# Eye Movement Analysis for Activity Recognition Using Electrooculography

Andreas Bulling, *Student Member, IEEE,* Jamie A. Ward,
Hans Gellersen, and Gerhard Tröster, *Senior Member, IEEE*

**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.

**Index Terms**—Ubiquitous computing, Feature evaluation and selection, Pattern analysis, Signal processing.

✦

---

## 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 activi-

ties 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 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 auxilliary 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.

### 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

- *A. Bulling and G. Tröster are with the Wearable Computing Laboratory, Department of Information Technology and Electrical Engineering, Swiss Federal Institute of Technology (ETH) Zurich, Gloriastrasse 35, 8092 Zurich, Switzerland. E-mail: {bulling, troester}@ife.ee.ethz.ch*
- *J. A. Ward and H. Gellersen are with the Computing Department, Lancaster University, InfoLab 21, South Drive, Lancaster, United Kingdom, LA1 4WA. E-mail: {j.ward, hwg}@comp.lancs.ac.uk*
- *Corresponding author: A. Bulling, bulling@ife.ee.ethz.ch*

1. An earlier version of this paper was published in [10].

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.

## 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 any specific activity (the NULL class). Finally, we discuss the findings gained from this experiment and give an outlook to future work.

## 2 RELATED WORK

### 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.

### 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.

### 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.

## 3 BACKGROUND

### 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).
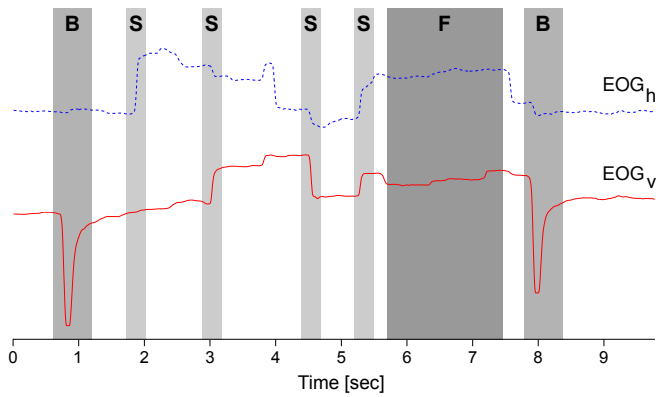
Fig. 1. Denoised and baseline drift removed horizontal ($EOG_h$) and vertical ($EOG_v$) signal components. Examples of the three main eye movement types are marked in grey: saccades (S), fixations (F), and blinks (B).

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].

## 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. 1).

### 3.2.1 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].

### 3.2.2 Fixations

Fixations are the stationary states of the eyes during which gaze is held upon a specific location in the visual
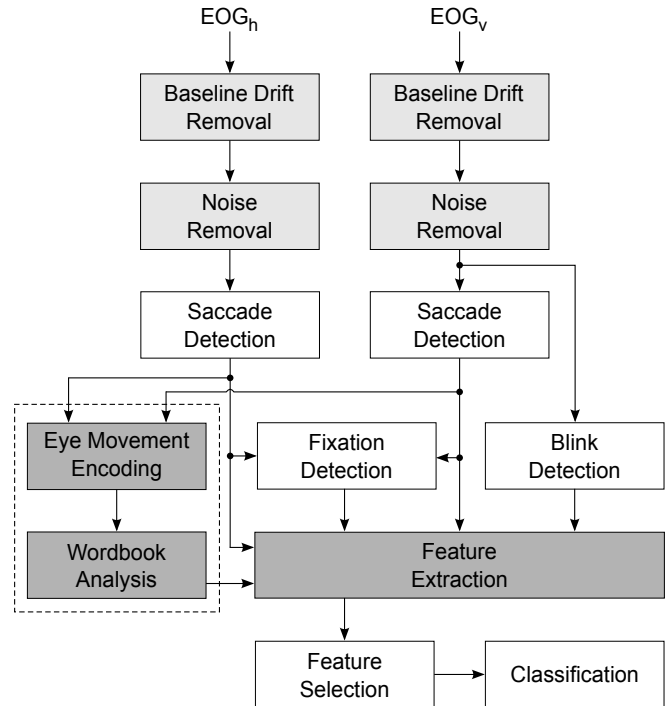


Fig. 2. Architecture for eye-based activity recognition on the example of EOG. Light grey indicates EOG signal processing; dark grey indicates use of a sliding window.

scene. Fixations are usually defined as the time between each two saccades. The average fixation duration lies between 100 ms and 200 ms [26].

### 3.2.3 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].

## 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.

### 4.1 Recognition Architecture

Fig. 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.

## 4.2 EOG Signal Processing

### 4.2.1 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. 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], [32], [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.

### 4.2.2 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 loose 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.

## 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. 2). In the following, we introduce our saccade and blink detection algorithms and characterise their performance on EOG signals recorded under constrained conditions.

### 4.3.1 Saccade and Fixation Detection

For saccade detection, we developed the so-called *Continuous Wavelet Transform - Saccade Detection* (CWT-SD) algorithm (see Fig. 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 $C_b^a$ of $s$ at scale $a$ and position $b$ is defined:

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

By applying an application-specific threshold $th_{sd}$ on the coefficients $C_i(s) = C_i^{20}(s)$, CWT-SD creates a vector $M$ with elements $M_i$:
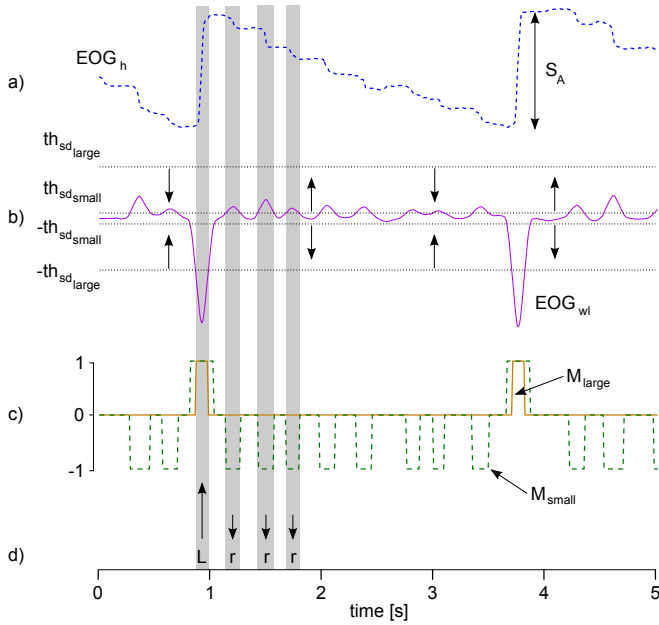
Fig. 3. *Continuous Wavelet Transform - Saccade Detection* (CWT-SD) algorithm. (a) Denoised and baseline drift removed horizontal EOG signal during reading with example saccade amplitude ($S_A$); (b) the transformed wavelet signal ($EOG_{wl}$), with application-specific small ($\pm th_{small}$) and large ($\pm th_{large}$) 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.

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

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. 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 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 threshold $th_{fd}$ of 10,000, or if its duration is below a minimum threshold $th_{fd_t}$ of 200 ms. The value of $th_{fd}$ was derived as part of the CWT-SD evaluation; that of $th_{fd_t}$ 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. 3): no saccade ($-th_{sd_{small}} < C < th_{sd_{small}}$), small saccade ($-th_{sd_{large}} < C < -th_{sd_{small}}$ or $th_{sd_{small}} < C < th_{sd_{large}}$), and large saccade ($C < -th_{sd_{large}}$ 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 ($2 * \frac{precision * recall}{precision + recall}$). 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. 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.
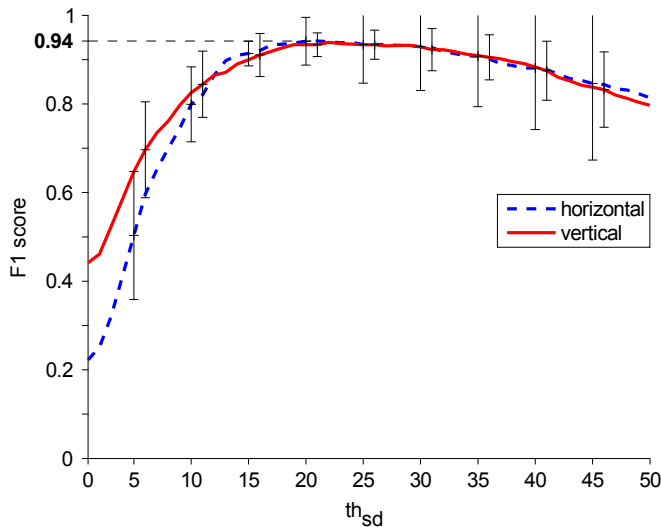
Fig. 4. Evaluation of the CWT-SD algorithm for both EOG signal components using a sweep of its main parameter, the threshold $th_{sd}$. The figure plots the mean F1 score over all five participants; vertical lines show the standard deviation for selected $th_{sd}$. Maximum F1 score is indicated by a dashed line.

### 4.3.2 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 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. 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
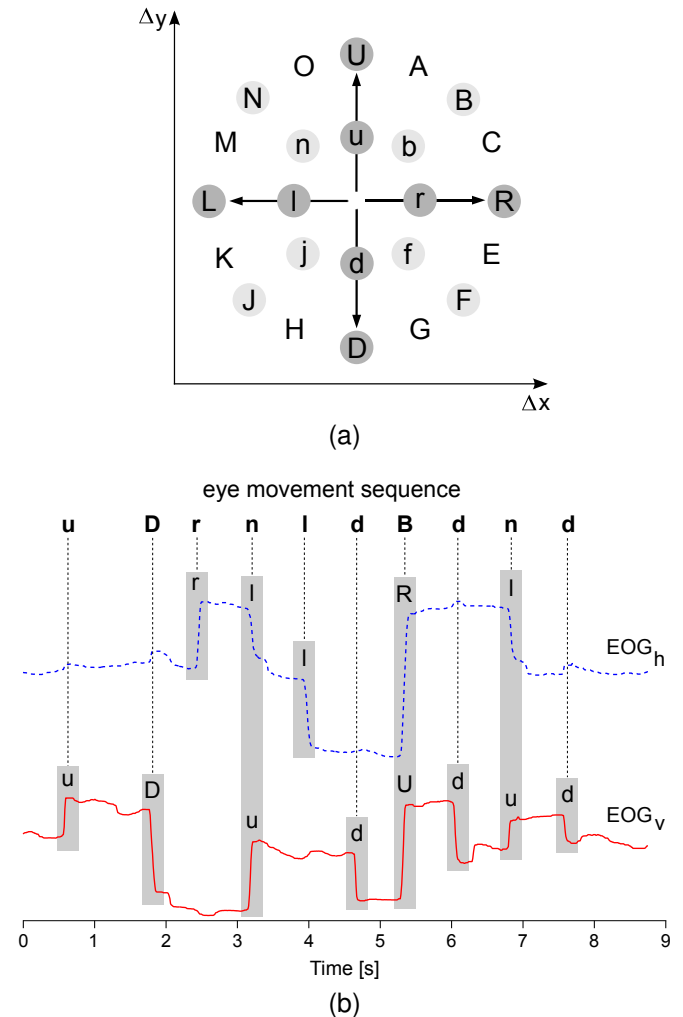


(a)



(b)

Fig. 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").

the average blink duration cited earlier from literature.

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

### 4.4.1 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
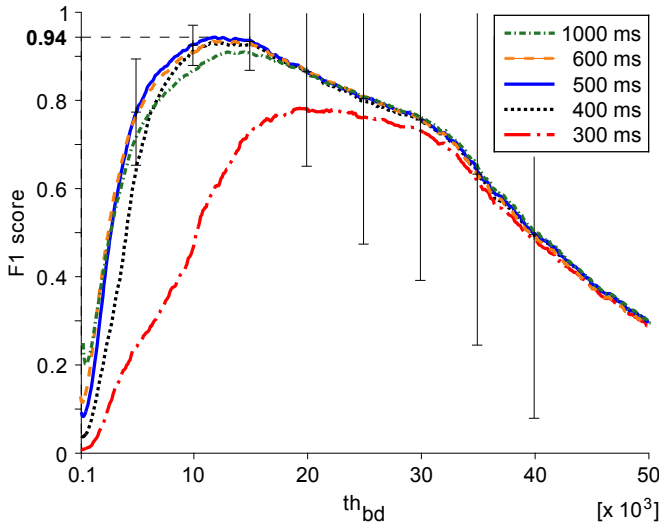
Fig. 6. Evaluation of the CWT-BD algorithm over a sweep of the blink threshold $th_{bd}$, for five different maximum time differences $th_{bd_t}$. The figure plots the mean F1 score over all participants; vertical lines show the standard deviation for selected $th_{bd}$. Maximum F1 score is indicated by a dashed line.



Fig. 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.

characters (see Fig. 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. 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.

### 4.4.2 Wordbook Analysis

Based on the encoded eye movement sequence, we propose a wordbook analysis to assess repetitive eye movement patterns (see Fig. 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
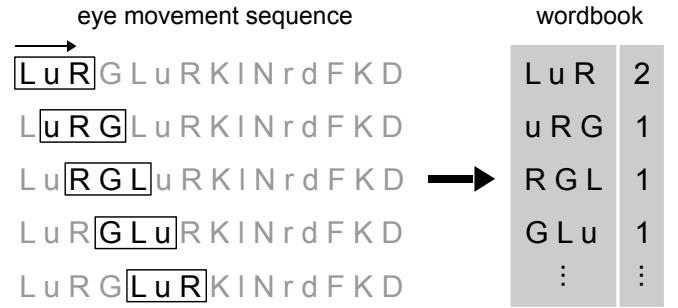
## TABLE 1
Naming scheme for the features used in this work. For a particular feature, e.g. *S-rateSPHor*, the capital letter represents the group - saccadic (S), blink (B), fixation (F) or wordbook(W) - and the combination of abbreviations after the dash describes the particular type of feature and the characteristics it covers.

| Group | Features |
|---|---|
| saccade (S-) | 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 |
| fixation (F-) | 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 |
| blink (B-) | mean (mean) or variance (var) of the blink duration or blink rate (rate) |
| wordbook (W-) | 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) |

$Wb_l$. For a pattern that is already included in $Wb_l$, its occurrence count is increased by one.

### 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 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 $W_{fe}$.

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.

### 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 $S$.

## 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
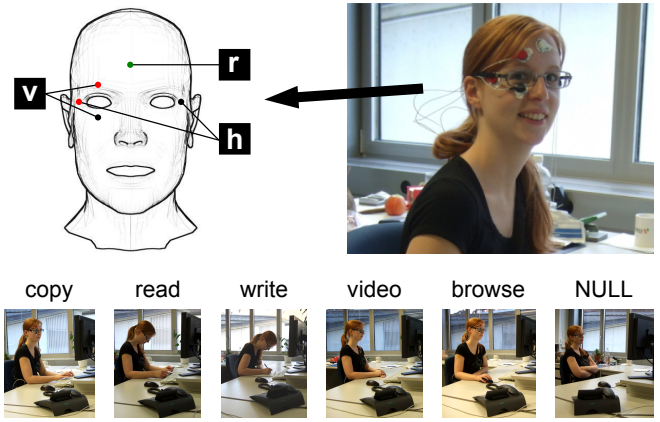


Fig. 8.  (top) Electrode placement for EOG data collection (h: horizontal, v: vertical, r: reference). (bottom) Continuous sequence of five typical office activities: copying a text, reading a printed paper, taking hand-written notes, watching a video, browsing the web, and periods of no specific activity (the NULL class).

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 asked to follow two continuous sequences, each composed of five different, randomly ordered activities, and a period of rest (see bottom of Fig. 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.

### 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. 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.

## 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.

## 5.3 Parameter Selection and Evaluation

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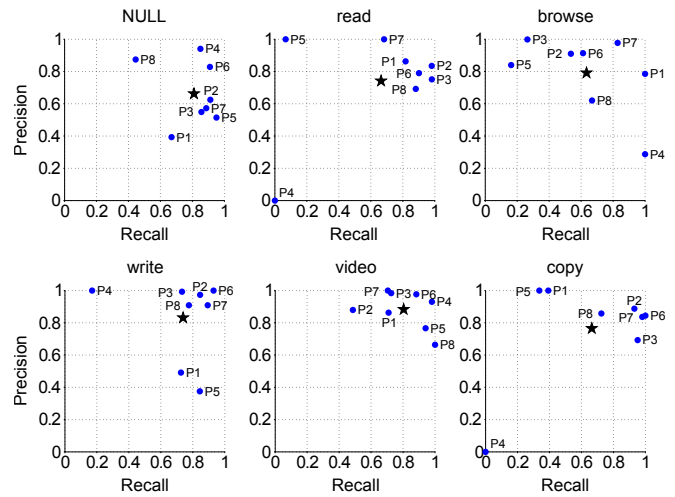


Fig. 9. Precision and recall for each activity and participant. Mean performance (P1 to P8) is marked by a star.



Fig. 10. Summed confusion matrix from all participants, normalised across ground truth rows.

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.

## 6 RESULTS

### 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 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. P4 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.

TABLE 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.

|  | P1 (m) | P2 (m) | P3 (m) | P4 (m) | P5 (m) | P6 (f) | P7 (f) | P8 (m) | Mean |
|---|---|---|---|---|---|---|---|---|---|
| Precision | 76.6 | 88.3 | 83.0 | **46.6** | 59.5 | 89.2 | **93.0** | 72.9 | 76.1 |
| Recall | 69.4 | 77.8 | 72.2 | 47.9 | **46.0** | 86.9 | 81.9 | 81.9 | 70.5 |
| # Features | **81** | 46 | 64 | 59 | 50 | 69 | 21 | **9** | 50 |

Fig. 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. 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.

### 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. 11 shows the top 15 features according to the median rank over all sets (see Table 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 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 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 DISCUSSION

### 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. 4 and 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.

### 7.2 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
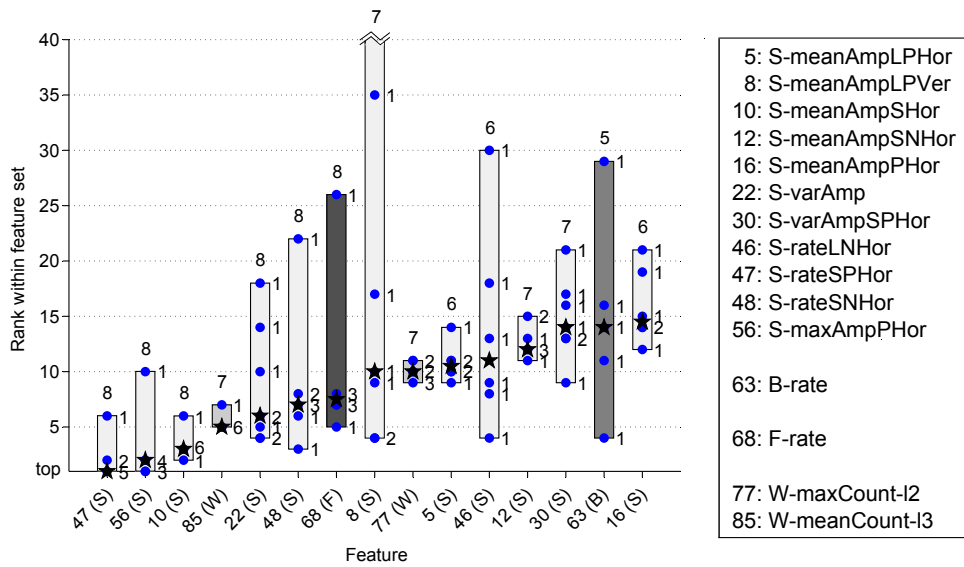
Fig. 11. Top 15 features selected by mRMR for all eight training sets. X-axis shows feature number and group; the key on the right shows the corresponding feature names as described in Table 1; Y-axis shows the rank (top = 1). For each feature, the bars show: the total number of training sets for which 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.

TABLE 3
Top five features selected by mRMR for each activity over all training sets (see Table 1 for details on feature names).

| rank | read | browse | write | video | copy |
|------|------|--------|-------|-------|------|
| 1 | W-maxCount-l2 | S-rateSPHor | W-varCount-l4 | F-meanVarVertAmp | S-varAmp |
| 2 | W-meanCount-l4 | W-varCount-l4 | F-meanVarVertAmp | F-meanVarHorAmp | S-meanAmpSNHor |
| 3 | W-varCount-l2 | W-varCount-l3 | F-varDuration | B-rate | S-meanAmpLPHor |
| 4 | F-varDuration | W-varCount-l2 | F-meanDuration | S-varAmpNHor | S-rateS |
| 5 | B-rate | W-meanCount-l1 | S-rateLPVer | S-meanAmpSPHor | F-meanVarHorAmp |

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. 10.

## 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.

## 7.4 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 wordbook 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.

## 7.5 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.

## 7.6 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 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.

## 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.

## REFERENCES

[1] S. Mitra and T. Acharya, "Gesture recognition: A survey," *IEEE Trans. on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 37, no. 3, pp. 311–324, 2007.

[2] P. Turaga, R. Chellappa, V. S. Subrahmanian, and O. Udrea, "Machine recognition of human activities: A survey," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 18, no. 11, pp. 1473–1488, 2008.

[3] B. Najafi, K. Aminian, A. Paraschiv-Ionescu, F. Loew, C. J. Bula, and P. Robert, "Ambulatory system for human motion analysis using a kinematic sensor: monitoring of daily physical activity in the elderly," *IEEE Trans. on Biomedical Eng.*, vol. 50, no. 6, pp. 711–723, 2003.

[4] J. A. Ward, P. Lukowicz, G. Tröster, and T. E. Starner, "Activity recognition of assembly tasks using body-worn microphones and accelerometers," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1553–1567, 2006.

[5] N. Kern, B. Schiele, and A. Schmidt, "Recognizing context for annotating a live life recording," *Personal and Ubiquitous Computing*, vol. 11, no. 4, pp. 251–263, 2007.

[6] A. Bulling, J. A. Ward, H. Gellersen, and G. Tröster, "Robust Recognition of Reading Activity in Transit Using Wearable Electrooculography," in *Proc. 6th Int'l Conf. on Pervasive Computing (Pervasive 2008)*. Springer, 2008, pp. 19–37.

[7] S. P. Liversedge and J. M. Findlay, "Saccadic eye movements and cognition," *Trends in Cognitive Sciences*, vol. 4, no. 1, pp. 6–14, 2000.

[8] J. M. Henderson, "Human gaze control during real-world scene perception," *Trends in Cognitive Sciences*, vol. 7, no. 11, pp. 498–504, 2003.

[9] A. Bulling, D. Roggen, and G. Tröster, "Wearable EOG goggles: Seamless sensing and context-awareness in everyday environments," *Journal of Ambient Intelligence and Smart Environments*, vol. 1, no. 2, pp. 157–171, 2009.

[10] A. Bulling, J. A. Ward, H. Gellersen, and G. Tröster, "Eye movement analysis for activity recognition," in *Proc. 11th Int'l Conf. on Ubiquitous Computing (UbiComp 2009)*, 2009, pp. 41–50.

[11] Q. Ding, K. Tong, and G. Li, "Development of an EOG (electrooculography) based human-computer interface," in *Proc. 27th Int'l Conf. of the Eng. in Medicine and Biology Soc. (EMBS 2005)*, 2005, pp. 6829–6831.

[12] Y. Chen and W. S. Newman, "A human-robot interface based on electrooculography," in *Proc. 2004 IEEE Int'l Conf. on Robotics and Automation (ICRA 2004)*, vol. 1, 2004, pp. 243–248.

[13] W. S. Wijesoma, K. S. Wee, O. C. Wee, A. P. Balasuriya, K. T. San, and K. K. Soon, "EOG based control of mobile assistive platforms for the severely disabled," in *Proc. 2005 IEEE Int'l Conf. on Robotics and Biomimetics (ROBIO 2005)*, 2005, pp. 490–494.

[14] R. Barea, L. Boquete, M. Mazo, and E. Lopez, "System for assisted mobility using eye movements based on electrooculography," *IEEE Trans. Neural Systems and Rehabilitation Eng.*, vol. 10, no. 4, pp. 209–218, 2002.

[15] M. M. Hayhoe and D. H. Ballard, "Eye movements in natural behavior," *Trends in Cognitive Sciences*, vol. 9, pp. 188–194, 2005.

[16] S. S. Hacisalihzade, L. W. Stark, and J. S. Allen, "Visual perception and sequences of eye movement fixations: a stochastic modeling approach," *IEEE Trans. Systems, Man and Cybernetics*, vol. 22, no. 3, pp. 474–481, 1992.

[17] M. Elhelw, M. Nicolaou, A. Chung, G.-Z. Yang, and M. S. Atkins, "A gaze-based study for investigating the perception of visual realism in simulated scenes," *ACM Trans. Applied Perception*, vol. 5, no. 1, pp. 1–20, 2008.

[18] L. Dempere-Marco, X. Hu, S. L. S. MacDonald, S. M. Ellis, D. M. Hansell, and G.-Z. Yang, "The use of visual search for knowledge gathering in image decision support," *IEEE Trans. Medical Imaging*, vol. 21, no. 7, pp. 741–754, 2002.

[19] D. D. Salvucci and J. R. Anderson, "Automated eye-movement protocol analysis," *Human-Computer Interaction*, vol. 16, no. 1, pp. 39–86, 2001.

[20] D. Abowd, A. Dey, R. Orr, and J. Brotherton, "Context-awareness in wearable and ubiquitous computing," *Virtual Reality*, vol. 3, no. 3, pp. 200–211, 1998.

[21] L. Bao and S. S. Intille, "Activity Recognition from User-Annotated Acceleration Data," in *Proc. 2nd Int'l Conf. on Pervasive Computing (Pervasive 2004)*. Springer, 2004, pp. 1–17.

[22] B. Logan, J. Healey, M. Philipose, E. Tapia, and S. S. Intille, "A Long-Term Evaluation of Sensing Modalities for Activity Recognition," in *Proc. 9th Int'l Conf. on Ubiquitous Computing (UbiComp 2007)*, 2007, pp. 483–500.

[23] F. T. Keat, S. Ranganath, and Y. V. Venkatesh, "Eye gaze based reading detection," in *Proc. 2004 IEEE Conf. on Convergent Technologies for the Asia-Pacific Region*, vol. 2, 2003, pp. 825–828.

[24] M. Brown, M. Marmor, and Vaegan, "ISCEV Standard for Clinical Electro-oculography (EOG)," *Documenta Ophthalmologica*, vol. 113, no. 3, pp. 205–212, 2006.

[25] A. T. Duchowski, *Eye Tracking Methodology: Theory and Practice*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2007.

[26] B. R. Manor and E. Gordon, "Defining the temporal threshold for ocular fixation in free-viewing visuocognitive tasks," *Journal of Neuroscience Methods*, vol. 128, no. 1-2, pp. 85 – 93, 2003.

[27] C. N. Karson, K. F. Berman, E. F. Donnelly, W. B. Mendelson, J. E. Kleinman, and R. J. Wyatt, "Speaking, thinking, and blinking," *Psychiatry Research*, vol. 5, no. 3, pp. 243–246, 1981.

[28] R. Schleicher, N. Galley, S. Briest, and L. Galley, "Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired?" *Ergonomics*, vol. 51, no. 7, pp. 982 – 1010, 2008.

[29] H. R. Schiffman, *Sensation and Perception: An Integrated Approach*, 5th ed. New York: John Wiley and Sons, Inc., 2001.

[30] J. J. Gu, M. Meng, A. Cook, and G. Faulkner, "A study of natural eye movement detection and ocular implant movement control using processed EOG signals," in *Proc. 2001 IEEE Int'l Conf. on Robotics and Automation (ICRA 2001)*, vol. 2, 2001, pp. 1555–1560.

[31] N. Pan, V. M. I, M. P. Un, and P. S. Hang, "Accurate Removal of Baseline Wander in ECG Using Empirical Mode Decomposition," in *Joint Meeting of the 6th Int'l Symp. on Noninvasive Functional Source Imaging of the Brain and Heart and the Int'l Conf. on Functional Biomedical Imaging*, 2007, pp. 177–180.

[32] V. S. Chouhan and S. S. Mehta, "Total removal of baseline drift from ECG signal," in *Proc. 17th Int'l Conf. on Computer Theory and Applications*, 2007, pp. 512–515.

[33] L. Xu, D. Zhang, and K. Wang, "Wavelet-based cascaded adaptive filter for removing baseline drift in pulse waveforms," *IEEE Trans. Biomedical Eng.*, vol. 52, no. 11, pp. 1973–1975, 2005.

[34] M. A. Tinati and B. Mozaffary, "A wavelet packets approach to electrocardiograph baseline drift cancellation," *Int'l Journal of Biomedical Imaging*, vol. Article ID 97157, p. 9 pages, 2006.

[35] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.

[36] D. D. Salvucci and J. H. Goldberg, "Identifying fixations and saccades in eye-tracking protocols," in *Proc. 2000 Symp. Eye Tracking Research & Applications (ETRA 2000)*, 2000, pp. 71–78.

[37] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.

[38] H. Peng, "mRMR Feature Selection Toolbox for MATLAB," February 2008; http://research.janelia.org/peng/proj/mRMR/.

[39] K. Crammer and Y. Singer, "Ultraconservative online algorithms for multiclass problems," *Journal of Machine Learning Research*, vol. 3, pp. 951–991, 2003.

[40] C.-J. Lin, "LIBLINEAR - a library for large linear classification," February 2008; http://www.csie.ntu.edu.tw/~cjlin/liblinear/.

[41] D. Bannach, P. Lukowicz, and O. Amft, "Rapid Prototyping of Activity Recognition Applications," *IEEE Pervasive Computing*, vol. 7, no. 2, pp. 22–31, 2008.

[42] R. L. Canosa, "Real-world vision: Selective perception and task," *ACM Trans. Applied Perception*, vol. 6, no. 2, pp. 1–34, 2009.

[43] D. Palomba, M. Sarlo, A. Angrilli, A. Mini, and L. Stegagno, "Cardiac responses associated with affective processing of unpleasant film stimuli," *Int'l Journal of Psychophysiology*, vol. 36, no. 1, pp. 45 – 57, 2000.

[44] A. Bulling, D. Roggen, and G. Tröster, "It's in Your Eyes - Towards Context-Awareness and Mobile HCI Using Wearable EOG Goggles," in *Proc. of the 10th Int'l Conf. on Ubiquitous Computing (UbiComp 2008)*, 2008, pp. 84–93.

[45] T. Huynh, M. Fritz, and B. Schiele, "Discovery of Activity Patterns using Topic Models," in *Proc. 10th Int'l Conf. on Ubiquitous Computing (UbiComp 2008)*, 2008, pp. 10–19.

**Andreas Bulling** received the MSc degree in computer science from the Technical University of Karlsruhe, Germany, in 2006. In the same year, he joined the Swiss Federal Institute of Technology (ETH) in Zurich, Switzerland, as a doctoral candidate and research assistant. His research interests are in activity recognition and context-awareness, particularly using multi-modal sensing and inference, eye movement analysis, novel methods for eye-based human-computer interaction, and the development of embedded sensors for wearable eye tracking. He is a student member of the IEEE. For more details please visit: http://www.andreas-bulling.eu/

**Jamie A. Ward** received the BEng degree with joint honors in computer science and electronics from the University of Edinburgh, Scotland, in 2000. He spent a year working as an analogue circuit designer in Austria before joining the wearable computing lab at the Swiss Federal Institute of Technology (ETH) in Zurich, Switzerland. In 2006, he received the PhD degree from ETH on the topic of activity recognition (AR) using body-worn sensors. In addition to his work on AR, Jamie is actively involved in developing improved methods for evaluating AR performance. He is currently a Marie Curie research fellow at Lancaster University, United Kingdom.

**Hans Gellersen** received the MSc and PhD degrees in computer science from the Technical University of Karlsruhe, Germany, in 1996. He was a research assistant at the telematics institute (1993 to 1996) and director of the Telecooperation Office (1996 to 2000). Since 2001, he is a professor for interactive systems at Lancaster University, United Kingdom. His research interest is in ubiquitous computing and context-aware systems. He is actively involved in the formation of the ubiquitous computing research community and has initiated the UbiComp conference series. He also serves as an editor of the Personal and Ubiquitous Computing Journal.

**Gerhard Tröster** received the MSc degree from the Technical University of Karlsruhe, Germany, in 1978, and the PhD degree from the Technical University of Darmstadt, Germany, in 1984, both in electrical engineering. During eight years at Telefunken, Germany, he was responsible for various research projects focused on key components of digital telephony. Since 1993, he is a professor for wearable computing at the Swiss Federal Institute of Technology (ETH) in Zurich, Switzerland. His fields of research include wearable computing, smart textiles, electronic packaging, and miniaturised digital signal processing. He is a senior member of the IEEE.