Automatic Detection of Driver Impairment Based on Pupillary Light Reflex

Alessandro Amodio, Michele Ermidoro, Davide Maggi, Simone Formentin, Member, IEEE, and Sergio Matteo Savaresi, Senior Member, IEEE

Abstract—The main objective of this paper is to determine the feasibility of designing a driver drunkenness detection system based on the dynamic analysis of a subject’s pupillary light reflex (PLR). This involuntary reaction is widely utilized in the medical field to diagnose a variety of diseases, and in this paper, the effectiveness of such a method to reveal an impairment condition due to alcohol abuse is evaluated. The test method consists in applying a light stimulus to one eye of the subject and to capture the dynamics of constriction of both eyes; for extracting the pupil size profiles from the video sequences, a two-step methodology is described, where in the first phase, the iris/pupil search within the image is performed, and in the second phase, the image is cropped to perform pupil detection on a smaller image to improve time efficiency. The undesired pupil dynamics arising in the PLR are defined and evaluated; a spontaneous oscillation of the pupil diameter is observed in the range [0, 2] Hz and the accommodation reflex causes pupil constriction of about 10% of the iris diameter. A database of pupillary light responses is acquired on different subjects in baseline condition and after alcohol consumption, and for each one, a first-order model is identified. A set of features is introduced to compare the two populations of responses and is used to design a support vector machine classifier to discriminate between “Sober” and “Drunk” states.

Index Terms—ADAS, system identification, video processing, pupil dynamics, classification, support vector machine.

I. INTRODUCTION AND MOTIVATION

Road accidents are recognized to be among the major causes of human death in the world; according to the World Health Organization,1 in 2015 road accidents were among the top-10 causes of death, and the only one which is not a disease.

The causes of road accidents are manifold; among them, one of the most important is certainly “drowsiness”, often also called “fatigue” or “sleepiness”. Sleepiness can be defined as the neuro-biological need for sleep, while fatigue is associated with physical labor; although the causes of fatigue and sleepiness may be different, their effects on driving performance are very similar.

Several in-depth studies [1], [2] revealed the impact of drowsiness on road accidents to be around 10% – 20% of the total; although about 84% involves cars and the remaining light trucks and heavy goods vehicles [1], the number of fatal road accidents involving lorries is almost 2 times that for cars [3]. In order to respond to this need for safety, car manufacturers are bringing to the market the class of the so-called Advanced Driver Assistance Systems (ADAS, [4]; examples are the Adaptive Cruise Control [5], the Road Departure Avoidance [6], the and the Lane Change Decision Aid Systems (LCDAS, [7]) and the Driver Drowsiness Detection Systems (DDDS).

The DDDSs are in-vehicle systems that monitor driver and/or vehicle behavior, evaluate some performance indices, and provide alerts or stimulation if the driver seems to be impaired, i.e. drowsy [8]. A number of solutions, with different methods have been proposed so far [9]; a classification based on the method divides them into three categories:

1) methods based on bio-medical signals, like cerebral, muscular and cardiovascular activity;
2) methods based on driver behavior, which evaluate variations in the lateral position of the vehicle, in the velocity, in the steering wheel angle and in other signals recorded;
3) methods based on driver visual analysis using image processing techniques.

The first category [10]–[14] currently requires electrodes attached to the driver’s body, which makes it not commercially viable unless ways are found to measure such body signals without electrodes being directly attached to the body.

The advantage of the second approach is that the useful signals are much easier to acquire and this makes it the most widely used in the commercial market. On the other hand, there are several limitations connected to vehicle type, driver experience, geometric characteristics, condition of the road, etc. These techniques usually require a considerable amount of time to analyze user behavior and don’t perform well in case of the so-called micro-sleeps, that is when a drowsy driver falls asleep for a few seconds on a very straight road section without changing the vehicle signals [15].

The main advantage of computer vision-based techniques is in its non-intrusive nature for monitoring driver’s sleepiness through images acquired with cameras placed in front of the user. The idea at the basis of the third method is to exploit

---

the clear connection between the occurrence of sleepiness and the driver’s face appearance and head/eyes activity; in particular, the most used methods rely on eye closure detection (and then eye closure duration or eye closure rate), yawning detection or nodding detection through head movement. A sub-classification of these methods can be made based on the type of camera:
- methods based on visible spectrum;
- methods based on infrared (IR) camera;
- methods based on stereo camera.

Although being very limited in terms of commercial products, wide interest has been dedicated so far to these techniques in the scientific literature; among the latest works, Zhang et al. [16] employ infrared video imagery to which methods based on Convolutional Neural Networks (CNN) are applied to reveal eye closure. Then, the PERCLOS index and the blink frequence are employed to reveal drowsiness state of the driver. Rigane et al. [17] reveal eye closure with a Neural Network and apply fuzzy logic to the PERCLOS index, eye closure duration and percentage of oriented head to detect driver’s drowsiness. Artranto et al. [18] employ electromyography to monitor eyelid muscles and measure duration of eyelid closure, then starts an alarm when it exceeds a limit; Zhao et al. [19] extract from a video sequence landmarks and textures of eye and mouth regions, to which a Deep Belief Network (DBN) is applied to reveal driver’s drowsiness. Reddy et al. [20] utilize a Multi-Task Cascaded Convolutional Network (MTCNN) for face detection and propose three Neural Network-based solutions to perform 3-class classification. Mandal et al. [21] propose a modular fatigue detection system for bus driver monitoring exploiting a framework that integrates head-shoulder detection, multi-pose face detection, multi-model eye detection, eye openness estimation, fusion and PERCLOS estimation. Pradhan et al. [22] show that eye closure detection is feasible with nearly 98% accuracy even in presence of eye rotation and slight variations in light intensity. Finally, Von Jan et al. [23] present an interesting customer survey explaining the motivations that lead to the development of fatigue detection systems, together with an insight into a prediction algorithm based on blinking status and fatigue indicators.

To the authors knowledge, the only implementation on a commercial vehicle of a DDDS that exploits driver visual analysis is represented by the DS Driver Attention Monitoring System on the DS7 Crossback.

This paper considers the third of the abovementioned methods and proposes a feasibility analysis for a system that is, as a matter of fact, a variant of the classic Driver Drowsiness Detection Systems; instead of detecting the driver drowsiness, the goal is here to detect a different form of impairment, that is the state of drunkenness of the driver. This state is easier to quantify than drowsiness, as it needs a simple alcohol tester, and has very similar effects on the driving capabilities, compared to drowsiness [24]. Moreover, the impact of alcohol abuse on road accidents is significant, as highlighted by a number of studies [25]; according to [2], the percentage of accidents related to alcohol is more than doubled with respect to the ones related to fatigue. Finally, alcohol consumption is well known to cause driving performance degradation mainly through reaction time increase [26].

The analysis of the involuntary reflex of the human pupil reacting to an increase of light condition is called Pupillary Light Reflex (PLR) and is widely utilized in the medical field as a diagnostic tool for a variety of possible lesions and impairment conditions [27]. Pupil’s dynamics have been already investigated in the scientific literature, and successfully used on different applications; Villalobos-Castaldi and Suaste-Gómez [28] demonstrate that spontaneous pupil diameter oscillations have the capability to provide enough discriminative information to authenticate the identity of a subject. Czajka [29] proposes a methodology for eye liveness detection based on the Kohn and Clynès pupil dynamics model; each observation is mapped into a feature space defined by the model parameters and a Support Vector Machine (SVM) classifier is utilized to recognize the behavior of a living eye.

This paper proposes a methodology to analyze the dynamic behavior of the human pupil in order to recognize alcohol abuse. Before the start of the drive, a drunkenness check is performed on the driver, who is requested to correctly pose in front of the acquisition equipment; in this way, the subject is not distracted by the test while driving, and the equipment is able to acquire a clean video sequence in standard pose. Then, in case of positive result of the drunkenness check, a simple warning is issued and the engine start is aborted.

In this paper, the effects of high values of Breath Alcohol Concentration (BrAC) on the Pupillary Light Reflex (PLR) are investigated by applying stepwise light stimuli to one eye of different individuals in a controlled environment, and by recording the pupillary response of both (illuminated and not illuminated) eyes. Several tests are performed in baseline condition and after alcohol intake, in order to acquire a statistically relevant database that allows to discriminate the subject status. To the best of the authors knowledge, this is the first work presenting such an analysis.

The contribution of this paper is threefold. First, a time-efficient methodology is proposed for processing video sequences and extract a time profile of the pupil diameter, consisting in two steps. In the first one, the region of the image containing the eye is detected and, in parallel, iris size is measured until detection stabilization; in the second step, the image is cropped to contain only the pupil, which allows for a more time-efficient pupil size measurement. The cropped region position is constantly adapted in order to compensate for undesired eye/face movements.

Second, an analysis on different subjects is performed to observe the pupil size variation in absence of external stimulation; in this analysis, the typical phenomena that are involved in the pupil dynamics in absence of external stimulation are observed and quantified and some considerations on the intersubject variability of the static values of the pupil size are made.

Third, the paper assesses the problem of determining whether it is feasible to detect the drunkenness state of a subject based on the dynamic characteristics of his/her pupillary light response. In order to perform such study, a light stimulus is applied to one eye of different subjects and the pupillary
response of both eyes is recorded through video cameras. Experimental data are collected from different subjects in baseline condition and after alcohol assumption; for each observation, a simple model describing the pupil constriction dynamics is identified, whose parameters are considered as features. Finally, a SVM-based classifier is designed on such features to discriminate between “Sober” and “Drunk” subjects, obtaining a maximum misclassification rate below 10%, where also 75% of the cases is below 5%.

The paper is organized as follows: Section II describes the proposed algorithm for extracting from the video sequence the time profile of the pupil diameter, together with the experimental set-up. Section III introduces the system model and discusses the main phenomena that affect pupil’s dynamics; among them, the ones that act as a disturbance to the analysis the pupillary light reflex are isolated and quantified in Section IV. Then, Section V describes the tests performed to analyze the correlation between BrAC and pupillary light response and the results are discussed in Section VI; finally, Section VII presents the final conclusions of this study.

II. VIDEO ACQUISITION AND PROCESSING

In this Section, the technique for extracting the time profile of the pupil’s dynamics from a video sequence is presented. At first, the analysis method for the single image is described and then the methodology for processing the entire video sequence is explained. Finally, the details of the experimental set-up are given.

A. Single Image Analysis

The problem of detecting iris and pupil within an image has been widely investigated in the scientific literature [30]–[35]. The most utilized techniques involve the use of the Circular Hough Transform (CHT, [36]–[38]): this is an efficient method for detecting circles within an image that is also utilized in this paper. Supposing the radius of the circle to be known, the method determines the position of the circle’s center and the main steps can be summarized in the following.

1) Candidate pixels that are supposed to lie on the circle are selected as the high-gradient ones within the image and are thus allowed to “cast votes” (Fig. 1a, small yellow circles) to select the center.

2) Each candidate votes every pixel that lies at a distance given by the pre-determined radius (Fig. 1b, red dashed circles).

3) The pixel where the most votes accumulate is the selected center (Fig. 1b, white circle).

In the proposed solution, the original image is first cropped to isolate the eye of interest from the rest of the image; this operation is sufficiently robust thanks to the face support (item 1. in Fig. 3), that ensures correct positioning at each test, guaranteeing that the eye always lies in the same section of the image.

In this context, the pupil’s radius is unknown, and determining it is an objective of the algorithm; for this reason, instead of a single radius value, a range of values is given as input to the algorithm; for each radius value, the algorithm the searches a circle with such radius within the image, returning the one whose center received most votes.

B. Video Analysis

The video-processing algorithm can be divided into two steps, as shown in Fig. 2:

1) Pupil Localization The algorithm locates the iris and pupil within the cropped image using the method described in Section II-A and the two radius ranges are, in this step, defined a-priori (\([r_{P,min}, r_{P,max}]\) and \([r_{I,min}, r_{I,max}]\)). This is repeated for each video frame until the position and size of the iris and pupil are stabilized.
2) **High Speed Processing** The image is once more cropped to work only on the image section containing the pupil: by doing so, the algorithm stops detecting the iris and supposes its radius to remain constant to speed up the image processing and work on a smaller image. The crop area within the image is adaptively selected by setting its center to coincide with the center of the pupil’s circle detected at the previous step: this guarantees the pupil to be always centered within the cropped image. The radius range for pupil detection is also adapted at each frame, by considering the pupil’s radius at the previous step. If we denote by $r_w$ the pupil range width (which is a parameter of the algorithm) and the pupil radius at frame $k$ by $r(k)$, then the radius range at frame $k$ is computed as follows:

$$[r_{\min}(k), r_{\max}(k)] = [r(k-1) - \frac{r_w}{2}, r(k-1) + \frac{r_w}{2}]$$

(1)

C. Experimental Set-Up

The experimental set-up is schematically represented in Fig. 3; the main components are summarized in the following:

1) Face support
2) Separator; this object has the goal of isolating the left eye from the visible light
3) Visible light LED
4) Infrared light LED
5) Visible-spectrum camera
6) Infrared filter
7) Full-spectrum camera
8) Board for video acquisition and LEDs control

The experiments consist in applying a visible-light stimulus (item 3.) to only one of the two eyes to cause the pupillary reflex; this reaction is captured with a high-definition camera at 100 fps, sensible to the spectrum of visible light (item 5.). The other eye is isolated from the light source thanks to the separator (item 2.) in order to minimize its reaction to a light stimulus; since this eye is not illuminated, the pupil movements are recorded with a full-spectrum camera at 100 fps, equipped with an infrared filter to enhance the pupil visibility within the images. An infrared LED (item 4.) illuminates this eye to improve video quality; the two LEDs light up at the same instant to allow for synchronization of the two videos, as described in Section V-B. An electronic board (item 8.) acts as control unit to drive the LEDs and control video acquisition; finally, the subject’s eyes are kept in place thanks to a face support (1.). The instrumentation is located in a room with no ambient light, which means that the two LEDs (visible and infrared) are the only light sources.

III. PUPIL DYNAMICS AND MODELING

This Section describes the various phenomena that affect the pupil dynamics and proposes a simple model that focuses only on the dynamics of interest.

The pupil is a hole located in the center of the iris; it allows light to strike the retina and appears black mainly because the light rays entering the pupil are absorbed by the tissues inside the eye. Within the eye, the pupil is the element that is responsible for the regulation of the light reaching the retina: it constricts when the light amount increases and, vice versa, dilates in case of low-light condition.

The causes of a pupillary diameter variation are manifold; one cause is the so-called Hippus, or spontaneous oscillations of the pupil size, which is a normal and persistent behavior in every subject, occurring without any stimulation. Other causes are the involuntary reflex reactions, such as the Accomodation Reflex or the Light Reflex; the first is a reflex action of the eye, in response to focusing on a near object or looking at a distant one, causing constriction or dilation of the pupil respectively. The Light Reflex can be Direct or Consensual; the Direct Light Reflex is responsible for the constriction of the pupil due to a light stimulus. In case of healthy subject, the sizes of the two pupils are always kept at the same value by the Consensual Light Reflex, responsible of the constriction of one pupil, when the other one is illuminated and subject to Direct Light Reflex. In other words, when a light stimulus is applied to one eye, its pupil constricts due to the Direct Light Reflex, while the other pupil constricts due to Consensual Light Reflex. A pupil size variation can also be induced by other factors, such as emotional stimuli, i.e. by recalling a particularly emotional event [39] or by subliminal auditory stimuli [40].

This study focuses on inducing a Direct and Consensual Light Reflex on a subject by applying a stepwise light stimulus on one of the two eyes; the reactions of the two eyes in normal condition are compared to the reactions when the subject is impaired due to alcohol assumption. No side or oblique views will be considered in this work.

Several models can be found in literature for describing the dynamics of the pupillary reflex [41]; one very simple model and among the most popular ones is the Kohn and Clynes model [42] shown in Fig. 4. In the model, the input $x$ represents the light hitting the retina of the illuminated eye and the output $y$ stands for the observed eye’s pupil diameter; when the light amount increases, $x > 0$ holds and vice versa when the light decreases. Notice that the constants $K_r$ and $K_i$...
This allows to consider a simplified gray-box first light stimulus by truncating the response after the first instants.

oscillations are isolated from the pupil constriction due to the processing level, the slow dynamics and the spontaneous transfer function with delay.

positive and negative light stimuli, modeled as a first order term and persistent changes in pupil size, active for both stimuli. The lower channel, on the other hand, models “long-real poles and a delay, acting only in case of positive light behavior of the pupil as a second order transfer function depending on the sign of $x$ decreases.

are positive, so that for a positive light step, the pupil diameter decreases.

This model features an asymmetry in pupil response depending on the sign of $x$; the upper channel models a “fast” behavior of the pupil as a second order transfer function with real poles and a delay, acting only in case of positive light stimuli. The lower channel, on the other hand, models “long-term” and persistent changes in pupil size, active for both positive and negative light stimuli, modeled as a first order transfer function with delay.

As will be discussed in the following chapter, at data-processing level, the slow dynamics and the spontaneous oscillations are isolated from the pupil constriction due to the light stimulus by truncating the response after the first instants. This allows to consider a simplified gray-box 1st-order model to describe only the “fast” dynamics between the variables $x$ and $y$ of Fig. 7, which can be represented by the following transfer function:

$$G(s) = \frac{\mu e^{-\tau s}}{1 + Ts} \tag{2}$$

The model’s parameters, with the relative constraints are described in the following:

- $\mu \leq 0$: Gain
- $\tau \geq 0$: Delay
- $T \geq 0$: Time constant

Every model belonging to this class can be associated to the row vector of parameters $\phi = [\mu, \tau, T]$, and can be seen as a single point in a three-dimensional space. Hence, the set of all possible models, denoted by $\Phi$, can be described by the following class of vectors:

$$\Phi = \{[\mu, \tau, T] \in \mathbb{R}^3 | \mu \leq 0, \tau \geq 0, T \geq 0\}. \tag{3}$$

IV. ANALYSIS OF THE DISTURBANCE DYNAMICS

This Section is dedicated to the study of the main phenomena that can affect the pupillary light reflex and alter the pupil size profile. In particular, Section III highlighted that, even in absence of light stimuli, the Hippus and Accomodation Reflex can occur, hence they need to be studied to evaluate their potential effect if eventually superimposed to the PLR. In this Section, these two phenomena are evaluated and quantified on different subjects by performing the tests that will be described in the following.

A. Tests Description

Two tests are performed on four different subjects (three dark-eyed and one light-eyed), with the set-up described in Section II-C; in these tests, only the right eye is recorded, which is constantly illuminated by the visible-light LED. After the LED is turned on to illuminate the eye, a sufficient time interval passes before starting the recording, to let the pupil adapt to the new light condition; then the camera is switched on to obtain a 7 seconds recording. This protocol is repeated for the four subjects.

Each subject performs Test 1 and Test 2 after a ten minutes interval; the subjects are given the following instructions for the two different tests:

- **Test 1** focus to a point at virtually infinite distance;
- **Test 2** move focus between a point at virtually infinite distance and a close point.

B. Results

1) Hippus: Fig. 5a shows the time profile of the pupil diameters of four subjects during Test 1, that is in absence of light stimuli and when no accomodation reflex is present (assumption based on the feedbacks of the subjects, reporting that they were able to avoid focus change during the test). Since the iris diameter is supposed to be constant, the pupil diameter (PD) is expressed in percentage of the iris diameter. The four behaviors are quite different: Dark Eyes 3 exhibits very low spontaneous oscillations, while the others present very clear diameter variations, even in absence of light stimuli; in general, such variations are contained within less then 10%, with a frequency bandwidth lower then 2 Hz, as shown in Fig. 5b. Since such variations are due to spontaneous phenomena, it is not possible to avoid their occurrence, which are always visible in the diameter profile of a pupillary light response.

The influence of the Hippus on the PLR is visible in Fig. 7, which shows a typical diameter profile after the application of a light stimulus. The profile can be divided into two parts: the first one, comprising the diameter reduction due to the light stimulus and a second one, dominated by spontaneous oscillations. From inspection of Fig. 7, the most of the dynamics due to the light reflex can be considered as expired after around 1.5 s (highlighted, in the figure, by the black vertical line), which is also consistent with the results of the previous studies in [29]. For this reason, in order to remove the effects of the spontaneous oscillation, the pupillary response observation is truncated at 1.5 s and, for identification purposes, only this portion of the response, representing the “fast dynamics” of the system, is considered.
2) Accomodation: In Test 2, the subjects are asked to move focus between a close and a far point, in order to force an accommodation reflex in their pupil; Fig. 6 shows in continuous lines the time profile of the pupil diameter of the four subjects during Test 2. The three dark-eyed subjects present a profile where the instant in which the accommodation occurs is clearly recognizable and is marked with small circles in Fig. 6: in such instants the pupil constricts of around 10% and dilates again after some seconds. In the light-eyed subject such constriction is not evident, which is consistent with the feedback collected from the subject reporting difficulties in changing focus. In general, the effects of the accommodation reflex are evident and may strongly affect the dynamics of the pupil response to a light stimulus; for this reason, during light response tests the subjects are asked to always focus on a point at infinite distance and to pay attention not to change the focus during the test, to avoid undesired pupil diameter variations due to accommodation.

3) Static Analysis: From the already described results, some conclusions can be made about the static values of the pupil diameter. As can be seen from Fig. 5a, the static pupil diameter can vary a lot between different subjects (inter-subject variability): in particular, the pupil size of Dark Eyes 3 is always below 25% of iris diameter, while Dark Eyes 1 reaches 40% of iris diameter. Those differences are due to the natural characteristics of individuals that present higher or lower sensitivity to light, thus resulting in high inter-subject variability in terms of static pupil diameter. Another significant observation comes from Fig. 6 where, together with Test 2 profile in continuous lines, also Test 1 profiles are shown in dashed lines; the plot shows that even the same subject presents slightly different static values of pupil diameter, with differences that reach more than 5% in the case of Dark Eyes 3 (intra-subject variability). We recall that Test 1 and Test 2 are performed in the same light condition and at a distance of 10 minutes from each other.

V. ANALYSIS OF ALCOHOL EFFECTS ON THE PUPILLARY LIGHT REFLEX

This Section investigates the effects of alcohol consumption on the Pupillary Light Reflex; in particular, Direct and Consensual Reflex are induced by applying a stepwise light stimulus to one eye of the subject. In the following, the test protocol is first described, then the method for processing the video sequences is explained. Finally the technique for analyzing the diameter profiles is presented. The tests are performed on different subjects at the same time.

A. Test Protocol

Each test is conducted in absence of ambient light: when the light source is off, both eyes are in complete darkness. Then, a sequence of 3 light impulses is generated, directed to the right eye of the subject (as shown in Fig. 3), each 4 seconds long, with intervals of 10 seconds, in order to let the eye adapt again to the dark condition between two consecutive steps. The tests sequence is schematically represented in Fig. 8, where each single test follows the structure already mentioned, comprising 3 light impulses. In Phase 1, three tests for each subject are performed in normal condition; since, as discussed in Section IV, every person presents his own pupil characteristics, these tests are performed in order to obtain a baseline database for each subject. In the following, data acquired in this phase will be labeled as “Sober” condition. Phase 2 is dedicated to alcohol assumption, in order to let
the BAC of the subjects overcome the maximum value set by the italian law for driving (0.5 g/L); in this phase, each person is asked to consume alcohol in the form of red wine. Phase 3 is a 45 min waiting phase, introduced to let the Blood Alcohol Concentration (BAC) reach its maximum value; this time is, in general, depending on the subject and on the alcohol type [43] and after the peak the BAC tends to slowly reduce. During Phase 4 the subjects’ BrAC level is monitored, and the measurements are reported in Fig. 8 for the three subjects; the graphic shows how those levels reach significantly higher values than the 0.5 g/L threshold designated by the italian law as the limit for driving. After each BrAC measurement, a total of 7 tests under drunk condition are performed; in the following, data acquired in this phase will be labeled as “Drunk” condition.

B. Video Sequence Processing

For each of the observed subjects, a total of 60 video sequences have been captured, 30 for each eye (9 baseline responses and 21 drunk responses), which have been analyzed with the method described in Section II. For each light step, the first operation in the video processing is the synchronization of the video sequences of the two eyes.

The test is performed in dark condition and the two LEDs (visible and infrared light) light up in the same instant, thus allowing to set the zero-instant in both videos when the respective image is illuminated. The problem is then moved to detecting the first illuminated frame of a video sequence; this can be done by extracting from each frame a “luminosity index” \( \lambda \) and comparing it with a threshold \( \lambda_T \). For such index, many definitions can be found on the literature; if we denote with \( R, G \) and \( B \) the average values of the red, green and blue channels of an image pixels, and with \( R', G' \) and \( B' \) their gamma-corrected values:

- **Intensity** (HSI model): \( \lambda = \frac{1}{4} R + \frac{1}{4} G + \frac{1}{4} B \)
- **Value** (HSV model): \( \lambda = \max(R, G, B) \)
- **Lightness** (HSI model): \( \lambda = \frac{1}{4} \max(R, G, B) + \frac{1}{4} \min(R, G, B) \)
- **Luma**
  - Rec.709: \( \lambda = 0.21 R' + 0.72 G' + 0.07 B' \)
  - Rec.601: \( \lambda = 0.30 R' + 0.59 G' + 0.11 B' \)

Fig. 9 shows the luminosity index for a sample case of the illuminated eye, together with the chosen value for the threshold \( \lambda_T = 50 \). As it can be seen, with a proper choice of the threshold, the first illuminated frame is fairly insensitive to the particular definition; for this reason, the Intensity definition is selected for its simplicity. On the other hand, due to the LED dynamics, variations of \( +/−10\text{ms} \) may be observed for different threshold values; for space reasons, we do not show the dynamics of the infra-red LED, but the conclusions are similar.

Considering a sample light response, Fig. 10 shows the cropped images on which the algorithm works in three different instants, for both eyes; on the images, the detected circles representing the iris (blue line) and the pupil (green line) are depicted. The illuminated eye (IE) and the not illuminated eye (NIE) are shown in the left and right columns respectively: since the latter is in dark condition, the images are captured with the infrared camera, with the eye being illuminated with infrared light. The not illuminated eye appears slightly bigger because the camera was positioned a little closer to obtain best focus condition and the light condition seems to change due to the camera adjusting its aperture; for both eyes the pupil size reduction is visible.

The same three instants are indicated in Fig. 11 with the three black markers: asterisk for Instant 1, square for Instant 2 and triangle for Instant 3. In this plot, the illuminated pupil diameter is computed for each frame and the results are shown in percentage of the iris diameter. The light impulse is applied after 0.1 s (the gray region represents the dark condition) and, after some delay (yellow region), the pupil
responds with a diameter reduction; the constriction dynamics last around 1.5 which, as already mentioned, corresponds to the selected time window of observation.

C. Diameter Profiles Processing

Once the time profile of the pupil size is derived from each of the video sequences, parametric identification can be employed to obtain a model of the class introduced in Section III (Eq. 2) from the data.

For each time profile of the pupillary response, the first 1.5 s are extracted and then the optimal parameters \( \hat{\phi} = [\hat{\mu}, \hat{\tau}, \hat{T}] \) are computed by solving a nonlinear least-squares curve fitting problem. Supposing \( N \) to be the total number of considered samples, if we denote by \( y(k), k = 1, \ldots, N \) the computed values of the pupil diameter in each sample and \( \hat{y}(k, \phi) \) the step response of the model \( \phi \) at instant \( k \), \( \hat{\phi} \) is determined as follows:

\[
\hat{\phi} = \arg\min_{\phi \in \Phi} \sum_{k=1}^{N} (\hat{y}(k, \phi) - y(k))^2.
\]

Without loss of generality, the step amplitude for computing the model step response is taken as 1. The result of this optimization performed on a sample pupillary response test is shown in Fig. 11, where the blue dots represent the actual data obtained the video processing algorithm and the red line is the estimated model’s step response.

Some invalid data may be present in the database, for example as a consequence of the subject partially closing the eye as a response to light stimulus application. In order to remove such invalid data, the Normalized Root Mean Square Error (NRMSE) is employed to evaluate the goodness of the fit, defined as:

\[
NRMSE = 1 - \frac{\|y - \bar{y}\|^2}{\|y - \bar{y}\|^2}
\]

where \( \bar{y} \) denotes the mean of the data \( y \). The data are considered to be valid if \( NRMSE \geq 0.8 \).

In order to compare the pupillary response of a subject in normal condition to the ones of an alcohol-impaired subject, the eight-element feature vector \( \phi = [\mu_{IE}, \mu_{NIE}, \tau_{IE}, \tau_{NIE}, \omega_{IE}, \omega_{NIE}, \tau_{IE}, \tau_{NIE}] \) is associated to every valid step response. The feature vector of a step response contains the gains of the identified models for the IE and NIE (\( \mu_{IE}, \mu_{NIE} \)), together with the relative delays (\( \tau_{IE}, \tau_{NIE} \)) and pole frequencies (\( \omega_{IE}, \omega_{NIE} \)). Moreover, the 1%-settling times (\( t_{1IE}, t_{1NIE} \)) are considered, computed as \( t_i = t_i + 5T_i, i \in \{IE, NIE\} \); this last parameter is useful in evaluating the overall pupil reactivity, since it takes into account both the delay time and the time constant. All the described features are time-related and express dynamic characteristic of the pupillary response, except the gains \( \mu_{IE} \) and \( \mu_{NIE} \) which are the only static features.

VI. RESULTS

This Section shows the results of the data analysis described in Section V performed on the entire acquired data-set. At first, all the features, for the IE and NIE, are shown and their values in the two states are compared and discussed; then, a classifier based on such features is designed and the classification results are shown.

A. Analysis of the Extracted Features

In this Section, the distributions of the extracted features for three different subjects are shown, and the two states, before and after alcohol consumption, are compared.

The entire population of step responses is first subdivided by subject, and for each subject we compute the eight features of the vector \( \phi \); Fig. 12-15 show through box plots the extracted features by subject, where the plots denoted by (a) show the IE and those denoted by (b) illustrate the NIE (each diagram comparing the baseline condition to the cases after alcohol consumption). Each baseline box refers to a 9-elements population, while the drunk state boxes refer to a 21-elements population; the illuminated eye is represented in blue, while the other in red. For each subject and eye, the darker boxes represent the baseline and the lighter is the drunk case.

For each box, the red crosses are the outliers, and the interval delimited by the dashed lines contains all elements...
The pole frequencies are presented in Fig. 13a and 13b, where the median values range in the interval \([2.6, 5.8]\) rad/sec. Considering the NIE, the three subjects present a pole frequency reduction, meaning again an overall slowdown of the response, even though only Subject 2 presents a very clear distinction between the two populations. On the other hand, only Subjects 1 and 2 present this clear behavior also on the IE.

In Fig. 14a and 14b, both the time delay and the dynamics determined by the pole are taken into account to compute the 1%-settling time, which assumes values that range in the \([1.1, 2.2]\) s interval. The analysis of this feature reflects the considerations just made for the pole frequency: Subject 2 presents a clearly slower response in drunk state with respect to baseline condition both in terms of IE and NIE, while Subject 1 presents a slower response only in terms of median value for both eyes. Subject 3 clearly shows this behavior only on the NIE.

Among the elements in the feature vector, the gain is the only one not related to time; this feature is intended to reveal a possible increase or reduction in pupil sensitivity to light, that may increment or reduce the intensity of the pupil constriction. As shown in Fig. 15a and 15b, the median value of the difference between initial and final values of the pupil diameter range in the \([16, 25]\)% interval. The analysis of this feature highlights an increment of the pupil’s sensitivity to light after alcohol assumption; in fact, the gain appears to be higher in terms of median values both for the IE and for the NIE on subjects 1 and 2. This means that the intensity of pupil constriction increases, with the pupil diameter reaching smaller values when the subject is drunk.

B. Classification Algorithm

In this Section, a classifier is designed that, based on the 8-element feature vector \(\phi\) associated to a single observation, has the goal of determining the subject’s state between “Sober” and “Drunk”. The problem is thus cast into a 2-class classification problem and four classification techniques are compared: Support Vector Machine (SVM) with Linear, Polynomial and Radial Basis Function (RBF) kernels, and Decision Tree (DT).

The database of observations comprises three subjects, with a total of 27 “Sober” and 63 “Drunk” observations; however, in order to perform training and validation of the classifier, it is useful to consider a symmetric database, comprising the same number of samples from the two classes. To do so, a total of \(N = 1000\) databases \(D_k\) are created, where \(k = 1, \ldots, N\); each of the databases comprises all the \(N_s\) available “Sober” samples and \(N_d = N_s\) “Drunk” samples randomly selected according to a uniform distribution. Each database \(D_k\) is then split into two equal-sized subsets \(D_{k,T}\) (training dataset) and \(D_{k,V}\) (validation dataset), each containing an equal number of “Sober” and “Drunk” samples, randomly chosen with uniform
probability. Then, a classifier is trained on $D_{k,T}$ and tested on $D_{i,V}$; this data separation strategy has been preferred to others (for example, k-fold cross-validation) due to the reduced amount of available data.

Four different classification strategies have been considered, which have been compared in terms of misclassification rate for each of the $N$ validation tests performed. Fig. 16 shows the misclassification rate distributions for the four strategies, from which the SVM with polynomial kernel clearly results from which the SVM with polynomial kernel clearly results in a lower average misclassification rate than the others, with a maximum misclassification rate of 9.52%, with 75% of values below 4.76%.

VII. CONCLUSIONS

This paper presents a feasibility analysis for the design of a Driver Drunkenness Detection System, in the class of the standard DDDSs, by evaluating the subject’s Pupillary Light Reflex, by means of a methodology based on image-processing techniques.

At first, a two-steps method for extracting the pupil diameter profile from a video-sequence is presented, that makes use of the Circular Hough Transform; in the first step, a pupil/iris search is performed to identify the image region that contains the elements of interest. Then, in the second step, pupil size estimation is performed only on the cropped image containing the pupil, whose size is expressed in terms of the predetermined iris diameter.

An analysis of the main phenomena influencing pupil size variability is performed; as a result, a $\pm 5\%$ spontaneous pupil size variability at [0, 2] Hz and a significant pupil constriction ($> 5\%$) due to the accommodation reflex when focusing on a close object have been observed.

Some tests have been performed on different subjects by applying a light stimulus on one eye and measuring the Direct and Consensual Light response on both eyes both in normal and drunk state. The results showed a general increment of the response time of the observed subjects in terms of delay, pole frequency and settling time, together with a general increment in the sensitivity to light. A polynomial-kernel SVM is designed, that is able to discriminate between “Sober” and “Drunk” state of a subject based on a set of 8 features extracted from his/her pupil diameter profile. Based on the acquired dataset, the classifier provides a misclassification rate below 10%, which falls below 5% in the 75% of the cases.

As a conclusion, this paper addressed the problem of determining the feasibility of designing a robust Driver Drunkenness Detection System based on the pupillary reflex measurement; a set of features have been isolated and a classifier has been designed on a dataset of reduced dimension, with very encouraging results. As a consequence, future work based on a statistically relevant dataset acquired on a variety of subjects will be dedicated to design a robust classifier to be effectively deployed in real-life applications.

REFERENCES


