Image and Vision Computing 108 (2021) 104124

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Improving eye movement biometrics in low frame rate eye-tracking devices using periocular and eye blinking features

Sherif Nagib Abbas Seha [a,*], Dimitrios Hatzinakos [a], Ali Shahidi Zandi [b], Felix J.E. Comeau [b]

[a] Electrical and Computer Engineering Department, University of Toronto, Canada
[b] Alcohol Countermeasure Systems (ACS) Corp., Toronto, Canada

ARTICLE INFO

Article history:
Received 31 January 2020
Received in revised form 21 September 2020
Accepted 30 January 2021
Available online 6 February 2021

Keywords:
Eye movements
Eye blinking
Periocular features
Multi-modal biometrics
Continuous driver authentication

ABSTRACT

In this paper, the biometric potential of eye movement patterns extracted from low frame rate eye-tracking devices is evaluated. Also, possible improvement in recognition rates is investigated using other static and dynamic features extracted from the eyes including eye blinking patterns and periocular shape features. These modalities can be applicable for specific biometric applications like continuous driver authentication for law enforcement. For this purpose, two databases are collected with two low frame rate eye-tracking systems that capture the eye movements. Data were recorded from 55 participants while watching real driving sessions. For eye gaze, features from fixations and saccades are extracted separately including duration, amplitude, and statistical features. For eye blinking, features from the blinking pattern, its speed, acceleration, and power per unit mass profiles are extracted. Periocular features include the eye-opening height, width, axial ratio, etc. Each modality is evaluated first, then, these modalities are combined in a multi-modal setup for performance improvement. While each trait achieved a moderate performance in a single-modality setup, the fusion of the static and the dynamic features from the eye provides a great performance improvement up to 98.5% recognition rate and 0% error rate in both modes of authentication. Although the single-modality setup might not be secure enough, the fusion of these traits achieves high levels of identification making these traits effective for continuous driver authentication application.

1. Introduction

In the automobile industry, vehicle security is considered crucial to maintaining safe and responsible driving. Recently, advanced technologies adopting machine learning and computer vision techniques are integrated into smart cars to provide a comfortable and safe driving experience. Examples of these technologies include collision detection systems and autonomous driving. However, one of the key aspects of vehicle security that has not yet been fully resolved is continuous driver authentication [1]. Driver authentication is an important aspect that enhances passengers' safety in special vehicles (e.g. cash trucks and special force vehicles) and public transport (e.g. busses or metro). Only authorized people are allowed to drive these kinds of vehicles. Additionally, this technology can be integrated into personal cars for law enforcement which leverage responsible driving experience as it will help to prevent impersonation for devices such as Breathalyzers.

While conventional biometrics traits like facial characteristics, fingerprints, and iris can be employed for driver authentication applications, most of these traits are intrusive and require the drivers' cooperation to provide their biometric trait. For instance, fingerprints can be used in one-time authentication mode (e.g. to ignite the car) but it will increase the distraction levels in continuous authentication mode as the driver has to touch a fingerprint sensor every few minutes. Additionally, static traits lack the issue of security as they can be easily synthesized to spoof a system (e.g. gummy fingers, face photo, and 3D masks [2,3]). This hinders these modalities from integration in smart vehicles for continuous driver authentication applications.

Recently, the research community has directed its efforts to investigate the uniqueness levels inherited in behavioral and physiological biometrics. This group of signals are almost impossible to be replicated as they are caused by subconscious processes. A perfect modality for continuous driver authentication application is the eye movements that have been recently shown to have subject-unique behavioral and physiological patterns [4,5]. Eye movements can be simply and remotely registered using a camera sensor without any intervention from the driver's side. Additionally, these traits can provide an accurate assessment of driving behavior by measuring distraction/vigilance levels of the driver. Therefore, eye movements can be effectively deployed in vehicles to continuously verify the driver's identity and also assess their driving behavior.

* Corresponding author.
E-mail address: sherif.seha@mail.utoronto.ca (S.N.A. Seha).

https://doi.org/10.1016/j.imavis.2021.104124
0262-8856/

Eye movements study tries to analyze the behavioral patterns in addition to characterizing the physiological features of the tissues and the muscles generating eye movements. This helped researchers in understanding human behaviors in different domains like reading, image screening, sleeping, solving arithmetic problems, and driving [6,7]. Although eye movements have shown high levels of subject uniqueness, the current setup is considered impractical for biometric authentication applications, especially driver authentication. Previous studies of eye movement-based biometrics incorporated a complicated setup that involves expensive and high-accuracy eye-tracking devices. Other limitations include the requirement of data collection under a controlled environment where participants have to place their head on a chin-rest with minimum or no head movements allowed. Also, eye movements are registered under specific visual stimulation like reading texts or following a randomly moving target on a screen. All these obstacles have prevented eye movement biometrics from taking part in practical applications.

In this paper, we investigate the potential of eye movement biometrics using cost-effective and low frame rate eye-tracking devices in the wild. Our approach is to evaluate cost-effective eye-tracking devices that can be easily integrated into a car for continuous driver authentication. According to our knowledge, this is the first study to evaluate eye movements in the wild for a biometric authentication application. However, due to the relaxed conditions in the acquisition process, lower recognition rates are expected. This is mainly because low frame rate devices can only extract behavioral eye movement characteristics but not capable enough to characterize the physiological traits of the eye. Therefore, in this paper, we propose two additional traits that can be extracted from the periocular region to improve the recognition rates. These traits rely on both the static and the dynamic features of the eyes; dynamic features comprise the blinking behavior while static features include 2D features from the periocular region alignment. These additional traits rely on eye landmark detection (or periocular region detection) which is an essential step in eye-tracking devices and can be applied to any IR camera. Additionally, these modalities do not require a high frame rate and share extensive discriminant behavioral and physiological features as shown in previous studies [8–10]. This inspired us to investigate the feasibility of building a multi-modal system that adopts these traits for a practical application like continuous driver authentication.

This paper is a continuation of our earlier work published in [8]. We extend our work through the following main contribution points:

1. Adopting a new technique for characterizing the blinking behavior of the eye using periocular region alignment. This technique is more accurate than estimating the blinking pattern using the reflected light from the eyelid which is adopted in [8]. For this task, A customized eye alignment model is trained to work on images of incomplete or partial faces which is according to our knowledge is a novelty in face alignment techniques.
2. New features are extracted from the power profile of the blinking patterns which showed improvement in the performance of our proposed system (in the single- and multi-modal setups) compared to [8].
3. Proposing a new set of static features based on the 2D shape of the eye.
4. Evaluation of the dynamic and static features of the eye on a new dataset built using the lower frame-rate device over a larger population size.

2. Related work

2.1. Eye-based dynamic biometrics

Regarding eye movements, various approaches have been previously proposed in the context of biometric authentication. However, to provide a brief overview of previous approaches, we focus on the various methods proposed in the 2015 BioEye competition [11,12]. The most common approach for extracting features from eye movements is to classify them into fixations or saccades and extract different features from each class [13,14]. This approach was followed by A. George and A. Routray where their system achieved the best identification performance in a two-session setup [13]. Common features for both classes like duration, path length, average velocity, angle with previous fix/saccade instance were extracted. Other features like velocity and acceleration profiles were extracted solely from saccades. Using this setup a correct recognition rate of 89.54% over a population of 153 users was achieved. Other proposed systems followed a different approach where signal processing techniques are applied directly to the estimated gaze position in time or frequency domain. Examples of these techniques are the Fourier Transform (FT), Gaussian Mixture Models (GMM), or Cepstrum Transform (CST) [15,16].

On the other hand, biometric authentication based on eye blinking patterns was not investigated until recently with the most approaches adopting the invasive way for acquisition using EOG. The first system to use EOG blinking patterns for biometrics was proposed by S. Seha and M. Abo-Zahhad [17,18]. Features based on the time delineation of the blinking pattern where extracted like amplitude, duration and sloe of the eyelid closure and opening phases. Using this setup, a recognition rate of 93.75% and an equal error rate of 7.45% were achieved in both modes of authentication over a population of 40 users. Moreover, EOG eye blinking patterns were found applicable in multi-modal systems where they were combined with brainwaves to improve the overall system's performance [19,20]. A different technique for registering eye blinking patterns was followed by J. Espinosa el al. [9], where the blinking patterns were captured remotely using a high-speed camera. The amount of light reflected from the eye region during blinking can easily reflect the blinking behavior of an eye. Features from the registered blinking pattern along with speed, acceleration and power profiles were employed to generate a subject-unique template. A correct recognition rate of up to 99.7% was achieved in identification mode. Finally, eye blinking detection was showed to be helpful in liveliness detection to prevent face spoofing in facial recognition systems [21].

2.2. Previous approaches for continuous driver authentication

Approaches adopted in previous studies for continuous driver authentication mainly rely on the concept of driver fingerprinting; i.e. collecting features from the Controller Area Network (CAN) data in the vehicle including steering wheel angle, speed variation, acceleration status and gas/brake pedals pressure. These features can assess the driver style, and recently employed for driver authentication [22]. For instance, the authors in [23] have modeled the raw and the Cepstrum versions of the driver fingerprinting data, especially the following-distance patterns, using GMM for driver identification purposes. The best recognition rates achieved were 89.6% (over 12 drivers) and 76.8% (over 276 drivers) using simulated and real driving data, respectively. Again in [24], the authors used GMM to model the acceleration patterns, brake pedal pressure, and their derivatives for driver identification and verification tasks. Over different subsets of 10 drivers, the best correct identification and verification error rates were 93.75% and 2.5%, respectively, using Bayesian optimization. The study in [25] evaluated driver fingerprinting data acquired from simulated driving sessions for a driver verification application. Features included brake/acceleration pressure, vehicle speed, steering wheel angles, and total distance traveled. Using a one-class SVM model, an equal error rate of 14.7% was achieved over 10 drivers. Recently in [1], the authors adopted a Convolutional Neural Network (CNN) to extract subject-unique feature maps from the characteristic data extracted from the CAN. Using real driving sessions with two different cars, an average identification rate of $\approx$ 98% was achieved over 15 drivers.

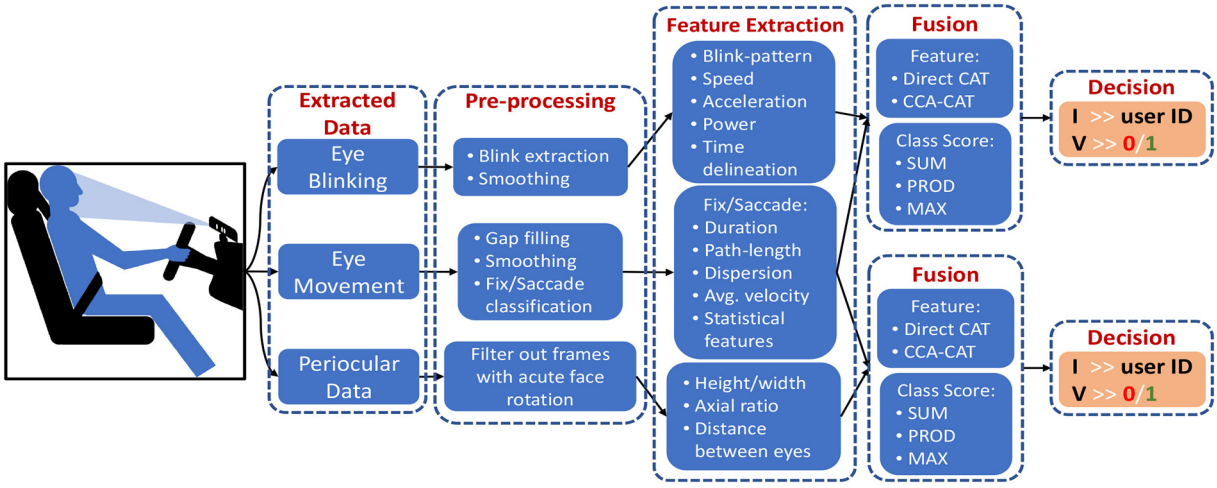


Fig. 1. Block diagram of the proposed multi-modal biometric authentication system.

Based on the above overview, although eye movement biometrics showed promising results for human identification, it has not been previously evaluated in the context of a practical application, e.g. driver authentication, due to the complexity in the acquisition and the expensive eye-tracking devices required to model eye movement characteristics. To reduce the gap between these two worlds, this paper investigates the employment of eye movement biometrics in the wild (i.e. using low frame rate and cost-effective eye-tracking devices without visual stimulation task) for driver authentication. Additionally, in this paper, we take advantage of the eye alignment feature implemented in the eye-tracking devices to extract additional dynamic (blinking) and static (periocular shape) biometric features to improve the performance.

3. Proposed system

The proposed multi-modal biometric authentication system comprises four main steps; a) raw data acquisition and eye-related data extraction (Sections 3.1 and 3.2), b) pre-processing (Section 3.3), c) feature extraction (Section 3.4), d) feature or class score fusion (Sections 3.5 and 3.6). Fig. 1 shows a block diagram that summarizes the proposed system.

3.1. Databases

The databases used in this paper were collected in the Biometric Security (BioSec.) Lab after having approval from the research ethics board of the University of Toronto (Protocol # 00035655). The purpose and the procedure of the experiment were discussed with the participants who signed a consent form before conducting the experiment. The participants were 18 years or older and have a normal or corrected-to-normal vision.

Table 1
Summary of the properties for the eye tracking devices that are used in this paper.

	GP3 system	AC system
Camera sensor	IR	IR
Gaze frame rate (Hz)	60	13–14
Gaze accuracy (degree)	0.5–1	3–5
Face and eye alignment	No	Yes
Gaze data	2D projected gaze for each eye	Angular gaze (average of both eyes)
Driver monitoring	No	Yes
Calibration method & time	9 points (around 30 *s*)	Direct gaze to camera (less than 10 *s*)

3.1.1. Eye-tracking devices

Two commercial eye-tracking devices with low frame rates were evaluated for the eye movement biometrics in this paper. The first device is the Gazepoint (GP3) with a 60 *Hz* frame rate and the second one is the Autocruis (AC) driver monitoring system with a maximum frame rate of 30 *Hz*. Although the AC system has a 30 *Hz* frame rate, activating the built-in real-time gaze detection algorithm lowers the frame rate to 13–14 *Hz*. Table 1 shows a brief comparison between these two devices.

3.1.2. Acquisition protocol

Data was captured from participants while watching videos of a driving session. The videos attempt to create a similar environment to that of a driver, hence, the visual stimulation is in the form of video clips of a real driving session (from the driver's point of view) in downtown Toronto.[1] The video is trimmed into different clips each of duration 5 *min*. For the GP3, each subject has performed three trials by watching a different clip each trial. A 1–2 *min* relaxation period was conducted between trials. For the AC system, since it has lower accuracy and lower frame-rate, more data was collected from participants where they conducted a fourth trial. The exact timing for a one trial recording is illustrated in Fig. 2. Both devices were calibrated at the beginning of the recording. GP3 uses a 9-point calibration method which takes approximately 30 *s* to calibrate. AC is calibrated differently where users are required to fix their gaze to the camera directly for about 10 *s*. A 15 *s* introduction is displayed before each video clip with the instructions as well as a cross sign for the subjects to center their gaze. After that, the video clip of the driving session is displayed for 5 *min*. The subjects were seated about 0.5–0.6 *m* away from the IR camera. Based on the installation instructions, the GP3 system was fixed below the display screen facing up towards the subject at an angle of about 30 degrees. For the AC system, it was placed on the top of the display screen with a down angle $< 20^{\circ}$. The videos are displayed on a 4:3 widescreen with size 20 in. and 1680×1050 resolution. Another laptop was used to run the recording software and save the data. The total number of subjects who participated in this experiment is 55. Data from 35 subjects were recorded using the AC system and 22 subjects using the GP3 (two subjects conducted the session twice using both devices).

[1] Driving session video: https://www.youtube.com/watch?v=rM8dbiH0kfY

Fig. 2. Time-line of a one trial recording.

3.2. Eye-related data extraction

For the eye movement data, the companion software for both systems (GP3 and AC) provides gaze estimation data. For the GP3 system, the 2D gaze coordinates (X and Y) for each eye are provided as a projection on the stimulus plane (along with their validation flags). For the AC system, only the average gaze estimation of both eyes is provided as angular values in the horizontal and the vertical planes (with validation flags). These values are extracted for pre-processing and feature extraction stages.

For the eye blinking patterns, they can be remotely registered using two approaches; a) light reflection [9], b) area within the eye [26]. Method (a), which was used in our previous work in [8], uses the average light intensity reflected from the periocular region to estimate the blinking behavior. This method requires only eye detection, however, it is not an effective estimator as it is affected by the lighting conditions, reflections from glasses, and the size of the periocular region. Method (b), which is adopted here, estimates the blinking behavior by calculating the area within the eye during eye blinking. This method requires an eye alignment step to accurately define the borders of the eye-opening.

As mentioned above, to use method (b), eye alignment is required first. This step is already implemented in the AC system as the companion software has a built-in eye landmarks detection feature that provides 8 landmarks for each eye. These landmarks were used to estimate the blinking patterns. However, this feature is not implemented for the GP3 system. Therefore, in this paper, we propose a Machine Learning (ML) model that is trained to detect the eye landmarks from the videos of the users' faces. These videos were captured from the GP3 control window using a third-party screen capturing software[2] at the same frame rate of the GP3 system (60 *Hz*). The details of the ML model are described in the next paragraph.

The proposed ML model is based on an ensemble of regression trees that was used before for fast and accurate face alignment [27]. This model was selected because it can run in real-time making it practical for our proposed application. While eye and face landmarks detection models are already available, they can't be applied directly to our recorded videos because these models work in a hierarchical fashion that requires all the features of the face to be present, however, GP3 frames capture only the top part of the face (forehead, eyes and nose). Therefore, a custom model is trained in this paper using the shape predictor function from the dlib package [28]. To train the model, around 600 frame images from the recorded videos were annotated manually using imglab[3] where each eye was aligned by 8 points. Then, the model was tested on about 5 K more frames which are then checked visually for manual correction. Finally, the model was trained again on all the labeled frames (≈5600 frames). Table 2 shows the parameters our ML model for the eye alignment. Fig. 3 shows our custom trained model in action.

Finally, for the periocular data, they were extracted from the same 8-point eye landmarks which are either provided directly by the AC software or extracted using our ML model from the GP3 data.

3.3. Pre-processing

3.3.1. Eye blinking data

Using the eye landmarks extracted from the previous section, the eye blinking pattern can be easily characterized using the area of the polygon defined by the landmarks within each frame. When eyes are open, the area of this polygon is large and it decreases noticeably when the eye closes as shown in Fig. 4. Based on the technical specifications, both devices have accurate detection for eye blinking status (≈98%). However, the definition of the blink validation flag (BKF) is different for both devices. For the AC system, it is an accumulator to estimate the blinking rate, so it is incremented by one whenever blink is started (Fig. 4a (top)). Therefore, the sequential difference (i.e. BKF(f +1) − BKF(f), where f is the frame index) of this accumulator is employed as a blink flag, this why the spike in the flag occurs before the blink (Fig. 4a (bottom)). In this case, a window is taken around this spike to find the maximum area within this window which defines the peak of a blink. For the GP3 system (Fig. 4b), the blink validation flag is assigned a value of 1 during the blinking state and 0 elsewhere. So, the peak of the blink occurs within the BKF value of 1. Therefore, no extra steps are required, and the validation flag is used directly.

One issue is noticed with the AC system, that is the eye landmarks are not accurately updated for the people wearing glasses during

Table 2
Parameters of our custom trained eye alignment detector.

Parameter	Value
Tree depth	5
Cascade depth	15
Feature pool size	5000
Number of test splits	500
Oversampling amount	50
Oversampling translation jitter	0.1

Fig. 3. Eye landmarks detection during different eye states; open, half-open, closed for the GP3 system.

[2] OBS studio: https://obsproject.com/download
[3] Image annotation website: https://imglab.in/

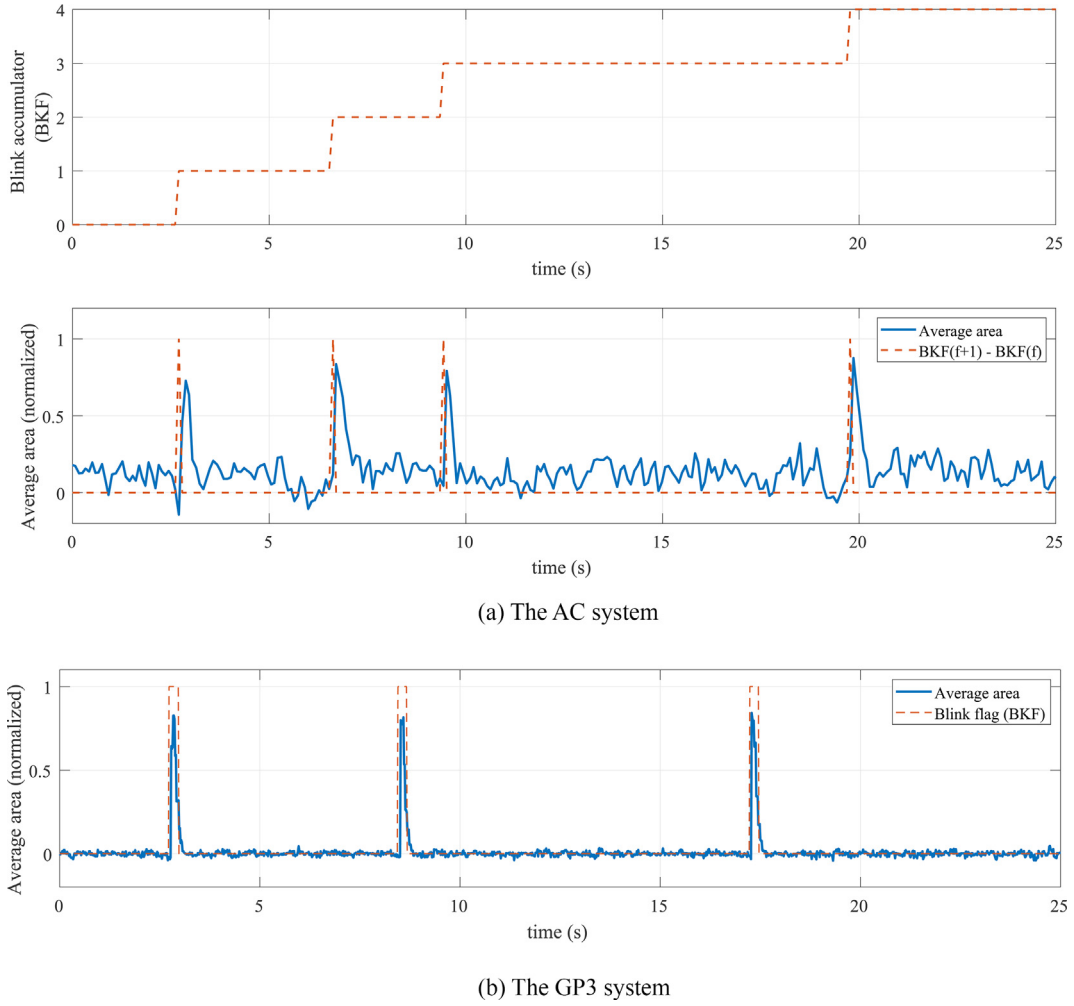


Fig. 4. Blink pattern estimation and detection using eye alignment. The vertical axis shows the inverted area within the polygon of the eye landmarks.

blinking making these patterns undetectable in most cases. To tackle this issue, a simple yet effective technique is applied to improve the detection of the eye blinking patterns. The blinking pattern estimated from the eye landmarks is multiplied by the eye confidence levels (left and right). This helps in correcting the eye blinking patterns estimated from users with glasses. The AC system can detect if the users are wearing glasses, and hence this information can be used to decide if this additional step is required or not. Fig. 5 shows the blinking patterns with and without correction for users with glasses.

3.3.1.1. Blink extraction. To extract eye blinks, the eye blinking pattern from both eyes were averaged and the BKF provided by each system was used to identify the blinking incidents within the frames. Within a window around the non-zero values of the BKF, the maximum peak of the area is detected. Then, the eye blinking is assumed to start before that peak by 15 frames and ends after 30 frames from the peak (eye blink average duration is about 300 ms). The blinking pattern from the AC system is linearly interpolated by 5 to have a similar frame rate as the GP system.

3.3.1.2. Blink smoothing. After extracting the blinking patterns from the GP3 and the AC systems, they are smoothed using a Spline smoothing curve (Figs. 6a (left) and 6b (left)). This process is done by creating continuous piece-wise 3^{rd} degree polynomials to create a smoothed curve representing each blink.

3.3.2. Eye movement data

As mentioned earlier, the GP3 companion software provides an estimation of the gaze position projected on the stimulus screen (2D plane). The best gaze position (which is the average of the left and right eye

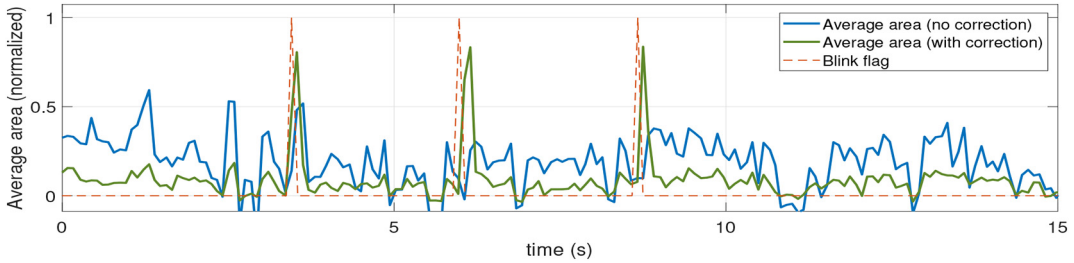
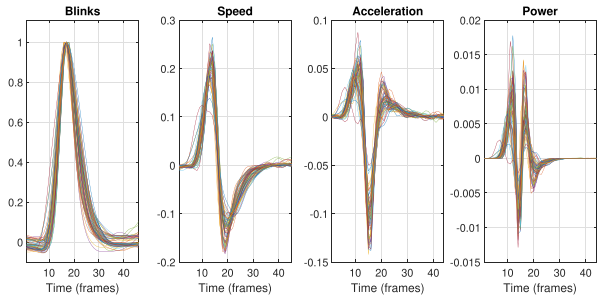
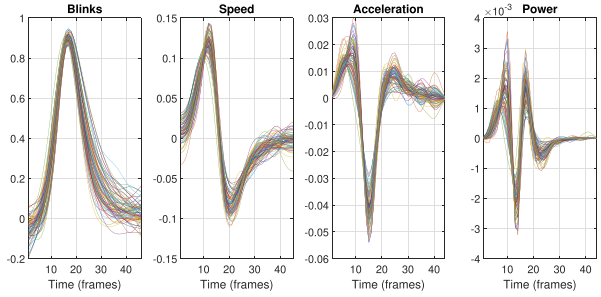


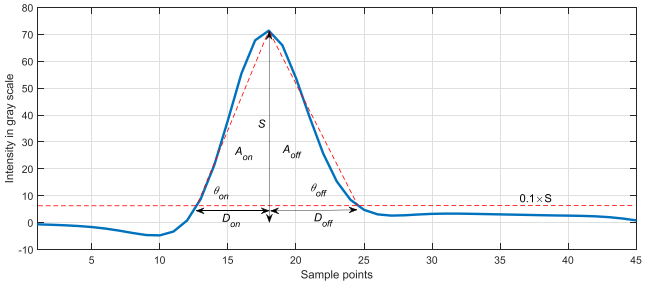
Fig. 5. Blink pattern correction for the users with glasses recorded by the AC system.



(a) Eye blinking pattern and features captured from the GP3 system



(b) Eye blinking pattern and features captured from the AC system



(c) Time delineation features extracted from an eye blinking pattern (symbols are explained in Table 3)

Fig. 6. Extraction of eye blinking patterns and features.

gaze) is used as the raw gaze position for our eye movement biometric system. This signal is denoted as BPOGX and BPOGY for the x and y coordinates, respectively. In contrast, the AC system provides two angular values to characterize the gaze direction within a frame, denoted here as *angX* and *angY*, that span the horizontal (azimuthal) and vertical (elevation) planes (eyeball centre is the origin of the coordinates). These two values were used as the raw gaze from the AC system. Before extracting features from these signals, a pre-processing stage comprises gap-filling, smoothing, and fixation/saccade classification is conducted as follows.

3.3.2.1. Gap filling. It is used to estimate the gaze position in the short-time instances when eyes are not detected. This can happen due to user- or algorithm-related issues. For example, this can happen due to eye blinking, rapid face rotation, eye blocking by the user, or strong light reflections [29]. These cases can be easily detected using the gaze validation flag (BPGV) in the GP3 system which provides a confidence level for the detected eyes. Similarly, the AC system provides a confidence level for each eye as well as face pose angles (yaw, pitch and roll). In these instances, the gaze position is estimated by linear interpolation using the position values before and after the window of eye detection failure.

3.3.2.2. Smoothing. It is applied to the whole recording per trial (5 *min*) using the Savitzky–Golay filter which fits a polynomial of order 3 in a window of 17 frames. This helps to minimize the noise in the velocity profile for better classification of fixation and saccadic instances.

3.3.2.3. Fix/saccade classification. In this paper, the velocity-based method (I-VT) [29,30] is adopted for fixation/saccade classification. For the GP3 system, the eye movement velocity is calculated based on the euclidean distance between the projected gaze of the current and previous frame as follows

$$V(i) = \frac{\sqrt{\left(X_f(i+1) - X_f(i)\right)^2 + \left(Y_f(i+1) - Y_f(i)\right)^2}}{F_t(i+1) - F_t(i)} \tag{1}$$

where V, X_f, Y_f, and F_t denote the 2D velocity, filtered gaze in X and Y directions, and frame time, $i = 1, 2, \ldots, N-1$ denotes the frame index and N is the total number of frames. A threshold of unit pixels/s is used for eye movement classification. In detail, a velocity greater than 1.8 *pixels/ms* identifies the onset of the saccade and below this value is considered a fixation. Moreover, short fixations that are less than 100 *ms* are discarded [13]. In contrast, for the AC system, the angle between the two vectors represented by the gaze angles in the 3D space is calculated. The vectors in two consecutive frames are transferred into Cartesian coordinates (along unit length vector) and the angular velocity is calculated as follows

$$V(i) = \cos^{-1}\left(\frac{\mathbf{g}(i+1).\mathbf{g}(i)}{\|\mathbf{g}(i+1)\|\|\mathbf{g}(i)\|}\right) \tag{2}$$

where $a.b$ is the dot product between a and b, and $\mathbf{g}(i)$ is a 3D vector representing the x, y, z coordinates transformation from the Spherical

coordinates defined by *angX* and *angY* which are provided by the AC system. A threshold of $25^{\circ}/s$ was set to classify fixations and saccades for the AC system.

3.3.3. Periocular data

The pre-processing stage simply filters out the frames where the face or the eye detection confidence levels are low. In detail, the frames with face and eye detection levels below 0.9 and 0.7, respectively, are discarded. Additionally, for the AC system, frames with acute face rotation are filtered out; i.e. frames with yaw and pitch angles greater than 35° and 25° are discarded.

3.4. Feature extraction

3.4.1. Eye blinking data

After extracting the blinks from each subject, features are extracted for classification. The set of features characterizes the time-delineation of the blinking pattern and its derivatives. The first and the second derivatives which represent the speed and the acceleration profiles of the blinking behavior (Fig. 6a and b) are estimated using the following two equations

$$S(f) = \frac{B(f+1) - B(f)}{1/f_r} \tag{3}$$

$$A(f) = \frac{S(f+1) - S(f)}{1/f_r} \tag{4}$$

where, f is the frame index, f_r is the frame rate and B, S, A are the blink, speed and acceleration patterns, respectively. The time-delineation process includes features from the opening and the closing phases of the eye blinking pattern like area, duration, slope, energy, and the strength of the blink. Fig. 6c and Table 3 provide more details about how these features are estimated from the blinking patterns. Moreover, the power per unit of mass, $P(f)$, which describes the force of the eyelid muscles per unit of time during blinking is computed (Figs. 6a (right) and 6b (right)). This value is estimated as follows [9].

$$P(f) = S(f) \times A(f) \tag{5}$$

Time delineation features also included the extraction of the peaks and valleys and their time indices from the speed (one peak and one valley), acceleration (two peaks and one valley), and power (two peaks and two valleys) profiles of the eye blinking pattern.

3.4.2. Eye movement data

The features employed here were inspired mainly by the previous works in the 2015 BioEye competition, specifically, the work by A. George and A. Routray [13] as it achieved the best recognition accuracy, and the overview of eye movement features evaluation in [31]. This step is based on extracting different features for fixation and saccadic periods separately. Tables 4 and 5 provide the list of features that are extracted from fixation and saccade, respectively. Some features are common for both classes like duration, path length, dispersion, average velocity, etc. Other statistical features like mean, median, maximum, standard deviation, skewness and kurtosis are extracted only from the speed and acceleration patterns of the saccades. Finally, the two sets of features are concatenated together to form one feature vector.

3.4.3. Periocular data

The third type of features extracted from the eyes is based on the 2D shape of the left and the right eyes. These features are extracted from the detected landmarks as described in Section 3.2. These features include the width and the height of the opened eye (left and right), axial ratio (left and right), the distance between both eyes (calculated from the center, the nearest and the furthest edges) where the eye centre is determined by the centroid of the polygon defined by the eye landmarks. Two more features are used for the AC system include the left and the right iris diameter.

3.5. Multi-modal system: fusion of two modalities

3.5.1. Feature-level fusion

For feature fusion, two techniques are adopted; feature concatenation (F_{CAT}) and Canonical Correlation Analysis (CCA - F_{CCA}).

- F_{CAT}: This is simply achieved by concatenating together the training feature matrices extracted from each modality as described in Sections 3.4.1, 3.4.2, and 3.4.3. Before training with the fused matrix, the features are *z*-score normalized because the concatenated features have different dynamic ranges and units. The same procedure is followed for the testing feature matrices before classification.
- F_{CCA}: The basic idea behind CCA is that it maximizes the correlation between two sets of random vectors [32,33]. Assume we have two data sets, $X \in \mathbb{R}^{d_x \times n}$ and $Y \in \mathbb{R}^{d_y \times n}$, representing the eye movement and the eye blinking training features from all subjects. Then CCA finds a pair of projecting matrices $W_X = \{\alpha_1, \alpha_2, \ldots, \alpha_d\}$ ($W_X \in \mathbb{R}^{d_x \times d}$) and $W_Y = \{\beta_1, \beta_2, \ldots, \beta_d\}$ ($W_Y \in \mathbb{R}^{d_y \times d}$) where $d = min\,(rank(X), rank(Y))$. CCA maximizes the correlation between the projections $\alpha_1^T X$ and $\beta_1^T Y$ which are known as the first pair of canonical variates. Similarly, the second pair of canonical variates $\alpha_2^T X$ and $\beta_2^T Y$ are maximally correlated, however, these pairs are uncorrelated with the first pair. This continues until all the correlation features of X and Y are extracted. More details about CCA along with its mathematical formulation are provided in [33]. Using CCA, a new feature space is estimated where similar information from both datasets is preserved. Moreover, in

Table 3
Time delineation features of eye blinking patterns (Fig. 9c).

Symbol	Definition
S	Strength of a blink (max value in a blink pattern) defined as max(B)
D_{on}	Duration of the eye closing phase (blinking onset is defined at the time when the intensity (or area) is 10% of S). D_{on} is the time from $0.1 \times S$ to S.
D_{off}	Duration of the eye opening phase (similar to D_{on}). D_{off} is the time from S to $0.1 \times S$.
θ_{on}	Slope at the onset of a blink defined as $\arctan\left(\frac{S}{D_{on}}\right)$ (the blink pattern is approximated as a triangle)
θ_{off}	Slope at the offset of a blink defined as $\arctan\left(\frac{S}{D_{off}}\right)$ (similar to θ_{on})
A_{on}	Area under the eye blinking curve during the closing phase
A_{off}	Area under the eye blinking curve during the opening phase
E_{on}	Energy of the signal in the closing phase defined as $\sum_{f_0.1\times S}^{f_S} B(f)^2$ (where $\sum_{f_0.1\times S}^{f_S}$ defines the frames within $0.1 \times S$ and S periods)
E_{off}	Energy of the signal in the opening phase (similar to E_{on})

Table 4
List of features extracted from fixations (H, V, and R stands for horizontal, vertical and radial components of a gaze).

Feature	Definition
Fixation duration	Obtained from I-VT result (one feature)
Fixation centroid (HV)	Mean of the gaze during a fixation segment (2 features)
Drift displacement (HV)	The absolute difference between the first and the last gaze in a fixation segment (2 features)
Drift distance (HVR)	Total distance covered during a fixation segment (in pixels for GP3 and in angles for AC - 3 features)
Drift average speed (HVR)	Defined as *drift distance/fixation duration* (3 features)
Dispersion	Dispersion during a fixation segment defined as $\frac{Max(X_f) - Min(X_f)}{Max(Y_f) - Min(Y_f)}$ (one feature)
Distance (or angle) from last fixation (HVR)	Distance (for GP3) or angle (AC) between the centroids of the current and the previous fixation segments (3 features)
Statistical features (HV)	Standard deviation, skewness, and kurtosis of the gaze during a fixation segment (6 features)

Table 5
List of features extracted from saccades (H, V, and R stands for horizontal, vertical and radial components of a gaze).

Feature	Definition
Saccadic duration	Obtained from I-VT result (one feature)
Standard deviation (HV)	Standard deviation of gaze during a saccadic segment (two features)
Drift distance (R)	Total distance covered during a saccadic segment (in pixels for GP3 and in angles for AC - one feature)
Dispersion	As in fixation (one feature)
Saccadic angle (HV - for AC only)	Angle between the first and the last gaze in a saccadic segment (two feature)
Saccdic angle and amplitude (R - for GP3 only)	Obtained from the first and the last gaze in a saccadic segment defined as $\tan^{-1}\left(\frac{Y_f(N_s)-Y_f(1)}{X_f(N_s)-X_f(1)}\right)$ and as $\sqrt{(X_f(N_s)-X_f(1))^2+(Y_f(N_s)-Y_f(1))^2}$, respectively (two features)
Distance (or angle) from last saccade (HVR)	Same as fixation (3 features)
Statistical features from velocity and acceleration (HV)	Mean, median, max, Std, skewness and Kurtosis of the velocity and the acceleration patterns during a saccadic segment (12 features)

the new space, information gain from features is maximized by minimizing the correlation of the canonical variates.
In this paper, the pair of projecting matrices W_X and W_Y are obtained from the training feature matrices of the eye movement and the other modality (blinking or the periocular data). Then, the new pair of projected feature sets $\widehat{X} = W_X^T X$ and $\widehat{Y} = W_Y^T Y$ are concatenated together to form the fused feature set (parallel fusion) before feeding to the classifier. Then, the testing feature set from each modality is projected using the same W_X and W_Y (estimated during training) and then concatenated together.

3.5.2. Score-level fusion

In score-level fusion, separate classifiers are trained using each training feature matrix and the testing score (matching score) from both classifiers are combined to predict the output class (decision). In this paper, three straightforward fusion techniques are employed: sum or average (S_{SUM}), product (S_{PROD}), and max (S_{MAX}). For S_{SUM} and S_{MAX}, the fused score ranges from 0 to 1, however, for S_{PROD}, normalization is required to get the same range.

3.6. Classification

Different classifiers were tested for the proposed system including Linear Discriminant Analysis (LDA), multi-class Support Vector Machines (SVM), and Radial Basis Function (RBF) Neural Networks. However, LDA was the most computationally efficient classifier with similar performance as the multi-class SVM, hence, it is the one used in this paper. LDA models the training feature matrix for each subject (or class c) with a Gaussian distribution that has a mean, μ_c, and a standard deviation, Σ_c. These are estimated using maximum likelihood estimation. LDA also assumes that every class has the same covariance. This is equivalent to projecting the features into a new space that maximizes the inter-class distance and minimizes the intra-class variance [34]. Finally, it uses the optimum Bayes rule to make the classifier decision to maximize the logarithm of the posterior probability given by

$$\log(P(c|x_s)) = -\frac{1}{2}(x_s-\mu_c)^T\Sigma_c^{-1}(x_s-\mu_c) + \log(P(c)) \quad (6)$$

where $c = 1, 2, \ldots, N_s$. In the equation above, $\Sigma_c = \Sigma$ for LDA. N_s is the total number of subjects, where $P(c|x_s)$ is the probability of matching the features from an unknown subject x_s to a class c. $P(c)$ is the prior probability. The equation itself represents a linear equation where the boundaries between the different classes is determined by as straight line.

4. Experimental setup and results

As mentioned in Section 3.1, two databases are collected from two low frame rate eye-tracking devices. Two different eye modality features are proposed and evaluated for improving the eye movement based-biometrics in these devices. Firstly, each modality is evaluated solely in a single modality setup along with the eye movements' performance. Then, these modalities are combined with the eye movement patterns in a multi-modal setup. Moreover, the proposed system is evaluated in both modes of authentication; identification and verification. The Correct Recognition Rate (CRR), which is also known as rank-1 accuracy, is used as a performance metric in identification mode. In verification mode, the Equal Error Rate (EER) is used for performance evaluation.

The system's performance is evaluated using the hold-one-out cross-validation technique. For each subject, one trial is held for building the test pool of features and the remaining trials (2 for the GP3 system and 3 for the AC system) are used for building the training pool. This is repeated multiple runs (3 runs for the GP3 system and 4 runs for the AC system) and in each run, a different trial is selected for testing. Moreover, since eye features are varying in time even for the same subject, multiple feature vectors from each modality, which are selected randomly within the pool, are averaged to form one testing or training example. The random selection is based on sampling without replacement $C(N_t, N_f)$, where N_t is the total number of feature vectors available in the training/testing pools and N_f is the number of the selected feature vectors to average. This is also known as random sub-sampling or boosting which helps in increasing the number of training and testing examples specifically for small-sized databases. Based on the databases recorded, the average number of eye fixations/saccades and blinks per minute are 50 and 15, respectively.

Using random sub-sampling, 500 training and 40 testing examples are generated for each subject. The N_f value reflects the total time required for user authentication (i.e. testing time, T_a). In this paper, the system was evaluated using different values of N_f; for eye movements $N_f = [25,50,75,100]$, and for eye blinking, $N_f = [8,15,23,30]$. Based on our databases, these values represent an average recording time (T_a) of 30, 60, 90, and 120 *s*, respectively. This is the time window required to perform authentication. Only two subjects in the AC database have less than 30 blinks per trial. In this case, their blinking data are doubled by adding random noise with a small variance to their original blinks.

4.1. Single modality evaluation

Table 6 summarizes the results for the eye blinking based biometric system in identification and verification modes for a different number of averaged feature vectors. As mentioned previously, the average number of eye blinks is 15 *blinks/min*. So, N_f values in this table represent an average authentication time (T_a) of 30, 60, 90, and 120 *s*. In the last row, the feature vectors within the whole testing pool trial (5 *min* duration) are averaged to form one testing example. Based on the achieved results, averaging a larger number of eye blinking features provides a better estimate of the eye blinking behavior for each subject, hence, improving the performance of the biometric authentication system. As expected, the best CRR and the lowest EER achieved for the GP3 and the AC systems were achieved using the longest authentication time

Table 6
Eye blinking based biometric system evaluation in single modality (the first and the second row in each N_f value represent CRR and EER in %, respectively).

N_f (T_a)	GP3 system				AC system				
	run = 1	run = 2	run = 3	Mean (std)	run = 1	run = 2	run = 3	run = 4	Mean (std)
8 (30 *s*)	72.27	78.45	76.82	75.85 (3.2)	61.15	68.29	67.62	61.22	64.57 (3.9)
	10.05	9.36	10.36	9.92 (0.5)	11.77	10.87	12.56	11.65	11.71 (0.7)
15 (60 *s*)	80.73	84.73	83	82.82 (2)	67.9	75.04	72.3	67.68	70.73 (3.6)
	8.18	8	9.09	8.42 (0.6)	11.2	8.96	10.3	9.9	10.09 (0.9)
23 (90 *s*)	84.36	90.64	86.36	87.12 (3)	71.2	76.57	74.51	69.84	73.03 (3.1)
	6.55	7.36	8.91	7.61 (1.2)	11.03	8.21	9.57	9.54	9.59 (1.2)
30 (120 *s*)	87.45	93.27	87.82	89.52 (3.3)	72.18	76.78	75.3	72.8	74.27 (2.2)
	5.73	6.91	8.27	6.97 (1.3)	11.07	7.31	9.02	9.49	9.22 (1.5)
all (5 min)	95.45	95.45	90.91	**93.94 (2.6)**	74.29	76.97	77.14	70.29	**74.67 (3.2)**
	0	4.55	9.09	**4.55 (4.5)**	7.6	2.86	6.29	8.57	**6.33 (2.5)**

(5 *min*). For the GP3 system, the best CRR and EER values achieved were 93.94% and 4.55%, respectively, over 22 subjects. For the AC system, the best CRR and EER values achieved were 74.67% and 6.33%, respectively, over 35 subjects (19 subjects without glasses and 16 subjects with glasses). These results were achieved using an authentication time window of 5 *min*.

The results for the eye movement-based biometric system in single modality are provided in Table 7. Again, N_f values in this table represent an average authentication time (T_a) of 30, 60, 90, and 120 *s* (average fixations/saccades rate is 50 per minute). An observation similar to the eye blinking system can be drawn here where averaging features from a larger number of eye movements per subject increases the discrimination ability for the biometric system. In identification mode, the highest CRR achieved was 89.39% and 79.29% for the GP3 and the AC systems, respectively, using all the 5 *min* recording. Similarly, Using the same T_a, an EER as low as 3.03% and 7.14% can be achieved. This test was conducted over all the participants in both databases (22 subjects for the GP3 system and 35 subjects for the AC system). Similar to the eye blinking evaluation, the GP3 system outperforms the AC system because it has higher frame rate and higher gaze estimation accuracy.

Fig. 7 evaluates the static features from the periocular regions extracted from the GP3 and the AC systems in the single modality. In comparison with the dynamic and the behavioral features of eye movements and blinking, this set of static features achieves stable recognition rates at much lower authentication time (frames). As shown in both figures, averaging features from more than 30 frames (about 0.5 *s* for the GP3 system and 2.5 *s* for the AC system) does not affect the CRR or EER rates obtained. Also, it is noticeable that our proposed eye alignment model for the GP3 system is more accurate than the built-in model for the AC system. This is reflected in the higher CRR and the lower EER values achieved for the GP3 system. The GP3 system achieved a consistent performance of CRR = 96.1% and EER = 3%, while the AC system achieved a CRR and EER values of 89.4% and 8.3%, respectively. This test is conducted over all the participants in both databases (22 for the GP3 system and 35 for the AC system).

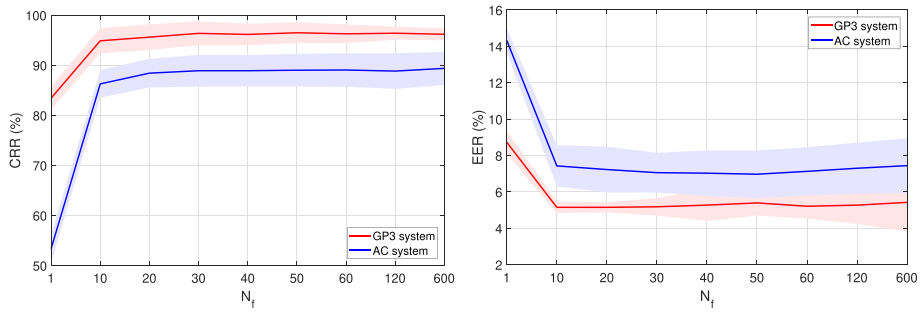
4.2. Multi-modal setup evaluation

Table 8 shows the performance of the multi-modal system combining dynamic eye features (eye movement and blinking) for the five fusion techniques discussed in Section 3.5. Noticeable performance improvement, in the identification and verification modes, is achieved for all the values of N_f in comparison with the single modality setup. For example, using CCA for feature fusion, an EER as low as 0.6% and a CRR as high as 96.67% can be achieved for the GP3 system. Similarly, the highest CRR and the lowest EER achieved for the AC system were 94.71% and 2.14%. In general, fusing eye movement and blinking features shows approximately 7% and 14% improvement in the recognition rates for the GP3 and the AC systems, respectively, compared to the single modality. On the other hand, EER values are lowered by about 2.5–5% in verification mode. Therefore, combining features from both traits makes our proposed system comparable to other biometric systems that use conventional biometric traits like fingerprint [35].

The performance of the multi-modal system combining dynamic eye features (eye movement) and the static periocular features is provided in Table 9. Again, noticeable performance improvement is gained in comparison to the single modality setup based solely on eye movements. The highest CRR values achieved are 98.48 and 94.57% for the GP3 and the AC systems, respectively. In terms of EER, the lowest error rates achieved are 0 and 2.57% for the GP3 and the AC systems, respectively. Another observation is that combining both static and dynamic features lead to high recognition rates at lower authentication times compared with Table 8. For example, acquiring both sets of features for 1 *min* lead to high recognition rates with CRR = 98.16% and EER = 1% using CCA for the GP system over 22 subjects. Similarly, for

Table 7
Eye movement based biometric system evaluation in single modality (the first and the second row in each N_f value represent CRR and EER in %, respectively). Bold signifies Highest CRR and lowest EER.

N_f (T_a)	GP3 system				AC system				
	run = 1	run = 2	run = 3	Mean (std)	run = 1	run = 2	run = 3	run = 4	Mean (std)
25 (30 *s*)	64.73	68.91	71.91	68.52 (3.6)	61.71	75.83	76.91	68	70.61 (7.1)
	11.64	11.09	9.82	10.85 (0.9)	13.26	9.83	11.43	13.14	11.91 (1.6)
50 (60 *s*)	74.09	79.73	83.27	79.03 (4.6)	65.89	80.23	84.51	74.4	76.26 (8.1)
	9.27	9.55	7.64	8.82 (1)	13.14	8.29	8.97	10.69	10.27 (2.2)
75 (90 *s*)	77.64	84.73	87.18	83.18 (5)	67.43	82	85.54	74.97	77.49 (8)
	9	8.09	5.45	7.52 (1.8)	11.77	6.91	9.2	9.54	9.36 (2)
100 (120 *s*)	79.45	87.45	89.09	85.33 (5.2)	68.74	81.26	87.43	76.23	78.41 (7.9)
	8.45	7	5.45	6.97 (1.5)	11.77	5.94	9.31	9.26	9.07 (2.4)
all (5 min)	86.36	90.91	90.91	**89.39 (2.6)**	68.57	80	88.57	80	**79.29 (8.2)**
	4.55	4.55	0	**3.03 (2.6)**	8.57	5.71	5.71	8.57	**7.14 (1.6)**



(a) Identification mode (b) Verification mode

Fig. 7. Periocular static features evaluation in single modality setup (shaded areas shows the standard deviation).

1 *min* authentication time using the AC system, a CRR and EER of values 93.23% and 3.5% can be obtained over a population of 35 subjects.

4.3. Effect of wearing glasses

In this section, we examine the effect of wearing glasses on the performance of the proposed system. For the GP3 system, only 3 participants (out of 22) were wearing glasses, however, for the AC system, 16 users out of 35 ($\approx$ 50%) were wearing glasses. Therefore, this experiment was conducted only for the AC system to provide a fair comparison. It is important to mention here that the AC system can detect glasses, hence, this information can be used to apply different preprocessing steps for users with glasses if required. The database collected by the AC system was split into two categories; users with glasses (G) and users without glasses (NG), and the performance metrics were evaluated for each category separately. The results are summarized in Fig. 8. Based on the achieved results, we can conclude that there is no consistent degradation in the performance due to wearing glasses. For example, eye blinking patterns (after correction as mentioned in Section 3.3.1) from users wearing glasses (G category) achieved approximately similar performance as the NG category in the single and the multi-modal setups. In contrast, the G category outperformed the NG category in single and multi-modal setups for the periocular features. Only the eye movement in single modality showed relatively lower but acceptable performance for users with glass (8% lower in CRR and 3% higher in EER). This issue might be tackled in the future by adding an extra pre-processing step to minimize the gap in the performance between these two classes.

5. Discussion

This section provides a discussion of the proposed system, its applicability and a comparison of the achieved results with previous works.

5.1. Single vs multi-modal setup

Based on the achieved results, it is surprising that the eye blinking based biometric system achieved comparable performance as the one based on eye movements. Again, the reason for this might be the low frame rate of the eye tracking device which is enough to capture blinking patterns (average duration 300 *ms*) but not enough to capture eye movement patterns, specifically, saccades (duration less than 100 ms). This attributes to the lower performance of the eye movement system as most of the features extracted are from saccades. Finally, the set of the static periocular features that characterizes the 2D shape of the eyes outperformed the aforementioned modalities as these features are more robust and time-consistent, however, they are more vulnerable to spoofing attacks. Therefore, combining static and dynamic features enhances accuracy and improves the security of the biometric system.

However, taking into account the small population size of the database, the performance of all the systems in the single-modality setup is considered low in comparison with conventional traits. Despite the degraded performance in the single-modality setup, combining static and dynamic features from the eyes significantly improves the system's performance and making it more precise in the discrimination between the identities of the enrolled users as given in Tables 8 and 9. This contributes to the significant increase in the CRR values with up to 8% higher than eye movements solely in identification mode. Similarly, in verification mode, the EER is lowered by about 3–5%.

5.2. Comparison with previous eye-related biometric systems

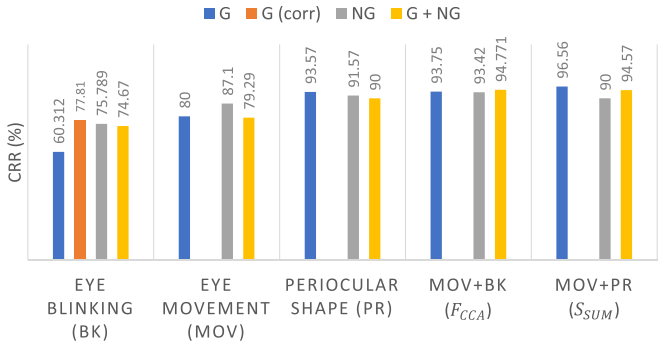
In comparison with the state of the art eye movement biometric systems in [13], our proposed system achieved noticeably lower CRR and higher EER values despite the lower number of enrolled subjects. However, our databases were built with cost-effective eye-tracking devices.

Table 8
Eye movement and eye blinking in multi-modal setup. The first and the second row in each fusion technique represent CRR and EER as mean(std) in %, respectively. Bold signifies Highest CRR and lowest EER.

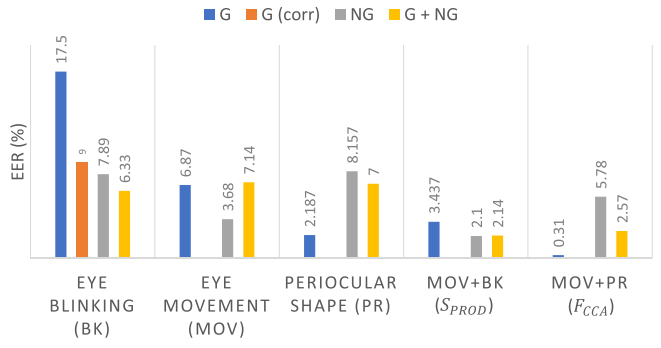
	GP3 system			AC system		
T_a (*min*)	1	2	5	1	2	5
F_{CAT}	90.91 (3.4)	93.36 (5.5)	93.94 (5.2)	92.79 (3.6)	93.73 (4)	94.28 (4)
	3.97 (1)	3.03 (1)	4.54 (0)	4.36 (1.9)	3.61 (1.7)	2.43 (1.8)
F_{CCA}	95.49 (3.9)	96.26 (5)	**96.67 (5)**	92.97 (3.6)	93.63 (4.2)	**94.71 (4.1)**
	2.5 (0.7)	1.96 (0.7)	**0.61 (1)**	4.35 (1.8)	3.65 (1.6)	2.71 (1.6)
S_{SUM}	90.3 (2.5)	92.7 (2.8)	92.42 (2.6)	83 (7.6)	86.03 (8)	87.86 (10.3)
	3.89 (0.7)	2.85 (0.6)	1.14 (2)	5.26 (1.4)	4.96 (1.7)	2.57 (1.9)
S_{PROD}	93.03 (2.8)	95.12 (3.8)	95.45 (4.5)	92.75 (3.6)	93.72 (3.9)	94.28 (4)
	3.55 (0.6)	2.73 (0.7)	1.52 (2.6)	4.37 (1.2)	3.9 (1)	**2.14 (1.1)**
S_{MAX}	90 (2.3)	93.03 (2.4)	92.42 (2.6)	83.28 (7.8)	85.65 (8.4)	87.86 (10.3)
	4.55 (0.8)	3.18 (0.6)	2.9 (2.5)	7.15 (0.9)	6.07 (1.1)	3.29 (1.7)

Table 9
Eye movement and periocular static features in multi-modal setup. The first and the second row in each fusion technique represent CRR and EER as mean(std) in %, respectively. Bold signifies Highest CRR and lowest EER.

	GP3 system			AC system		
T_a (*min*)	1	2	5	1	2	5
F_{CAT}	98.48 (2.5)	98.48 (2.6)	**98.48 (2.6)**	92.75 (3.7)	92.82 (2.8)	93.57 (3.6)
	1.6 (0.7)	1.68 (1.2)	**0 (0)**	4.63 (1.3)	4.4 (1.9)	3.3 (1.7)
F_{CCA}	98.16 (0.6)	98.11 (1.7)	**98.48 (1.4)**	89.38 (3.3)	89.42 (3.3)	89.57 (3.3)
	1.07 (0.4)	1 (0.2)	0.3 (0.5)	3.5 (0.7)	2.87 (1)	**2.57 (1.8)**
S_{SUM}	97.82 (2.9)	97.36 (3.8)	97.58 (3.4)	93.23 (2.5)	93.96 (2.2)	**94.57 (1.9)**
	2.7 (0.2)	2.65 (0.2)	0.3 (0.5)	4.14 (0.2)	3.82 (0.6)	3.45 (1.5)
S_{PROD}	98.12 (2.5)	98.61 (2.4)	**98.48 (2.6)**	92.7 (3.5)	92.58 (3.4)	92.71 (3.8)
	1.62 (0.6)	1.67 (1.4)	**0 (0)**	5 (0.9)	5.57 (0.4)	4 (1.2)
S_{MAX}	97.81 (2.9)	97.35 (3.8)	97.58 (3.4)	93.16 (2.6)	93.27 (1.5)	93.86 (0.9)
	3.18 (0.06)	3.14 (0.15)	0.3 (0.5)	5.31 (0.6)	5.08 (0.9)	6.34 (0.7)



(a) Identification mode



(b) Verification mode

Fig. 8. Effect of wearing glasses on the performance of the proposed system using the AC system in both modes of authentication (BK: eye blinking, MOV: eye movement, PR: periocular features).

These devices have a significantly lower frame rate compared to the systems used previously which operate at 1000 fame/s. Moreover, the database in [13] was built in a controlled environment where subjects have to rest their head on a chin-rest and follow a randomly moving target. Our databases were recorded while watching driving sessions which provides a more realistic scenario specifically for the application of driver authentication.

Regarding eyeblinking-based biometrics, according to our knowledge, remote registration of eye blinking was adopted for biometric authentication only by J. Espinosa et al. [9]. In comparison, our system shows lower correct recognition rates. In [9], the authors claimed to achieve a CRR of 99.7% by averaging features from 25 randomly selected eye blinks, however, in their experimental setup, they did not separate the data into training and testing before averaging, so, same original eye blinks contributed to the averaged blinks in the training and testing data. Moreover, similar to the eye movement recording procedure, they registered eye blinking patterns in a controlled environment with a higher frame rate camera (250 frames/s).

Comparing with our previous multi-modal setup based on eye movements and eye blinking in [8], adopting a new technique for eye blinking behavior characterization based on the eye alignment showed a noticeable improvement in the eye blinking based biometric setup with about 6% improvement in the identification mode (3% lower EER in the verification mode). The proposed technique in this paper mitigates the variations in the lighting conditions which alters the amount of light reflected back to the IR camera during blinking. Moreover, proposing a new set of periocular static features for the multi-modal setup shows about 2.5% improvement in the identification mode with even shorter authentication time (1.1% lower EER in the verification mode).

5.3. Comparison with previous driver authentication systems

As mentioned in Section 2.2, the main approach for driver authentication is to leverage the driving style assessment to human identity recognition. Features to assess the driving behavior like steering wheel angle, brake/gas pedal status, acceleration status, and vehicle speed were extracted from the CAN setup in the vehicle. While this approach does not require additional setup (i.e. for data acquisition), these modalities do not carry subject discriminative features. Although they can estimate the driving style, they are highly dependent on external conditions like the traffic/road status and weather conditions. For instance, the driving behavior of the same driver is significantly different in normal versus rush hours. In terms of performance, in general, most of the previous works evaluated this approach over a relatively small population size of 10–15 drivers with an average identification accuracy of $\approx$ 90% and an EER as low as 2% [23–25]. The best identification accuracy achieved was 98% using automatic feature learning with CNN over a population of 15 drives [1].

In contrast, our proposed system requires a simple setup of an IR camera on the dashboard to register different modalities from the eye like eye movements, eye blinking, and periocular features. These traits showed reliable performance in a practical setup when they are fused in a multi-modal system. In this paper, our proposed approach showed significant improvement in terms of CRR and EER values even under a database with larger population size (Table 10).

5.4. Applicability and limitations

While the achieved results in this paper are promising, the proposed system is still in the first phase as a proof of concept and many factors need to be considered to obtain reliable performance in a real-world scenario. First of all, averaging features from multiple

Table 10
Comparison between different approaches for driver authentication system (I: identification, V: Verification).

Paper	Modality	# of Participants	Mode	Performance (%)
[23]	Following-distance patterns	12	I	CRR = 89.6
[24]	- Acceleration patterns	10	I	CRR = 93.75
	- Brake pedal pressure		V	EER = 2.5
	- Their derivatives			
[25]	- Pedal pressure (acc/brake)	10	V	EER = 14.7
	- Vehicle speed			
	- Steering wheel angle			
	- Traveling distance			
[1]	CNN feature maps of driver fingerprinting data	15	I	CRR = 98
Ours	- Eye Movements (MOV)	22 (GP3)	I	MOV + BK: 96.7 (GP3) - 94.7 (AC)
	- Eye Blinking (BK)	35 (AC)		MOV + PR: 98.5 (GP3) - 94.6 (AC)
	- Periocular Shape (PR)		V	MOV + BK: 0.6 (GP3) - 2.1 (AC)
	- Fusion of modalities			MOV + PR: 0 (GP3) - 2.6 (AC)

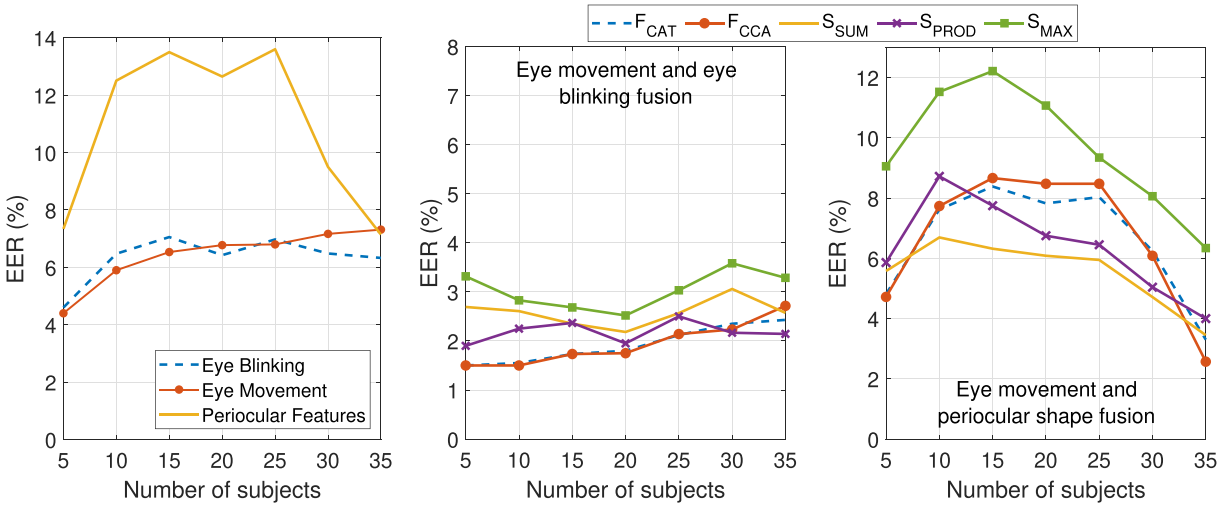
patterns (eye blinks or eye movement) is essential to achieve robust performance. In this paper, the authentication time is estimated from the average blinking and fixation/saccade rate per minute from the collected databases. However, these rates are not consistent even for the same subject as they depend on the human body conditions like fatigue, sleepiness or distraction. All these factors need to be addressed in future works as they also affect the behavior of the eye blinking and movement patterns not only the rate. Also, for a driver authentication application, lighting conditions vary from day-time to night-time which affects the eye-tracking system, hence, affect the extracted features. All these factors need to be addressed in the future by recording multiple sessions from the same subject under different human and lighting conditions.

Another limitation of this study is the population size of the collected databases. Increasing the number of enrolled subjects is expected to have a negative impact on the system's performance. This might be valid if the system operates in identification mode. As in identification mode, increasing the number of subjects increases the complexity of the feature space, and hence increases the difficulty of the classification task which leads to a decrease in the CRR. However, for verification, which is the practical mode for the driver authentication application, a separate classifier model is added for each user enrolled in the system. Additionally, enrolling more subjects increases the confidence of the system for imposter detection.

To proof our claim, we have run an experiment to assess the effect of increasing the number of enrolled users on the Equal Error Rate (EER). For this experiment, we used the data from the AC system only since it has a larger population size. We increased the number of enrolled subjects in the system by a step of 5 and evaluated the EER in each step. At each step, the subjects' IDs were randomly selected from the whole database. This is repeated 10 times and the average EER is reported. The findings from this test are illustrated in Fig. 9. As shown in the figure, EER does not show any significant increase as the number of the enrolled subjects increases above 15, in contrast, the EER decreases in some cases.

6. Conclusion and future work

In this paper, two sets of dynamic and static features extracted from the eyes are proposed to improve the performance of eye movement biometric systems based on low frame rate eye-tracking devices. Based on the achieved results, we can conclude that eye movement behavioral patterns extracted from such devices still carry subject-discriminating features, however, they achieve an under-performing recognition rate taking into account the small population size. Moreover, it was shown that other dynamic and static features from the eye like the eye blinking patterns and periocular 2D shape carry subject-unique features. By combining these modalities with eye movements, a noticeable improvement in the recognition performance was



(a) Single modality (b) Fusion of two modalities

Fig. 9. Effect of the number of enrolled subject on system's performance in verification mode.

achieved. These rates are considered reliable for various biometric applications like continuous driver authentication system in smart cars.

Future work can be extended in many directions. The first extension would be transferring the current setup to a real car with the IR camera facing the driver's face. This setup has many challenging issues to be addressed like various lighting conditions (day and night light) and increased head movements during real driving. This can be tackled by using more accurate eye-tracking devices with higher frame rates. The GP3 and the AC systems have a very low frame rate compared to the other cameras used in similar work [13]. This attributed to the reduced recognition rates in our experiments which could have been avoidedespecially for the eye movement modality. This affects the saccadic patterns specifically which need eye-tracking devices with a minimum frame rate of 200 *Hz*. Moreover, in this study, although the training and testing data came from different recordings, these recordings are conducted in the same session. However, to investigate the time-permanence of these patterns and different human conditions, multiple sessions should be recorded from the same subject. Finally, there is still a room for improving the eye movements-based biometrics by exploring new feature extraction techniques that can precisely characterize the neurological signals of the brain that controls the extraocular muscles controlling the eye movements [5,31].

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors would like to thank the Natural Sciences and Engineering Research Council (NSERC) of Canada, Alcohol Countermeasure Systems (ACS) Corp., and the Royal Bank of Canada (RBC) for their support (Grant numbers 503492 and 503493). Also, the authors would like to thank the participants who agreed to do the experiment.

References

[1] Yijie Xun, Jiajia Liu, Nei Kato, Yongqiang Fang, Yanning Zhang, Automobile driver fingerprinting: a new machine learning based authentication scheme, IEEE Transactions on Industrial Informatics 16 (2) (2019) 1417–1426.
[2] Javier Galbally, Sébastien Marcel, Julian Fierrez, Biometric antispoofing methods: a survey in face recognition, IEEE Access 2 (2014) 1530–1552.
[3] J. Galbally-Herrero, J. Fierrez-Aguilar, J.D. Rodriguez-Gonzalez, Fernando Alonso-Fernandez, Javier Ortega-Garcia, and M Tapiador. On the vulnerability of fingerprint verification systems to fake fingerprints attacks, *Carnahan Conferences Security Technology, Proceedings* 2006 *40th Annual IEEE International*, IEEE 2006, pp. 130–136.
[4] Chiara Galdi, Michele Nappi, Eye movement analysis in biometrics, Biometrics under Biomedical Considerations, Springer 2019, pp. 171–183.
[5] Ioannis Rigas, Lee Friedman, Oleg Komogortsev, Study of an extensive set of eye movement features: Extraction methods and statistical analysis, Journal of Eye Movement Research 11 (1) (2018).
[6] Keith Rayner, Eye Movements and Visual Cognition: Scene Perception and Reading, Springer Science & Business Media 2012.
[7] Panos Konstantopoulos, Peter Chapman, David Crundall, Driver's visual attention as a function of driving experience and visibility. using a driving simulator to explore drivers' eye movements in day, night and rain driving, Accident Analysis & Prevention 42 (3) (2010) 827–834.
[8] Sherif Seha, Georgios Papangelakis, Dimitrios Hatzinakos, Ali Shahidi Zandi, Felix J.E. Comeau, Improving eye movement biometrics using remote registration of eye blinking patterns, *ICASSP* 2019–*2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE 2019, pp. 2562–2566.
[9] Julián Espinosa, Begoña Domenech, Carmen Vázquez, Jorge Pérez, David Mas, Blinking characterization from high speed video records. application to biometric authentication, PloS One 13 (5) (2018) e0196125.
[10] Unsang Park, Raghavender Reddy Jillela, Arun Ross, and Anil K Jain. Periocular biometrics in the visible spectrum, IEEE Transactions on Information Forensics and Security 6 (1) (2010) 96–106.
[11] Oleg V. Komogortsev, Ioannis Rigas Bioeye, Competition on biometrics via eye movements, Biometrics Theory, Applications and Systems (BTAS), 2015 IEEE 7th International Conference on, IEEE 2015, pp. 1–8.
[12] Ioannis Rigas, Oleg V. Komogortsev, Current research in eye movement biometrics: An analysis based on BioEye 2015 competition, Image and Vision Computing 58 (2017) 129–141.
[13] Anjith George, Aurobinda Routray, A score level fusion method for eye movement biometrics, Pattern Recogn. Lett. 82 (2016) 207–215.
[14] Narishige Abe, Shigefumi Yamada, Takashi Shinzaki, A novel local feature for eye movement authentication, 2016 *International Conference of the Biometrics Special Interest Group (BIOSIG)*, IEEE 2016, pp. 1–5.
[15] Pawel Kasprowski and Katarzyna Harezlak. Using dissimilarity matrix for eye movement biometrics with a jumping point experiment. In *Intelligent Decision Technologies* 2016, pages 83–93. Springer, 2016.
[16] Pawel Kasprowski and Jozef Ober. Eye movements in biometrics. In *International Workshop on Biometric Authentication*, pages 248–258. Springer, 2004.
[17] Sherif N Abbas and M Abo-Zahhad. Eye blinking EOG signals as biometrics. In *Biometric Security and Privacy*, pages 121–140. Springer, 2017.
[18] Mohammed Abo-Zahhad, Sabah M. Ahmed, Sherif N. Abbas, A novel biometric approach for human identification and verification using eye blinking signal, IEEE Signal Processing Letters 22 (7) (2014) 876–880.
[19] Mohammed Abo-Zahhad, Sabah M. Ahmed, Sherif N. Abbas, A new multi-level approach to EEG based human authentication using eye blinking, Pattern Recognition Letters 82 (2016) 216–225.
[20] Qunjian Wu, Ying Zeng, Chi Zhang, Li Tong, Bin Yan, An EEG-based person authentication system with open-set capability combining eye blinking signals, Sensors 18 (2) (2018) 335.
[21] Santosh Tirunagari, Norman Poh, David Windridge, Aamo Iorliam, Nik Suki, Anthony T.S. Ho, Detection of face spoofing using visual dynamics, IEEE Transactions on Information Forensics and Security 10 (4) (2015) 762–777.
[22] Saad Ezzini, Ismail Berrada, Mounir Ghogho, Who is behind the wheel? driver identification and fingerprinting, Journal of Big Data 5 (1) (2018).
[23] Chiyomi Miyajima, Yoshihiro Nishiwaki, Koji Ozawa, Toshihiro Wakita, Katsunobu Itou, Kazuya Takeda, Fumitada Itakura, Driver modeling based on driving behavior and its evaluation in driver identification, Proc. IEEE 95 (2) (2007) 427–437.
[24] Abdul Wahab, Chai Quek, Chin Keong Tan, Kazuya Takeda, Driving profile modeling and recognition based on soft computing approach, IEEE Transactions on Neural Networks 20 (4) (2009) 563–582.
[25] Angela Burton, Tapan Parikh, Shannon Mascarenhas, Jue Zhang, Jonathan Voris, N. Sertac Artan, Wenjia Li, Driver identification and authentication with active behavior modeling, 2016 *12th International Conference on Network and Service Management (CNSM)*, IEEE 2016, pp. 388–393.
[26] Tereza Soukupova, Jan Cech, Eye blink detection using facial landmarks, 21st computer vision winter workshop, Rimske Toplice, Slovenia, 2016.
[27] Vahid Kazemi, Josephine Sullivan, One millisecond face alignment with an ensemble of regression trees, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2014, pp. 1867–1874.
[28] Davis E. King, Dlib-ml: a machine learning toolkit, J. Mach. Learn. Res. 10 (2009) 1755–1758.
[29] Anneli Olsen, The Tobii I-VT Fixation Filter: Algorithm Description, Available Online: https://www.tobiipro.com/siteassets/tobii-pro/learn-and-support/analyze/how-do-we-classify-eye-movements/tobii-pro-i-vt-fixation-filter.pdf/?v=2012. (Accessed 23 August 2018).
[30] Dario D Salvucci and Joseph H Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the* 2000 *symposium on Eye tracking research & applications*, pages 71–78. ACM, 2000.
[31] Ioannis Rigas, Oleg Komogortsev, Reza Shadmehr, Biometric recognition via eye movements: Saccadic vigor and acceleration cues, ACM Transactions on Applied Perception (TAP), 13(2), , 2016.
[32] Mohammad Haghighat, Mohamed Abdel-Mottaleb, Wadee Alhalabi, Fully automatic face normalization and single sample face recognition in unconstrained environments, Expert Syst. Appl. 47 (2016) 23–34.
[33] Quan-Sen Sun, Sheng-Gen Zeng, Yan Liu, Pheng-Ann Heng, De-Shen Xia, A new method of feature fusion and its application in image recognition, Pattern Recogn. 38 (12) (2005) 2437–2448.
[34] Sergios Theodoridis, Konstantinos Koutroumbas, Chapter 2 - classifiers based on bayes decision theory, in: Sergios Theodoridis, Konstantinos Koutroumbas (Eds.), Pattern Recognition (Fourth Edition), Academic Press, Boston 2009, pp. 13–89 , fourth edition.
[35] Raaele Cappelli, Matteo Ferrara, Annalisa Franco, Davide Maltoni, Fingerprint verification competition 2006, Bio- metric Technology Today 15 (7–8) (2007) 7–9.