Eyelid Contour Detection and Tracking for Startle Research related Eye-Blink Measurements from High-Speed Video Records

Florian Bernard\textsuperscript{a,\textdagger}, Christian Eric Deuter\textsuperscript{b}, Peter Gemmar\textsuperscript{a}, Hartmut Schachinger\textsuperscript{b}

\textsuperscript{a}Institute for Innovative Informatics Applications, Trier University of Applied Sciences, 54293 Trier, Germany
\textsuperscript{b}Institute of Psychobiology, University of Trier, 54290 Trier, Germany

Abstract

Using the positions of the eyelids is an effective and contact-free way for the measurement of startle induced eye-blinks, which plays an important role in human psychophysiological research. To the best of our knowledge, no methods for an efficient detection and tracking of the exact eyelid contours in image sequences captured at high-speed exist that are conveniently usable by psychophysiological researchers.

In this publication a semi-automatic model-based eyelid contour detection and tracking algorithm for the analysis of high-speed video recordings from an eye tracker is presented. As a large number of images have been acquired prior to method development it was important that our technique is able to deal with images that are recorded without any special parametrisation of the eye tracker. The method entails pupil detection, specular reflection removal and makes use of dynamic model adaption.

In a proof-of-concept study we could achieve a correct detection rate of 90.6\%. With this approach, we provide a feasible method to accurately assess eye-blinks from high-speed video recordings.

Keywords: Image processing, startle eye-blink, segmentation, eye-blink detection, eyelid detection, eyelid tracking

1. Introduction

Emotions are considered to be a readiness for action and a strong motivational force: emotional states activate the organism for a certain behaviour [1]. Furthermore, emotion is closely connected to other cognitive processes, to which environmental stimuli we attend to and how we process them [2]. For this reason, the study of emotion is of principal interest to the psychological science. However, the measurement of emotion is associated with certain methodological limitations [3]. While self-report remains useful for many experimental fields, it also entails some profound disadvantages. In many situations people are not totally aware of their affective states or unable to verbalise them. Furthermore, self-report might be biased by social desirability. These constraints led to the development of research tools that measure behavioural and physiological responses to emotional stimuli. Emotions can be inferred from vocal changes, facial expressions and body postures [3]. Reaction times and error rates in tasks that require fast behavioural responses to emotional stimuli allow conclusions about the individual’s affective states and evaluations [4]. In addition, affective states can also be measured via physiological markers—an object of research for a sub-discipline named psychophysiology. A typical experimental setting usually requires the participant to view or listen to emotional information while changes in physiological parameters are recorded. Such a psychophysiological method is the startle eye-blink response, which is modulated by the perceptual processing of affective information [5].

\textdagger Corresponding author.

Email addresses: f.bernard@hochschule-trier.de (Florian Bernard), deuter@uni-trier.de (Christian Eric Deuter), p.gemmar@hochschule-trier.de (Peter Gemmar), schachinger@uni-trier.de (Hartmut Schachinger)
2. Background

In this section the background of startle modulation and the eye-blink response as experimental standard measure will be described. Then, advantages and limitations of existing methods for eye-blink measurement will be given, followed by a description of the technical challenges of our proposed method.

2.1. Affective Startle Modulation

The startle response is a fast defensive reflex that protects the organism against potential injury. Affective context stimulation modulates the response: Negative stimuli potentiate, positive stimuli activate the organism’s defensive system while positive stimuli activate the appetitive, approach-related system [5]. Since the startle response itself is defensive, the pre-activation of a negative state potentiates the response. In an appetitive, pleasant mode, the elicitation of a defensive reaction will be inhibited—thereby attenuating the startle response. Affective startle modulation has been demonstrated in different paradigms using various stimuli in different modalities [6]. The human startle response follows a distinctive pattern of muscle contraction. Many different skeletal muscle systems are involved in the response. However, because of its reliable elicitation and short onset latencies, the startle induced eye-blink response evolved as an experimental standard measure for the general startle response [7, 8].

2.2. Eye-Blink Measurement

The most commonly employed and, up to recently, only available method available measure for startle eye-blink is the electrophysiological recording of the orbicularis oculi muscle via electromyography (EMG) [9]. Even though EMG is a reliable measure of the eye-blink, it is also limited in its range of applications. In experiments that make use of functional magnetic imaging, parallel EMG recording can be unfeasible. Also, the eye-blink response per se might be impaired by the measurement. The EMG recording is not contact-free, it requires the attachment of electrodes directly underneath the eye. These can hamper the lid and can be considered as tactile stimulus in itself. The eye-blink startle response serves to protect the eye against physical impact, therefore a physical stimulus in the proximity of the eye might have an influence on startle induced blinks.

Furthermore, the only site for EMG recording around the eye is the orbicularis oculi. While a blink is basically accomplished by rapid activation of the musculus orbicularis oculi, other antagonising muscles control the lid as well. The musculus levator palpebrae and the smooth Müller’s muscle, that runs from the musculus levator palpebrae to the upper margin of the tarsal plate, elevate the eyelid and regulate the palpebral aperture [10, 11]. Since Müller’s muscle is innervated by the sympathetic branch of the autonomous nervous system [12], factors such as stress could influence the upper eyelid’s muscle tonus and thereby also the movement during the blink. While the actual blink kinematics are underdetermined by orbicularis oculi activation alone, this muscle is the only possible site for EMG recordings. Hence, using EMG recordings merely the eye-blink magnitude—a relative measure describing the strength of an eye-blink—becomes accessible. With that, the exact distance between the eyelids remain inaccessible with the EMG-based measurement of the orbicularis oculi only, fortifying the necessity of other approaches.

Simultaneously to EMG recordings, in [13], the movement of the upper eyelid is measured using a magnetic search coil technique, with a fine wire coil being attached to the upper eyelid. Using this multimodal approach exact eye-blink kinematics become available. However, the attachment of a coil is invasive and uncomfortable for the subjects [14].

Besides non contact-free indirect measures of muscle activation or approaches using magnetic search coils, two lines of research have attempted to directly measure the blink via optical means, employing either infrared reflections or high-speed video recordings. With the first approach, infrared light is transmitted to the eye and the reflections are recorded by photo-sensitive sensors. This technique makes use of the fact that different surface structures, in that case the eye ball and the eyelid, have different reflectance rates to determine eyelid closure [15, 16]. The video-based technique makes use of a high-speed video camera that continuously records the participants eye [17].
In [18, 19], a region placed on the upper eyelid and the corner of the eye (as a reference point) are marked by a human observer and tracked via a motion tracking software. In these cases the exact kinematics of a blink remain not fully accessible because only two reference points are tracked. In [20], eye-blink kinematics have been analysed manually after transcranial magnetic stimulation (TMS) with four points (upper eyelid, lower eyelid, upper pupil border and lower pupil border), where the resulting kinematics are more detailed but yet not exhaustive because merely points, rather than the entire eyelids, are used.

While the mentioned techniques broaden the scope for general human startle research and provide interesting alternatives to EMG, the automatic assessment of the acquired images has still not been satisfactorily addressed.

Even though eye-blink magnitude, a standard measure in many experimental paradigms, can be reliably assessed, the exact palpebral aperture and upper to lower eyelid distance can not be readily measured.

The direct assessment of the eye-blink can be a valuable enhancement of the available methodology. We provide an alternative to the common EMG-based approach that carries several advantages. Since the video-based assessment is contact-free, we spare the electrodes beneath the eyelid that are oftentimes described as unpleasant to the participant. The direct assessment of the blink carries information about muscle systems other than the orbicularis oculi, which might reflect modulation by stress and autonomic arousal. Eventually, this method allows general startle assessment based on the eye-blink startle response in research situations where EMG measurement is interference-prone or even impossible such as fMRI settings and for laboratories that do not have the relevant EMG hardware available.

As a very high number of trials needs to be evaluated it is desirable to automatise the process of eye-blink measurement, leading to our main objective which is the improvement of the feasibility and practicability of the eyelid distance measurement during eye-blinks. We developed a method to detect the eyelid contours and thus measure startle induced eye-blinks by using a model-based automatic detection of the eye contours in sets of still images (referred to as trial), which were recorded by an eye tracking system. Using this approach the entire detection and tracking of the eyelids runs automatically without manual intervention. However, prior to the processing a model, describing the shape of the eyelid, needs to be created by the user. This offline analysis of the images enables the experimenter a high degree of control over the process while still offering convenient usability. The method was confirmed by manual assessment of the detected eyelid contours which revealed an accurate contour detection.

2.3. Technical Challenge

Our main purpose is the automatic eye-blink measurement in a high number of trials. The distance between the upper and the lower eyelid will be used as measurement, therefore we have chosen to detect the upper and the lower eyelid first and then determine the distance between them.

Biologically, the appearance of the eye differs amongst different persons in several aspects, such as eye shape, eyelash texture, eye colour or the width of the eyelid contour. Furthermore, the eyelid shape changes drastically during eye-blinks, which complicates the use of common eyelid contour detection approaches, such as model-based, appearance-based and feature-based methods [21]. Therefore the main challenge is to deal with the dynamic nature of the upper and lower eyelid shape due to eye-blinks. For entirely closed eyes, the upper eyelid contour coincides with the lower eyelid contour. The correct detection of both eyelids during re-opening is a further difficulty in these cases.

3. Related Work

Eyelid detection and tracking for the recognition of the eye’s position is a widely discussed research topic. Some of the possible applications are iris recognition techniques used in biometrics [22, 23, 24, 25], human-computer interaction [26, 27, 28], e.g. used in consumer electronics [21], or facial expression recognition [29, 30, 28].

However, the exact determination of the eyelid contour in order to determine the exact eyelid distance during eye-blinks, which is of main interest for us, is mostly neglected. On the one hand, the eyelids are harder to detect during eye-blinks as their shape changes significantly compared to an open eye. On the
other hand eye-blinks are one of the fastest human reflexes [21]. In our acquisitions we have observed a fully opened eye to close within approximately 20–40ms. Assuming a non high-speed camera capturing images at 50Hz—i.e. one frame is recorded every 20ms—the entire closing process of the eye would only be visible in a maximum of three frames. In the worst case, a fully opened eye will be entirely closed in the next image frame, without any intermediate states being captured. Due to this reason it is very difficult to track eye-blinks precisely in non high-speed videos because the difference of the eyelid position is large between two image frames. In this section a brief overview of research in the area of eye-blink detection and tracking and related areas that are considered to be the most relevant for us are given.

In [31], the detection of the occurrence and the duration of eye-blinks is regarded as classification problem, tackled by template matching using an open-eye template and a closed-eye template, where an exact measurement of the eye-blink magnitude is simply not possible. A two stage procedure, divided into the detection of the eyelids using a minimal path algorithm followed by eyelid tracking using active contours is given in [28]. In this approach the extraction and the tracking of the eyelids due to gaze change in general face recognition applications is the main focus. The tracking of the eyelid contour is in particular problematical during eye-blinks because the eye contours are lost frequently after the eye is closed, which was a general observation in our experiments with active contours approaches. Additionally, with active contours, we have observed multiple cases where the iris contour was detected rather than the eyelid contour. To overcome the problem of lost contours a prediction of the eyelid patterns for reopening eyes are used in [26].

The approach pursued in [21] is the determination of model parameters that are able to characterise the degree of an eye-blink by using Active Appearance Models [32]. In order to enable robust eye tracking an extensive set of points for both eyes is used for training the shape model, including points on the eyelid contour, the iris, the eye socket as well as the eyebrows. As in our case only one image of a single eye is available and both the eye socket and the eyebrows are mostly not visible due to a smaller field of view, this method can not be applied in our case with similar robustness.

Eyelid tracking for eye-blink measurements is conducted in [18] by using a commercial motion tracking software that requires the manual selection of a point on the upper eyelid and the corner of the eye. On the one hand the manual selection of these points in each trial requires time-consuming user intervention and on the other hand only relative eye-blink measurements are performed, rather than the determination of the exact distance between the upper and the lower eyelid.

In [14], the eyelid position and eye retraction are measured by the automatic processing of images of the anterior chamber, illuminated by a slit lamp and recorded at high-speed. The processing comprises various filtering techniques in order to minimise irrelevant features and noise, the subsequent arrangement of the image sequence in a three-dimensional matrix and 3D re-slicing. The results are used in [17] in order to deduce a general mathematical model for eye-blink kinematics. This approach constitutes another method for eyelid distance measurement during eye-blinks. The main difference to our approach is that instead of a high-speed camera a slit lamp is required, which may not be available in all laboratories. The images used for eyelid distance measurement are different to our images as they exhibit a better contrast and thus their method is not applicable in our case.

Several methods exist that make use of the iris for eyelid contour detection. One of them is proposed in [33], where an energy function, that is defined for a deformable template incorporating the eyelid and the iris, is minimised. The iris is also used for the detection of the lower eyelid in [33] by fitting a polynomial through of the eye corner points and the lowest point of the iris circle, which is taken as approximation of the lower eyelid position. In [23] a linear Hough transform is used for eyelid contour segmentation by the detection of multiple line segments lying within the iris area, which then form a part of the eyelid contour. Nevertheless, relying on the iris as part of the eyelid detection is problematical for the detection and tracking of the eyelid contours during eye-blinks because the iris disappears during eye-blinks.

Deformable templates have proved to be an efficient model for the eyelid contour [26]. We have chosen to use a simple model describing the curvature of the eyelids as proposed by [22]. In addition, to cope with the change of the eyelid shape during eye-blinks a dynamic model adaption is performed for eyelid contour tracking.
4. Methods

This section describes the technical aspects of the developed algorithm for eye-blink measurement including eyelid detection and tracking. At first the existing input data is described and the definition of a valid trial is given. Then the principal idea of the entire detection and tracking system is given followed by an in-depth explanation.

4.1. Data

Our eyelid detection and tracking method has been developed in a data-driven manner by the use of trials that have already been acquired. The description of these trials and the resulting general requirements for the acquisition of new trials are given in this section. A typical trial of our studies is a high-speed video record with a length of 1s having a temporal resolution of 500 images/s, giving a total of 500 images per trial.

The head of a person in a trial must not move, thereby giving approximately a constant position of the eye. In order to ensure that the head does not move, a chin and a forehead rest are used in order to stabilise the subject’s head. Even though the pupil movement and pupil size does not remain constant during a blink, this change of the position and the size of the pupil is negligible because the head of a subject is stabilised, the subject is instructed to look at a fixed position and additionally the illumination is constant.

A trial is only regarded as valid if the eye is open at the beginning and the end of a trial. As the main application of our algorithm is to detect the strength of startle-induced eye-blinks a trial usually embodies a full or partial close of the eye. In addition, a trial may only contain a single eye-blink, therefore a trial must be split up into several trials if more than one eye-blink is present.

Each frame of a trial is a 8-bit grey-scale image with a height of 160 pixels and a width of 224 pixels. The trials are recorded by an iView X Hi-Speed 500 eye tracking system manufactured by SensoMotoric Instruments, where a stabilised light source that mainly emits light in the infrared range and a camera which is sensitive to this frequency range are integrated into the eye tracker. Apart from initial adjustments of the brightness and the contrast, the images are used as they are delivered by the eye tracking system. Several images taken from a trial are shown in Fig. 1.

4.2. Algorithm Summary

For the measurement of the eyelid distance an eyelid detection method combined with eyelid tracking is deployed.

Our eyelid detection algorithm requires a model that describes the shape of the upper and the lower eyelid for each person as in [22], given as offset for each image column (Fig. 2 and Fig. 3(b)).

For simplification of the internal processing each trial is split into two sequences. For that, the sub-sequence that depicts the closing of the eye is mapped to an opening eye be reversing a trial from the end to the position where the eye is closed the most.

For the detection and tracking of the eyelid during a closing eye, the following is repeated for each image (Fig. 3(a)) of a sequence: The image columns are shifted according to the upper eyelid curvature model (Fig. 3(b)) as done in [22], leading to an image where the upper eyelid resembles a straight line (Fig. 3(c)). Then the vertical image gradient is detected (Fig. 3(d)), binarised and a line through the edge points is determined using a Hough transform (Fig. 3(e)) [34, 35]. After discarding edge points that are not in the proximity of this line, the edge image is shifted in the inverse direction according to the curvature model. Then a polynomial is fitted through the eyelid edge pixels using a least-squares approach, which is the eyelid approximation (Fig. 3(f)).

After each iteration the upper eyelid curvature model is adapted in order to handle the changing shape of the eyelid while the eye is closing. A summary of the algorithm is illustrated in Fig. 4.

Finally, the eyelid distance in an image can easily be calculated using the upper eyelid polynomial and the lower eyelid polynomial.
Figure 1: Samples taken from a trial. (a)-(f) Closing eye. (g)-(l) Opening eye.

Figure 2: Scheme of curvature model for upper eyelid. The horizontal axis (c) shows the according image column and the vertical axis (r) shows the displacement for each column c.

4.3. Notation

The coordinate system of an image has its origin in the top left corner, with the vertical axis extending from the top row to the bottom row of the image and the horizontal axis extending from the leftmost to the rightmost column of the image. The pixel with the vertical position \( r \) and the horizontal position \( c \) is indexed by \( I(r, c) \).

Parameter values of the algorithm that have been used for the evaluation of our data will be given when a parameter is introduced.
4.4. Model Creation

The model that is required for the eyelid detection is created by the user before eyelid detection. In this section the model creation is described in detail.

As the model is used to describe the curvature of the upper eyelid and the lower eyelid for each person it is necessary to create both an upper eyelid curvature model and a lower eyelid curvature model. As the image intensities in the eyelid region vary heavily among different subjects (e.g. due to different eyelash appearance and density) the curvature model provides a further cue required for a proper eyelid identification. Due to this high variance of the eyelid appearance the identification of the eyelids without these models is error-prone and thus the manual model creation has been chosen in this work.

Rather than manually segmenting the entire eyelid contour, merely a set of at least three points—denoting pixel coordinates—that lie on the eyelid contour are used for curvature model creation for each eyelid in order to keep the required user intervention to a minimum. Then a polynomial regression is applied that finds the best fitting second degree polynomial function \( q(c) \) through the selected contour pixels. This polynomial is used for the shape description of the eyelid contour. Our experiments have shown, in agreement with [36], that polynomials of second degree are sufficient to express the general shape of the upper eyelid as well as the lower eyelid.

For further processing the eyelid position values, given by \( q(c) \), for all \( 1 \leq c \leq C \), with \( C \) being the image width, are calculated, rounded (the values denote pixel coordinates) and stored in the vector \( cm_{abs} \).

In order to generalise the curvature model, i.e. storing the relative position offsets only, the curvature model for the upper eyelid \( cm_{ue}(c) \) and the curvature model for the lower eyelid \( cm_{le}(c) \) are given by

\[
cm_{\{ue,le\}}(c) = cm_{abs}(c) - \min_{1 \leq c \leq C} (cm_{abs}(c)),
\]

respectively.

Finally, all values of \( cm_{\{ue,le\}}(c) \) where \( c \) is smaller than the leftmost or greater than the rightmost selected pixel are set to 0. This gives the structure of a curvature model for the upper eyelid as shown in Fig. 2. This curvature model clipping is done because the leftmost and rightmost parts of the eyelids have a shape that is considerably different from second degree polynomials. Also, these sections of the eyelids

![Figure 3: (a) Eye image. (b) Eye image with upper eyelid curvature model. (c) Eye image shifted by upper eyelid curvature model. (d) Gradient magnitude of upper and lower eyelid (shifted individually by their corresponding curvature models). (e) Binary edges of upper and lower eyelid (extracted from (d)) with best fitting lines. (f) Best fitting polynomial through upper eyelid.](image-url)
have a large variance amongst the trials and due to the resulting error-proneness these parts are not used for further processing. Furthermore, the leftmost and rightmost part of an image do not always show any parts of the eye. With that, the leftmost and the rightmost points should be chosen in such a way that the section of the eyelid between those two points can accurately approximated with a second degree polynomial. In order to guarantee a well-conditioned polynomial fitting the third point should have approximately the same distance from both border points.

Usually it is sufficient to create a single curvature model for a person. As the number of persons is small compared to the total number of trials it is feasible to create one model for each person, requiring three mouse clicks for the upper and lower eyelid each. However, in our implementation an option to create individual curvature models for particular trials is provided additionally, as it has proved useful for trials where the head is heavily rotated or moved to the left or right.

4.5. Eyelid Segmentation

4.5.1. Preprocessing

To overcome the problem of not correctly tracking an eye that is reopening, as described in [26], the reopening of an eye is reduced to a closing eye by partial reversion of the trial, done during a preprocessing stage for each trial. Let $T$ be a trial comprising a sequence of images $[I_1, I_2, \ldots, I_N]$ with $N$ denoting the total number of images in the trial $T$. Each image $I_i$, $1 \leq i \leq N$, has a width of $C$ and a height of $R$.

The eye-closed-index $i_{ec}$ of the image in the trial $T$, where the eye is entirely closed (for trials with full eye closes) or closed the most (for trials with partial eye closes), is determined as described below. Then the trial $T$ can be split into two sequences $S_f$ and $S_b$. The forward sequence $S_f$ is given by $[I_1, I_2, \ldots, I_{i_{ec}}]$ and the backward sequence $S_b$ is given by $[I_N, I_{N-1}, \ldots, I + i_{ec} + 1]$. Both $S_f$ and $S_b$ portray a sequence of a closing eye where the tracking of the eyelid can be handled using the same method.

The size of the pupil that is visible in an image provides sufficient hints for the determination of the eye-close-index $i_{ec}$.

First, the position of the pupil (constant in a trial) in the first image $I_1$ of the trial is identified, denoted by its image column index $c_p$. The column index of the pupil $c_p$ is defined to be the image column where the sum of each column of image $I_1$ has the minimum grey-scale value, i.e. corresponds to the most dark pixels:

![Eyelid detection algorithm](image)

Figure 4: Eyelid detection algorithm.
\[ c_p = \arg\min_{1 \leq c \leq C} \sum_{r=1}^{R} I_1(r, c). \]  

Then, for all images \( I_i \) of the trial the average grey-scale value \( v_i \) of image columns in the proximity of column \( c_p \) is determined:

\[ v_i = \frac{1}{5R} \sum_{\Delta c \in \{0, \pm 2, \pm 4\}} \sum_{r=1}^{R} I_i(r, c_p + \Delta c). \]  

Fig. 5 shows the graph of \( v_i \) for a trial. Eventually, the eye-close-index is defined as

\[ i_{ec} = \arg\max_{1 \leq i \leq N} v_i. \]  

As this estimate of the eye close state is merely used for splitting a trial into two sequences that portray a closing eye, this rough approximation is sufficient.

The subsequent processing deals with image sequences \( S \in \{S_f, S_b\} \) rather than with full trials. Each sequence \( S \) has a length of \( N_S \) and comprises the images \( I_i, 1 \leq i \leq N_S \).

The detection and tracking of the upper eyelid is performed separately from the lower eyelid. However, as the general problem of the upper and lower eyelid detection is equivalent, the following sections describe this task merely as eyelid detection with the differences between the upper and the lower eyelid being pointed out.

### 4.5.2. Curvature Model Application

For the reduction of the computational complexity, thereby enabling the detection of eyelids in dozens of trials within reasonable time using a standard desktop computer, the curved eyelids are transformed to a line-like shape. This is accomplished by shifting each column \( c \) of image \( I_i \) by the offset denoted by the curvature model \( cm_{ue}(c) \) for the upper eyelid (or \( cm_{le}(c) \) for the lower eyelid, analogously), leading to a shifted image \( I_i^{\text{shifted}} \). The result of the curvature model application for the upper eyelid can be seen in Fig. 3(c), with the upper eyelid resembling a straight line.

The procedure described in the sequel is applied for all images \( I_i \) of \( S \). However, after each iteration the curvature models \( cm_{\{ue,le\}} \) are adapted dynamically, leading to a curvature model \( cm_{\{ue,le\}}^i \) for image \( I_i \), as described in 4.5.9.

### 4.5.3. Edge Detection

The edge detection is performed within a region of interest (ROI) that is determined according to the last known eyelid position. Then the edges that are not part of the eyelid are removed and eventually the edge gradient image is binarised.

The ROI of the images \( I_i^{\text{shifted}}, 1 \leq i \leq N_S \), is defined to be a rectangular area with constant size that is
moved along the vertical direction according to the eyelid position of the previous image. The rectangular area is enclosed by the image points with coordinates

\[
\min_{1 \leq c \leq C} (e_{abs}^{i-1}(c) - \Delta v, \Delta h)
\]

and

\[
\min_{1 \leq c \leq C} (e_{abs}^{i-1}(c) + \Delta v, C - \Delta h).
\]

\(\Delta h\) is a constant offset for the horizontal distance from the left and the right image edge (set to 35 pixels) and \(\Delta v\) is an offset for the maximum deviation of the absolute vertical eyelid position in the preceding image \(I_{i-1}\), given by the polynomial \(e_{abs}^{i-1}(c)\), with \(1 \leq c \leq C\). Coordinate values that are not part of the image are set to the corresponding minimum (maximum) within the image.

For the processing of the first image \(I_1^{shifted}\) the offset \(\Delta v\) is set to 30 pixels for both upper and lower eyelid detection and \(e_{abs}^{i-1}\) is defined to be \(cm_{abs}\), which is the absolute eyelid position that was set during the curvature model creation stage (see 4.4). Fig. 6(a) portrays the ROI within the first image \(I_1^{shifted}\) of a sequence \(S\) and Fig. 6(b) shows the extracted ROI only. The restriction of the further processing to the ROI of the first image \(I_1^{shifted}\) is performed in order to avoid that upper eyelid edges are detected as lower eyelid edges and vice versa.

For subsequent images \(I_i^{shifted}\), \(2 \leq i \leq N_S\), the vertical extent of the ROI is smaller. The vertical offset \(\Delta v\) is set to 12 pixels for upper eyelid detection and 8 pixels for lower eyelid detection, which can be interpreted as a liberal estimation of the vertical displacement of the eyelid from an image \(I_i\) to the consecutive image \(I_{i+1}\). The vertical offset is symmetric (i.e. same value in both upwards and downwards direction) in order to account for a potentially mis-detected eyelid contour in a single image frame. The main purpose of the ROI definition for the images \(I_2^{shifted}\) to \(I_{N_S}^{shifted}\) is to reduce the number of pixels that are processed in order to decrease the number of computational steps.

Before actual edge detection, noise due to eyelashes is reduced using a horizontal rank filter with a filter size of 1 by 7 as done in [22] (Fig. 6(c)). One must point out that this filter only has the desired effect if the area with eyelashes is less dense than the non-eyelash area. Otherwise it has the opposite effect of fortifying the eyelashes.

The actual edge detection is performed by applying the Frei-Chen edge detector [35, 37] on image \(I_i^{gradient}\), resulting in a gradient image \(I_i^{gradient}\) (Fig. 6(d)).

### 4.5.4. Pupil and Specular Reflex Removal

In cases where the eyelid does not have a common edge with the pupil, i.e. the pupil is not covered by the eyelids (as in Fig. 6), the pupil edge is a strong edge that must be removed in the gradient image \(I_i^{gradient}\) in order to avoid detecting the pupil edge as eyelid edge. In cases where the eyelid partially covers the pupil, the edge between the eyelid and the pupil is usually a strong edge and merely the removal of the specular reflection in the gradient image \(I_i^{gradient}\) is performed.

For pupil edge removal in the gradient image \(I_i^{gradient}\) the entire pupil of the first image \(I_i\) in \(S\) is detected, resulting in a binary image \(I_i^{pupil}\). As a sequence always begins with an open eye it can be guaranteed that the pupil is visible in the first image. Furthermore, given that (i) the head does not move, (ii) the subject looks constantly at a fixed position and (iii) the illumination remains constant, the change of the position of the pupil in a sequence is marginal. Thus, the pupil that is detected in the first image is also considered as the pupil in all subsequent images. This approach is also backed up by [20], where pilot experiments have revealed that the difference of the pupil borders before TMS and after TMS is negligible.

The pupil image \(I_i^{pupil}\) is then shifted according to the curvature model \(cm_{(ue,le)}^i\), resulting in image \(I_i^{pupil,shifted}\) (see Fig. 6(e)). Finally, all image points of \(I_i^{pupil,shifted}\) are removed from \(I_i^{gradient}\) (Fig. 6(f)).
For the determination of $I_{1}^{\text{pupil}}$ the approximate centre coordinates $(r_p, c_p)$ of the pupil are determined first. $c_p$ is given by (2) and $r_p$ is determined analogously by finding the image row where the sum of each row of image $I_1$ has the minimum grey-scale value, i.e. corresponds to the most dark pixels:

$$r_p = \arg \min_{1 \leq r \leq R} \sum_{c=1}^{C} I_1(r, c).$$

(5)

These coordinates are used for the extraction of the region of interest, which is defined by the rectangular area that is enclosed by the image points $(r_p - \Delta o_p, c_p - \Delta o_p)$ and $(r_p + \Delta o_p, c_p + \Delta o_p)$. In our case $\Delta o_p$ is set to 40 which is a liberal estimate of the maximum radius of the pupil measured in pixel. Fig. 7(a) portrays the ROI within image $I_1$ and Fig. 7(b) shows the extracted ROI only. For noise reduction a median filter with filter size 5 by 5 is applied (Fig. 7(c)). Subsequently, holes are filled using an algorithm based on morphological reconstruction (Fig. 7(d)) [38]. Then the result is binarised by thresholding (Fig. 7(e)). As the dark parts of the image are formed by the pupil a threshold value of 40 has proved to be successful. In order to cope with a pupil that increases over time, a morphological dilation with a discoidal structure element of radius 6 is applied on the binarised image, leading to a larger pupil (Fig. 7(f)).

For the specular reflection removal a specular reflection map $I_1^{\text{SR}}$ is created for each image. For that,
all image pixels of image \( I_i^{\text{shifted}} \) that lie within the area of the pupil, as given by image \( I_i^{\text{pupil,shifted}} \), are extracted. The resulting image is binarised using a constant threshold of 200, only keeping bright pixel values. Then a morphological dilation with a square structure element of edge length 4 is applied on this binary image, resulting in the binary specular reflection map \( I_i^{\text{SR}} \).

4.5.5 Binarisation of Gradient Image

After the removal of the entire pupil or the specular reflection from the gradient image \( I_i^{\text{gradient}} \), the gradient image is binarised. However, in order to increase the robustness further edge points of the gradient images \( I_i^{\text{gradient}}, 2 \leq i \leq N_S \), are removed.

For that, it is made use of the eyelid contour of the preceding image \( I_{i-1} \) that was detected in the previous iteration (Fig. 8(a)). The eyelid contour of the previous image is shifted according to the curvature model \( cm_i^{\text{loc},le} \) (Fig. 8(b)). Then the eyelid contour in the shifted image is thickened \( \Delta u \) pixels to the top and \( \Delta d \) pixels to the bottom. For upper eyelid detection \( \Delta u \) is set to 3 and \( \Delta d \) is set to 8. For lower eyelid detection \( \Delta u \) is set to 6 and \( \Delta d \) is set to 3. These values can be interpreted as offsets for the maximum vertical displacement of the respective eyelids from the eyelid position in the previous frame, additionally accounting for minor inaccuracies of the determined eyelid position in the previous frame. The result of the eyelid contour enhancement is depicted in Fig. 8(c). All image points of \( I_i^{\text{gradient}} \) (Fig. 8(d)) that are outside the eyelid contour enhancement image are removed (Fig. 8(e)). Vertical edges of the resulting image are eliminated by applying a rank filter with filter size 1 by 7 (Fig. 8(f)). For the removal of very thin edges and to bring out strong edges an average filter with a filter size of 5 by 3 is applied (Fig. 8(g)). As the eyelid contour has a roughly horizontal orientation the width of the filter kernel is larger than the height in order to strengthen its effect in the horizontal direction.

To binarise the processed gradient image the maximum absolute value of each image column \( c \) of
Figure 8: (a) Extracted ROI with eyelid contour of preceding image $I_{i-1}$. (b) Eyelid contour of (a) shifted according to curvature model $cm_{[ue,le]}$. (c) Shifted eyelid contour of (b) thickened. (d) Extracted ROI of gradient image $I_{i}^{gradient}$ with pupil removed (as in Fig. 6(f)). (e) Removal of image points that are not part of (c) from (d). (f) Rank filter applied on (e). (g) Average filter applied on (f). (h) Edge image $I_{i}^{edge}$ (Binarisation of (g)).

$$I_{i}^{gradient}(r,c) \text{ is determined for } 1 \leq c \leq C_g:\n$$

$$v_{\text{max}}(c) = \max_{1 \leq r \leq R_g} |I_{i}^{gradient}(r,c)|. \quad (6)$$

$C_g$ denotes the width of image $I_{i}^{gradient}$ and $R_g$ its height, which correspond to the dimensions of the current ROI. Then the binarisation is performed using the product of a constant $t_g$ and the value $v_{\text{max}}(c)$ as a combined dynamic threshold for each image column $c$:

$$I_{i}^{edge}(r,c) = \begin{cases} 
1, & |I_{i}^{gradient}(r,c)| > t_g \cdot v_{\text{max}}(c) \\
0, & \text{otherwise} \end{cases} \quad (7)$$

A value $t_g = 0.7$ as constant threshold has proven effective in our case. Using this approach only the most prominent edges in each image column of the gradient image $I_{i}^{gradient}$ are regarded as eyelid contour points. With that, in each image column $c$ all pixels in the gradient image having a magnitude of more than $0.7 \cdot v_{\text{max}}(c)$ are considered as eyelid contour points. The dynamic determination of the threshold in each image column, depending on the largest gradient magnitude in this column, is used in order to account for differences in the strength of the gradient along the eyelid contour, as can be seen in Fig. 8(d). The result of this binarisation is depicted in Fig. 8(h). The preceding stages of applying a rank filter and an average filter to the gradient image have proved to lead to a more robust edge image $I_{i}^{edge}$.

4.5.6. Hough Transform

The best fitting line through the edge pixels of image $I_{i}^{edge}$ is determined using a standard Hough transform for lines (Fig. 9(a)). Subsequently, the line is thickened using a morphological dilation operation
using a square structure element with edge length 4 (Fig. 9(b)). The points of $I_{i}^{e_{d}}$ that do not concur with the thickened line are image points that do not belong to the actual eyelid and are therefore discarded (Fig. 9(c)).

Figure 9: (a) Binary edges (as in Fig. 8(h)) with best fitting line. (b) Line in (a) thickened. (c) Edge points that concur with the line in (b). (d) Image (c) shifted by inverse curvature model. (e) Best fitting polynomial through image points of (d).

4.5.7. Polynomial Regression

The remaining image points are shifted in the inverse direction of the curvature model $cm_{ue,le}^{i}$ (Fig. 9(d)). Finally, the best fitting second order polynomial through these points is determined using a polynomial regression (Fig. 9(c)). This polynomial $e_{abs}^{i}(c)$ is regarded to be the eyelid edge. It is uniquely determined by its coefficients $q_0$, $q_1$ and $q_2$ as:

$$e_{abs}^{i}(c) = q_2 c^2 + q_1 c + q_0.$$  \hspace{1cm} (8)

4.5.8. Eyelid Distance Determination

The distance between the upper eyelid and the lower eyelid in image $I_{i}$ is defined to be the distance of pixels between the polynomial for the upper eyelid $e_{abs,ue}^{i}(c)$ and the polynomial for the lower eyelid $e_{abs,le}^{i}(c)$ in image column $c_{d}$:

$$d_{i} = \max \left( 0, e_{abs,le}^{i}(c_{d}) - e_{abs,ue}^{i}(c_{d}) \right).$$  \hspace{1cm} (9)

The difference of $e_{abs,le}^{i}(c_{d})$ and $e_{abs,ue}^{i}(c_{d})$ would allow the eyelid distance $d_{i}$ to be negative. In particular this is the case if the lower eyelid in image column $c_{d}$ is detected to lie above the upper eyelid, which occurs in images depicting an entirely closed eye because the detection of the upper and the lower eyelid is problematic while the eye is entirely closed. To overcome this problem the smallest value allowed for $d_{i}$ is defined to be 0 as this is the correct distance between the upper and the lower eyelid for an entirely closed eye.

The image column $c_{d}$ is determined once after the first iteration of a trial by finding the image column that maximises the distance between the upper eyelid polynomial and the lower eyelid polynomial:

$$c_{d} = \arg \max_{1 \leq c \leq C} \left( e_{abs,le}^{i}(c) - e_{abs,ue}^{i}(c) \right).$$  \hspace{1cm} (10)
4.5.9. Curvature Model Adaption

In order to adapt the curvature model dynamically the ratio of the current eyelid distance \( d_i \) and the eyelid distance of the first processed image \( d_1 \) is determined as

\[
d_{ratio}^i = \frac{d_i}{d_1}.
\]  

(11)

The curvature model \( cm_{\{ue,le\}}^i \) of image \( I_i \) is defined as

\[
cm_{\{ue,le\}}^i = \text{round} \left( cm_{\{ue,le\}} \sqrt[3]{d_{ratio}^i} \right).
\]  

(12)

The eye is always closing (or at least remaining opened), hence \( d_i \leq d_1 \) (under the assumption that no noise is present). Furthermore, \( d_i \geq 0 \) and therefore the ratio always lies in the interval \([0, 1]\). This ensures that for an open eye \((d_{ratio}^i = 1)\) the original curvature model is used and that a straight line is used as curvature model for a fully closed eye \((d_{ratio}^i = 0)\), as described in [27]. For \( 0 < d_{ratio}^i < 1 \) the curvature model is compressed vertically. In Fig. 10 it can be seen that using the cube root of \( d_{ratio}^i \) is a good approximation of the eyelid deformation during closing.

![Figure 10: Comparison of eyelid deformation factors.](image)

Figure 10: Comparison of eyelid deformation factors. The horizontal axis shows the image number \( i \) and the vertical axis the corresponding value for the deformation factor. A trial where the detected eyelid polynomial \( e_{abs}^i \) is exactly placed on the actual eyelid contour has been used for creating this plot (manual inspection). The relative shape of the upper eyelid contour in image \( I_i \) is determined by \( e_{rel}^i = e_{abs}^i - \min_{1 \leq c \leq C} (e_{abs}^i(c)) \). The values of \( a_i \) in the equation \( e_{rel}^i = cm_{\{ue\}} \cdot a_i \) are determined for each \( i \) using a least-squares approach and the result is plotted as black line, which shows the linear factors that transform the original curvature model \( cm_{\{ue\}} \) best to the actual eyelid shape \( e_{rel}^i \) at image frame \( i \). The dashed grey line depicts the values for \( d_{ratio}^i \) (according to eq. 11) and the solid grey line depicts the values for \( \sqrt[3]{d_{ratio}^i} \). Note that the range where the eye is closing is the most relevant part \((i \in \{40, ..., 120\})\) because the dynamic model adaption is essential in this interval only.

5. Results

5.1. Ground-Truth Creation

For creating the ground truth two points in an image have been selected manually by the author. These points are selected in such a way that they lie on the eyelid contours of the upper (or the lower eyelid, respectively) and the distance of the two points in the vertical direction at image column \( c_d \) is maximised. With that, the horizontal position of the points is determined by the highest point on the upper eyelid contour and the lowest point on the lower eyelid contour, respectively. The vertical component of the distance between both points is used as ground-truth distance.

As the eyelid appearance differs amongst persons, in particular regarding the width of the eyelid contour, it is not possible for human observers to determine an exact and unique position of the eyelid contour with a vertical accuracy of a single pixel. By repeating the above described procedure for a subset of images of different persons with different eye states, an uncertainty range of approximately 0.5mm (corresponds to 3 pixels in our setting) in the vertical direction, has been found out experimentally.
As it is unfeasible to perform this procedure for every single image frame in a trial, a number of images have been left out and then the eyelid positions between two images that are annotated with these points are interpolated linearly. In order to ensure that the interpolated eyelid positions are properly placed on the actual eyelid contour, the intermediate images have been inspected visually. In cases where the interpolation was not accurate, additional intermediate images have been annotated manually until all eyelid positions were determined accurately.

5.2. Occurring Problems

Two classes of problems that occur during eyelid contour segmentation can be identified (images belonging to these classes are both labelled as \textit{Wrong} during evaluation).

5.2.1. Wrong Edges

The first problem is that an edge which is not the actual eyelid edge is detected as such. This mainly happens if either the eyelashes or wrinkles form another edge above the eyelid edge, as seen in Fig. 11.

![Figure 11: Wrong edge detected as upper eyelid.](image)

If the edge that is detected as eyelid is consistent in the forward sequence $S_f$ and the backward sequence $S_b$ then the eyelid distance has a constant error offset. After manual inspection and identification of this error, the offset can be subtracted from the determined eyelid distance.

However, if the eyelid contour determined in the forward sequence $S_f$ differs from the eyelid contour determined in the backward sequence $S_b$ the error can not easily corrected by subtracting a constant offset. It must be pointed out that this case rarely occurred in our experiments.

5.2.2. Eyelashes

The other problem are eyelashes that move in front of the eyelid while closing the eye. The eyelashes cover the eyeball partially and then the edge between the eyelashes and the eyeball is detected as eyelid contour. This usually happens if the eyelash area is more dense than the non-eyelash area, leading to a failure in the eyelash removal, as pointed out in 4.5.3. Fig. 12 portrays the mentioned problem with a mis-detected eyelid contour.

![Figure 12: Eyelash that covers the eyelid forms a new edge that is detected as eyelid contour.](image)
5.3. Evaluation of Method

A total of 103 trials were available during the development of the algorithm. However, 7 trials have been identified as invalid trials, due to a moving eye, a moving pupil or because the trial was beginning with a fully closed eye. With that, a total of 96 valid trials are regarded for further evaluation. The complete set of the resulting 96 valid trials has been evaluated qualitatively.

The eyelid contour detection and tracking has proved to be useful in 87 of the trials. In 9 trials the segmentation failed due to entirely mis-detected eyelid contours or inconsistencies between the eyelid contours detected in the forward sequence and the backward sequence.

In 62 of the 87 trials the eyelid contours have been detected perfectly. These cases, where a human observer assessed that a more accurate manual identification of the eyelids would not be possible, are labelled as Perfect. The remaining 25 cases, where a human observer assessed that a small constant error was present, are labelled as Good, because the small constant deviation can easily be corrected with a minimum of effort. The results are summarised in Table 1.

Additionally to this qualitative evaluation, a quantitative evaluation using the created ground-truth has been conducted for a subset of 9 of the 96 trials, each being from a different person. This quantitative evaluation has not been carried out for all 96 trials because the manual ground-truth creation is unfeasible. The manually measured eyelid distances as described in section 5.1 (grey) and the eyelid distances determined by our method (black) are shown in Fig. 13. The detected eyelid distances shown in Figs. 13(a)-13(h) are examples for trials labelled as Perfect and the trial in Fig. 13(i), having a constant error, is an example of a trial labelled as Good.

### Table 1: Evaluation of method summary

<table>
<thead>
<tr>
<th></th>
<th>Perfect</th>
<th>Good</th>
<th>Wrong</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>#Trials</td>
<td>62</td>
<td>25</td>
<td>9</td>
<td>96</td>
</tr>
<tr>
<td>Percentage</td>
<td>64.6%</td>
<td>26%</td>
<td>9.4%</td>
<td>100%</td>
</tr>
</tbody>
</table>

6. Conclusion

6.1. Practical Value

The algorithm has been implemented and evaluated in Matlab R2011b 64-bit on a Core 2 Duo 2.4 GHz processor with 4 GB RAM. The approximate processing of 96 trials, each comprising 500 frames, took 72 min, leading to an average processing time of 45 s per trial. As the processing of trials is performed after data acquisition, i.e. there is no requirement for real-time processing, a processing time of 45 s per trial is absolutely reasonable. The reduction of processing time is possible by processing only the part of a trial where the actual eye-blink occurred, which is determined during eyelid detection as described in section 4.5.1. However, in our case we preferred obtaining the eyelid distance over the entire trial in order to avoid the loss of potentially useful information for further assessment and thus this has not been implemented.

One factor for the achievement of a reasonable processing time was the complexity reduction by using curvature models. By transforming the eyelid, being originally an elliptic arc, to a straight line, a standard Hough transform for lines can be used for line detection. The elliptic Hough transform has fourth degree polynomial time complexity whereas the standard Hough transform for lines has quadratic time complexity.

The user interaction required for the algorithm to work correctly is limited to the model creation stage, where the experimenter has to define the shape of the upper and the lower eyelid once per subject. This creation of a curvature model for each person comprises 6 mouse clicks which can be performed within a few seconds. In most of the cases a single curvature model was suitable for each person in order to process all trials of this particular person. We have observed the cases that a single curvature model was not sufficient when the vertical eye position in an individual trial was considerably different than the vertical eye position seen in the image that was used for model creation. In prospective studies where the described method
is used for the automatic eye-blink measurement the number of trials for each person will be in the range of 50–200. Hence, the user intervention required for a precise eye-blink measurement of a high number of trials is kept to a minimum.

Our experiments have shown that using a single curvature model for all persons is not sufficient, in contrast to [22]. We believe that this has mainly two reasons: On the one hand the anatomical orientation of the eyes among the individual subjects is different (mathematically, the slope of the line that connects both corners of an eye varies between individual persons). On the other hand, a different curvature of the eyelids is present due to a different eye size of the subjects, leading to different curvatures of the eyelids. With that, using individual curvature models for each person results in a more accurate and more robust eyelid detection than using a single model that approximates the average shape of the individual eyelids. As in general both eyes of a person are symmetrical, the curvature model created for one eye can be mirrored and then used for eyelid detection and tracking of the other eye of the same person.

However, for improving the convenience of the analysis of trials further research regarding a more advanced model creation stage must be conducted. An approach would be to investigate whether a set of general curvature models is sufficient for a universal eyelid contour shape description. With that, no new curvature models need to be created if trials of previously unknown subjects are to be processed. However, a method for the selection of the best fitting curvature model of the entire set must be developed. Another approach would be to create the individual curvature models automatically by using an elliptic Hough

Figure 13: Comparison of manually measured ground-truth eyelid distance (grey) and detected eyelid distance (black) for trials (a)-(i) and their root-mean-square error (RMSE). The horizontal axis indicates the image index (1 ≤ i ≤ N, from left to right) and the vertical axis indicates the distance between the upper and the lower eyelid of image $I_i$, measured in image pixels. The results of trials (a)-(h) are labelled as Perfect and (i) is labelled as Good.
Further investigations for the solution of the mentioned problems must be conducted in order to reach a higher success rate. Our method has two major advantages compared to the tracking of a point on the upper eyelid and a point on the internal canthus [18]: each of the points that is tracked needs to be selected in each trial separately, leading to a high degree of user interactivity required for the analysis of eye-blinks in trials. Furthermore, by using these two points merely a relative measurement of the eye-close state is performed, rather than an exact absolute distance between the upper and the lower eyelid. By using point tracking a further third point that is placed on the lower eyelid is required for exact distance measurements, which would increase the required user intervention by 50%.

Our algorithm does not employ any ways of handling a moving head or a moving eye because these cases are eliminated by stabilising the persons head and the instruction to look at a fixed position. However, in cases where the head or the eye are allowed to move, the displacements due to movements can be compensated by the tracking of distinctive points that have a constant position relative to the eye, such as the corner of the eye as described in [18].

In summary, even with the mentioned problems that arise during processing the developed method has proved to be highly beneficial for the automatic measurement of the eyelid distance during eye-blinks in high-speed video records because merely a minimum of time-consuming user interaction is required. Given similar recording conditions, our proposed technique is most likely to be applicable to other eye images that have been acquired by a high-speed eye tracker.

7. Discussion

The main problems that occur during eyelid contour detection have been identified in section 5.2 as the detection of wrong edges and eyelashes.

The detection of wrong edges as eyelid edges is a challenging problem because our algorithm simply assumes that the strongest edge is the eyelid edge. We have addressed most of these cases by the restriction of the region where the eyelid contour is expected. Luckily, the remaining cases where wrong edges are detected were rather rare and thus no further elaborations for coping with these cases have been investigated.

On the other hand, eyelashes are a more severe problem because eyelashes are—with varying extent—in general present in each trial. Mainly two methodological approaches have been utilised in order to provide a good support of eyelid contour detection even with eyelashes. First, a horizontal rank filter has been used in order to (partially) eliminate eyelashes (see section 4.5.3). The second approach is based on the fact that in general the eyelashes do not interfere with the eyelid contour detection while the eye is open. With that, a proper identification of the eyelid contour is possible at the beginning of each sequence. However, during closing of the eye, the eyelashes move in front of the eyelid, covering the actual eyelid contour. This problem has been addressed by the restriction of potential edge points to a small region that depends on the eyelid contour position in the previous image frame as described in section 4.5.5.

8. Conflicts of Interest

There are no conflicts of interest.

9. Mode of Availability of Software

A prototype of the developed software can be issued for non-commercial use upon request per email to the corresponding author.


19
A. L. Yuille, P. W. Hallinan, D. S. Cohen, Feature extraction from faces using deformable templates, International Journal

Y. Wu, H. Liu, Z. Ma, A new method of detecting human eyelids based on deformable templates., in: SMC (1), IEEE,


URL http://dl.acm.org/citation.cfm?id=1704555.1704733


URL http://dx.doi.org/10.1109/ICASSP.2000.859314


URL http://dx.doi.org/10.1007/BF00127169
[34] P. Hough, Method and means for recognizing complex patterns (1962).