for Context-Awareness?

from Applied Science Laboratories (www.asleyetracking.com) and the iView X HED from SensoMotoric Instruments (www.smivision.com). Efforts to miniaturise video-based eye trackers led researchers to consider alternative measurement techniques. Among these, Electrooculography (EOG) is probably one of the more well-known ones. Using electrodes attached to the skin around the eyes, EOG measures changes in the electric potential field caused by eye movements. By analysing these changes, eye movements can be tracked. Several mobile systems have been presented such as headphones with integrated electrode arrays [13] or a head cap with EOG electrodes embroidered of silver coated thread [14]. We have demonstrated an EOG-based wearable eye tracker implemented as ordinary goggles [15]. This self-contained device uses dry electrodes integrated into the goggles frame and a small pocket-worn component with a digital signal processor for real-time EOG signal processing. Onboard data storage and low-power design allow for more than seven hours of mobile data recording and online eye movement analysis (see Figure 8.6).

8.12. Box Two: Eye Movement Characteristics and Features

To be able to use eye movement analysis for context recognition, it is important to understand the different types of eye movement. We identified six movement types potentially useful for context recognition; currently, however, we only rely on three of them: saccades, fixations, and blinks (see Figure 8.7 for some examples). For each movement type, different features can be extracted that reflect eye movement dynamics (see [8] for details on the signal processing required to extract these features).

8.12.1. Saccades

The eyes move constantly in saccades to build up a mental “map” from interesting parts of the visual scene. The main reason for this is that only a small central region of the retina, the fovea, is able to perceive with high acuity. We used a total of 62 saccadic features (S) such as the mean, variance and maximum EOG signal amplitudes of saccades, and normalised saccade rates. All of these features can be calculated for the horizontal and the vertical movement direction, for small and large saccades, for saccades in positive or negative direction, and for all combinations of these.

In addition, we developed a wordbook encoding scheme to analyse repetitive patterns of eye movements. This scheme creates wordbooks that hold statistics on the occurrence counts and type of all movement patterns of a particular length that occur in an eye movement dataset. For such a wordbook we used five features: the wordbook size, the maximum occurrence count, the difference between the maximum and minimum occurrence counts, and the variance and mean of all occurrence counts.
8.12. Box Two: Eye Movement Characteristics and Features

8.12.2. Fixations

A fixation is the static state of the eyes during which gaze is held upon a specific location. Humans typically alternate saccadic eye movements and fixations. For each fixation, we used the following five features (F): the mean and the variance of the EOG signal amplitude within the fixation; the mean and the variance of the fixation duration, and the fixation rate in the window.

8.12.3. Blinks

The frontal part of the cornea is coated with a thin liquid film, the “precornial tear film”. To spread this lacrimal fluid across the corneal surface regular blinking is required. The average blink rate is influenced by environmental factors (e.g. relative humidity, temperature, brightness), but also by physical activity, cognitive workload or fatigue. We used three different blink features (B): blink rate, and the mean and variance of blink duration.

8.12.4. Microsaccades

Microsaccades are fast involuntary eye movements of small amplitude that occur during prolonged visual fixations. The role of microsaccades in visual perception is still a highly debated topic among vision researchers. Typically, microsaccade amplitudes vary over only one to two minutes of arc [4]. While microsaccades can be detected with recent video-based eye trackers, signal artefacts still prevent their detection using EOG.

8.12.5. Vestibulo-Ocular Reflex

The vestibulo-ocular reflex (VOR) is a very fast eye movement triggered to stabilise gaze on a stationary object during head movements. The VOR compensates for these movements by moving the eye in the opposite direction to the head movement. The VOR is difficult to differentiate from saccades only using EOG, i.e. without any information on head movements. So, eye movements caused by the VOR were not explicitly used.

8.12.6. Smooth Pursuit Movements

Smooth pursuit movements are voluntarily performed when stabilising gaze on a moving visual target. Several psychological deficits have noticeable effects on the velocity of smooth pursuit movements, e.g. schizophrenia, autism, or post traumatic stress disorder. Due to similar signal characteristics, smooth pursuit movements are hard to separate from EOG signal drift.
Figure 8.1: Recording, preprocessing (details shown for electrooculography), event detection, and feature extraction architecture for context recognition using eye movement analysis.
Figure 8.2: (a) Reading during different modes of locomotion. (b) Precision and recall for recognising reading while sitting, standing, and walking.
Figure 8.3: Example sequence of office activities, EOG signals, example features, and classifier output.
Figure 8.4: Colour-coded classifier confusion for eye-based recognition of office activities (with an example confusion marked with black squares). This classification confirms specific eye movement characteristics in different activities.
Figure 8.5: (a) Example sequences with alternating pictures from four categories and Gaussian noise. (b) Mean fixation counts for faces with 0, 1, 2, and 3 prior exposures. The error bars represent the standard error of the mean across all six participants. The asterisks indicate a significance level of $p < 0.05$. 
Figure 8.6: Wearable EOG-based eye tracker integrated into ordinary safety goggles.

Figure 8.7: Example horizontal and vertical EOG signals showing saccades (S), fixations (F), and blinks (B).
Chapter 8: What’s in the Eyes for Context-Awareness?
Bibliography


Chapter 8: What’s in the Eyes for Context-Awareness?


Glossary

Abbreviations

ACC  Accelerometer/Acceleration
ADC  Analog-to-digital converter
AUI  Attentive user interface
BT   Bluetooth
\(cm\)  Centimetre
CMRR Common-mode rejection ratio
CRC  Cyclic redundancy check
CRNT Context Recognition Network Toolbox
CRP  Corneo-retinal potential difference
CWT-BD Continuous Wavelet Transform Blink Detection
CWT-SD Continuous Wavelet Transform Saccade Detection
\(dB\)  Decibel
DEL  Device layer
D-HMM Discrete hidden Markov model
DRL  Driven right leg
DSP  Digital signal processor
EAR  Eye-based activity recognition
EEPROM Electrically erasable programmable read-only memory
ECG  Electrocardiography
EDD  Error division diagram
EEG  Electroencephalography
EMG  Electromyography
EOG  Electrooculography
fNIRs Functional near infrared spectroscopy
FN  False negatives
FP  False positives
FPR False positive rate
G-HMM Gaussian hidden Markov model
HAL  Hardware abstraction layer
HCI  Human-computer interaction
HMM  hidden Markov models
HUD  Head-up display
\(Hz\)  Hertz
LED  Light-emitting diode
Li  Lithium
LIB  Library
MAUI Mobile attentive user interface
MIPS Million instructions per second
Glossary

**mm**  
Millimetre

**mRMR**  
Minimum redundancy maximum relevance

**ms**  
Millisecond

**mAh**  
Milliampere-hour

**mV**  
Millivolt

**mW**  
Milliwatt

**µV**  
Microvolt

**MMC**  
MultiMediaCard

**RFID**  
Radio-frequency identification

**ROC**  
Receiver operating characteristic

**SEL**  
Serious error level

**SNR**  
Signal-to-noise ratio

**STR**  
String matching

**SVM**  
Support vector machine

**TAL**  
Task layer

**TN**  
True negative

**TP**  
True positives

**TPR**  
True positive rate (recall)

**V**  
Volt

**WEPU**  
Wearable EOG processing unit
Acknowledgements

This work would not have been possible without the facilities and the support offered by my advisor Prof. Gerhard Tröster. I would like to thank him for the opportunity he has given me to work in such an interesting research field and the freedom to make my own way in research. I would also like to offer warm thanks to Daniel Roggen who has always been an invaluable source of both support and scientific inspiration. Equally, I would like to thank Jamie A. Ward and Prof. Hans Gellersen for their interest in my work, their support, and the very successful collaboration since the very beginning.

I wish to thank all those other colleagues at ETH who, either through direct assistance or through interesting discussions, have helped me finish this work. In particular, I thank my current and former office mates Clemens Lombriser, Patrick de la Hamette, Mirco Rossi, and Claudia Villalonga, as well as Bernd Tessendorf. I would also like to give credit to all the students I’ve supervised in the context of their semester or master theses for their contributions to my work: Roberto Celli and Christian Meyer, Roman Suter, Marcel Breu, Philip Herter and Martin Wirz, Stefan Dahinden and Stephan Lütolf, Michael Hardegger, and last but not least Dominik Kamm.

This work would also not have been possible without the support of Ruth Zähringer, our institute’s secretary, and Urs Egger who supported my students in developing the wearable EOG goggles.

Most importantly, I would like to thank my mum, my dad, and my sister for all the support and encouragement they have given me over the years. My girlfriend Julia, too, deserves a big thank you for her love, for proofreading my papers, for her understanding when I stayed at ETH until late at night or when I worked at weekends, as well as for cheering me up if MATLAB code or experiments didn’t work out as expected.

I want to dedicate this work to the memory of my father Hans-Jochen Bulling. Back in the 1990s, he was the one who gave me my first computer, a Commodore C64. That present initially got me onto computers and he continued to support me in using and understanding technology. I will do my very best to pass his interest and enthusiasm in technology on to my own children. He was taken from us only half a year after I started my PhD and had never been able to visit me at his former place of work - at least I will now be able to share part of his memories of Zurich for the rest of my life.

Zurich, June 2010

Andreas Bulling
Curriculum Vitae

Personal Information
Andreas Bulling
Date of birth 14 February 1981, Ludwigshafen/Rhein, Germany
Citizen of Germany

Education
2006–2010 Ph.D. studies (Dr. sc. ETH) in Information Technology and Electrical Engineering at ETH Zurich, Switzerland
2001–2006 M.Sc. studies (Dipl.-Inform.) in Computer Science at the Technical University of Karlsruhe, Germany
1991–2000 Abitur, Geschwister-Scholl-Gymnasium, Ludwigshafen/Rhein, Germany

Work Experience
2008 Research internship, Lancaster University, Lancaster, United Kingdom (3 months)
2006–2010 Research staff member, Electronics Laboratory, ETH Zurich, Switzerland (43 months)
2003–2005 Student research assistant, Institute for Process Control and Robotics, Technical University of Karlsruhe, Germany (24 months)
2002–2005 Project manager and software engineer, Krumedia GmbH, Karlsruhe, Germany (31 months)

Scientific Involvement
2011 Pervasive (Publicity Co-Chair)
2010 Pervasive (Proceedings Co-Chair), BodyNets (Publicity Co-Chair)
2009 UbiComp (Publicity Co-Chair)
2008 EuroSSC (Poster, Demonstration, and Proceedings Co-Chair)