The ULg Multimodality Drowsiness Database (called DROZY) and Examples of Use

Quentin Massoz   Thomas Langohr   Clémentine François   Jacques G. Verly

Laboratory for Signal and Image Exploitation, University of Liège, Liège, Belgium

quentin.massoz@ulg.ac.be

Abstract

Drowsiness is a major cause of accidents, in particular in road transportation. It is thus crucial to develop robust drowsiness monitoring systems. There is a widespread agreement that the best way to monitor drowsiness is by closely monitoring symptoms of drowsiness that are directly linked to the physiology of an operator such as a driver. The best systems are completely transparent to the operator until the moment he/she must react. In transportation, cameras placed in the passenger compartment and looking at least at the face of the driver are most likely the best way to sense physiology related symptoms such as facial expressions and the fine behavior of the eyeballs and eyelids. We present here the new database[1] called DROZY that provides multiple modalities of data to tackle the design of drowsiness monitoring systems and related experiments. We also present two novel systems developed using this database that can make predictions about the speed of reaction of an operator by using near-infrared intensity and range images of his/her face.

1. Introduction

Drowsiness kills thousands of drivers each year [11, 18]. It is thus critical to develop real-time, automatic drowsiness monitoring systems designed to issue timely warnings to the driver and/or automatic driving system (when applicable). In transportation, one generally distinguishes between approaches based on vehicle behavior (*e.g.* line crossing), operator behavior (*e.g.* steering wheel control), and operator physiology (*e.g.* brain signals and ocular parameters). Systems based on lane crossings do not work in the absence of lines, in significant road curves, and/or in snow conditions. Systems based on steering wheel control are not readily transferable between different types of vehicles, and certainly not to other domains such as the surveillance of nuclear plants. Systems based on operator physiology have

the significant advantage of being mostly application independent. These latter systems can exploit a variety of sensor modalities, and, depending on the modality, the system's sensor(s) can be either worn by the driver or positioned remotely in the vehicle's passenger compartment, such as in the dashboard. Clearly, the remote ("dashboard") solution with contactless sensors has the significant operational advantage of not requiring any action from the driver, except reacting to warnings. The most natural sensing means for the remote, contactless solution is one or more cameras (2D and/or 3D). These have the advantage of being useful to quantify, not only drowsiness, but also other operator conditions, such as cognitive distraction.

The development, test, and performance evaluation of drowsiness monitoring systems imply access to appropriate data, and such data is not readily available today. We thus introduce in this paper a new database intended to facilitate research and development on drowsiness monitoring systems using a 2D or 3D camera. The generic, full name of this database is "The ULg Multimodality Drowsiness Database", and the generic, short name "DROZY". Many different types of developments and experiments can be carried out with this database (for drowsiness, cognitive distraction, *etc.*). At this stage, the database does not provide performance figures against which other systems can measure themselves up. To illustrate the use of our database, we describe two innovative drowsiness monitoring systems developed and tested with it.

Section 2 describes the motivation and goals, Section 3 the data collection, Section 4 the DROZY database, Section 5 the examples of use, and Section 6 the conclusions.

2. Motivation and goals

We have shown above that cameras placed remotely in the control area of a vehicle are ideally suited for providing data to automatic systems that monitor drowsiness, distraction, *etc.* Here, we focus on a single dashboard-mounted 3D camera capturing intensity images (I-images) and range images (R-images) of the face of the driver. One of our current

---

[1]Available on http://www.drozy.ulg.ac.be

goals is to produce automatically, continuously, and in real-time a level of drowsiness (LoD) based on these images. Provided they have enough resolution, one can extract from them a variety of parameters that are indicative of drowsiness, such as ocular parameters [16] - *e.g.* the PERCLOS [3] - and yawning parameters. The literature provides articles showing how to obtain an LoD from ocular parameters [5, 9]. Besides the LoD, one can produce a variety of other parameter signals from the images, such as degrees of anger, surprise, *etc*.

If one develops a new drowsiness monitoring system from images of the (driver's) face, it is paramount to validate the results produced by the system, and to evaluate its performance. Polysomnography (PSG) is regarded as the "Gold Standard" to study sleep and, in particular, to score sleep stages [14]. By extension, PSG is generally - but arguably - viewed as the "Gold Standard" to assess drowsiness [1]. It is thus important that the database contain all PSG signals of interest, along with the images of the face, and perfectly time-synchronized with them. The literature provides articles showing how to obtain an LoD from PSG signals [1, 7, 13].

In experimental settings, one can also ask a subject to (self-)evaluate his/her LoD on the Karolinska Sleepiness Scale (KSS) [1, 6, 10]. Of course, this would be impractical in real-life driving, and would call upon the intervention of the driver, which we wish to avoid. Moreover, one must keep in mind that self-reported LoDs are generally more accurately perceived at high levels than at medium levels [8, 17].

PSG and the KSS lead to measures of drowsiness (objective and subjective, respectively). However, there is another test that allows one to measure the level of vigilance of a subject, *i.e.* his/her ability to perform a task: the psychomotor vigilance test (PVT). The PVT produces an objective measure of vigilance and, indirectly, of drowsiness.

The above discussion indicates that it would be very useful to have a database containing the following main ingredients, all perfectly synchronized in time: images of the face, PSG signals, KSS data, and PVT data. Below, we present how we obtained these raw ingredients, as well as other ingredients derived from them.

3. Data collection

3.1. Overview

Here, we describe how we collected the data used to build our database. Fourteen healthy subjects/participants (3 males, 11 females), aged 22.7 ± 2.3 (mean $\pm$ SD), took part in the data collection. The experimental protocol (approved by the Ethics Committee of our university) called for each subject to take three PVTs over two consecutive days, under conditions of increasing sleep deprivation induced by

acute, prolonged waking.

The protocol required subjects without any alcohol dependency, drug addition, or sleep disorder. Each was asked to maintain a normal sleep pattern for the week prior to taking the first PVT and to have a full night sleep (of 7 to 8 hours at least) just before this PVT. Each was also asked to maintain a sleep diary during that week, to allow us to verify that the sleep requirements were met. Once a participant took the first PVT, he/she was not allowed to sleep until after the third PVT, resulting in a total sleep deprivation of 28 to 30 hours. We organized several sessions of three PVTs, each with a few subjects successively taking each PVT.

Details of the tests follow. **Day 1** (starting just after midnight). At 8:30 (in 24 hour time), the scheduled participants arrived at the laboratory, and were equipped with the PSG electrodes. Between 10:00 and 11:00, they (successively) carried out the first PVT, called PVT1. Then, they were equipped with wrist actigraphs (to verify that they would not sleep) and were allowed to leave the laboratory. From 12:00 on, they were not allowed to consume any coffee, tea, energy drinks, or other stimulants. At 20:30, they returned and were equipped with the PSG electrodes. They stayed overnight in the laboratory, and until the end of the tests. During the night, they were allowed to use multimedia devices, to play card and board games, and to interact with the laboratory staff. **Day 2**. Between 3:30 and 4:00, they carried out the second PVT, called PVT2. At 8:30, we provided breakfast. Between 12:00 and 12:30, they carried out the third PVT, called PVT3. This concluded the tests. We strongly advised the participants not to drive home by themselves, and we offered alternate transportation solutions when necessary. Figure 1 summarizes the data collection schedule.

The PVTs were all performed in a quiet, isolated laboratory environment without any temporal cues (*e.g.* watch or smartphone). The room lights were turned off for PVT2 and PVT3. For a 15-minute period before each PVT, we instructed the subjects to part with their phones, computers, and any other screen devices. At the beginning of each PVT, we asked the participant to estimate his/her level of drowsiness (LoD) in terms of the Karolinska Sleepiness Scale (KSS), which is a popular, subjective means to estimate one's LoD [1, 6, 10]. During each PVT, we also recorded the PSG signals and the images of the camera looking at the face of the subject.

We now describe in more detail the various types of collected raw data, and some additional, useful data obtained from the images.

3.2. PVT data

The psychomotor vigilance test (PVT) has become one of the most widely used tools to measure performance impairments due to drowsiness. The PVT gives the reaction

Figure 1. Pictorial summary of data collection schedule.

times (RTs) to visual or auditory stimuli that occur at random inter-stimulus interval.

Multiple studies have shown that the PVT is valid, reliable, and extremely sensitive to sleep deprivation [2, 4]. Compared to other tests, the PVT has the advantage of being independent of aptitude (inter-subject variability) and learning (intra-subject variability) [4]. Moreover, the PVT can be considered to be an indicator of "real world" performance since it requires sustained attention and quick response to sudden events [4].

We implemented our own version of the 10-minute PVT proposed by Basner and Dinges [2]. The subjects were instructed to monitor a red rectangular box over a black background on a computer screen, and to press a response button as soon as they noticed the appearance within the box of a yellow stimulus counter (expressed in milliseconds). When the button was pressed, the counter stopped and the RT remained displayed for 1 second. RTs below 100 milliseconds were discarded as false starts (errors of commission). After 30 seconds without any response, the counter timed out and displayed a yellow "overrun" message inside the box for a few seconds. The inter-stimulus interval, defined as the time interval between the last response and the appearance of the next stimulus, was varied randomly between 2 and 10 seconds.

Among the several PVT outcome metrics that can be computed from RTs, the following are the most frequently found in the literature: number of lapses (generally defined as RTs $\geq$ 500 ms), mean RT, and mean 1/RT [2].

3.3. PSG signals

Polysomnography (PSG) uses a number of physiology-related signals to study sleep [14], i.e. the electroencephalogram (EEG), electrooculogram (EOG), electrocardiogram (ECG), and electromyogram (EMG). Such PSG signals are also useful to assess drowsiness since activity in the alpha (8-12 Hz) and theta (4-8 Hz) bands and/or slow eye movements during a task are strong, reliable indicators of drowsiness [1, 7]. We used the portable, laboratory Embla Titanium system to record the Fz, Pz, Cz, C3, and C4 EEG

Figure 2. Examples of cropped near-infrared (NIR) intensity images obtained with the Kinect v2 sensor.

channels (referenced to A1, in the international 10-20 system), the horizontal and vertical EOG channels, the EMG, and the ECG, all sampled at 512 Hz.

3.4. Camera images

To explore the benefit of using 3D data, we directly used a 3D range camera, *i.e.* the Microsoft Kinect v2 sensor, which provides - for each "video" frame - a pair of near infrared (NIR) intensity and range images (I-image and R-image), and a color image. Since drowsiness monitoring systems must generally operate in all lighting conditions, including in total darkness, we only retained the I- & R-images (corresponding to active NIR illumination). The NIR images are of size 512×424 pixels, have 16-bit values, and are recorded at 30 fps. The camera was positioned just below the computer screen used for the PVTs, at a distance of about 0.7 m from the subject. Figure 2 shows examples of NIR intensity images.

3.5. Annotations of face landmarks

Some users of the database will likely want to use the I- & R-images of the face, *e.g.* to characterize the eyelids movements and/or facial expressions. Therefore, for each frame taken, the database provides, not only the corresponding I- & R-images, but also the positions of 68 (predefined) landmarks of the face, *e.g.* the left corner of the left eye. We
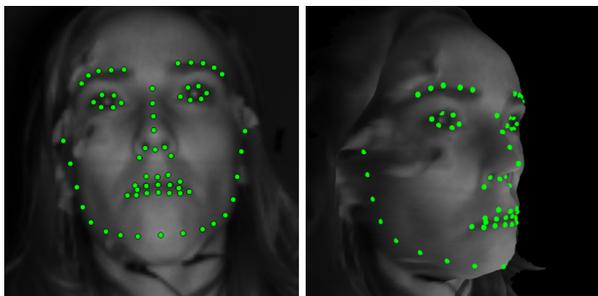
Figure 3. Example of manual annotations for one frame in 2D (left) and in 3D (right).

now describe how we located (or "annotated") these landmarks in all frames. We proceeded in two successive steps. The first - manual - resulted in the annotation of 720 manually selected frames, and the second - automatic - in the annotation of all frames.

Manual annotation: We manually selected 720 frames spread across all 14 subjects - with about 51 per subject - and judged to represent rather completely the various appearances of the face of each. For each of these frames, we built (automatically) a 3D surface of the face from the (corresponding) R-image, and we draped over it the (co-registered) I-image. By looking jointly at the I-image and the rotatable, textured 3D surface, we pinpointed the location of each visible landmark in the frame, in terms of pixel coordinates. Since the I and R-images are co-registered, the location of each such landmark was also immediately known in the three camera axes, thus in terms of three coordinates expressed in millimeters. Figure 3 shows manual annotation results.

Automatic annotation: The annotation of all frames can - reasonably - only be done in an automated way. We chose the following automatic approach. For each sequence of frames for a given PVT, we first detect the face region in the first frame by applying the method of Viola and Jones [19] to its I-image. We then fit a statistical, deformable shape model - a point density model (PDM) - to the successive frames in the sequence using subject-specific constrained local models (CLMs), *i.e.* one specific CLM per subject, with the regularized landmark mean-shift (RLMS) fitting strategy introduced by Saragih et al. [15]. The fitting leads to some optimal positioning of the 68 landmarks in all successive frames in both 2D and 3D. If the frame-to-frame tracking of the landmarks fails, we reinitialize the process with the detection step. Figure 4 shows automatic annotation results.

To be more precise, in the subject-specific CLM learning stage, we learn a 68-landmarks PDM and the corre-

sponding 68 independent, local appearance experts from the manually landmark-annotated frames of the corresponding subject. These local experts are discriminative regressors giving the probability that their corresponding landmark is correctly located based on a small appearance patch around it [15].

In the RLMS fitting stage, we start from coarse landmark positions (*e.g.* their positions obtained at the previous frame) and independently extract one response map around each landmark using the local experts. Then, to refine the landmark positions, the RLMS fitting strategy implements the EM algorithm on the response maps. The E-step applies one iteration of the mean-shift algorithm independently on each of the 68 response maps. The M-step applies one iteration of the Gauss-Newton method where the PDM is used to apply a global regularization to the 68 mean-shift vectors obtained at the E-step.

4. The DROZY database

4.1. Content

The DROZY database contains the following perfectly time-synchronized data for each of the 14 subjects, and for each of the three PVTs they each took:

- **KSS:** the self-reported LoD, as a score on this scale, just before taking the PVT;
- **PVT:** stimulus times and corresponding RTs;
- **PSG:** 5 EEG channels, 2 EOG channels, ECG, and EMG;
- **Kinect v2 sensor:** NIR intensity and range images;
- **Face landmarks (68):** 2D coordinates in the image axes expressed in real pixels, and 3D coordinates in the camera axes expressed in millimeters, all obtained manually for 720 selected, key images, and automatically for all frames.

The database contains about 500,000 frames, and has a total size of about 283 Gigaoctets.

4.2. Usage

Based on the many different types of data provided by the (multimodality) database, one can conduct various experiments, and develop and validate various systems. Examples of systems follow: (1) obtain LoD from PSG signals; (2) characterize facial expressions and predict RTs; (3) characterize facial expressions and predict LoDs (obtained from PSG or KSS).

The examples of use below concern the analysis of the behavior of the eyes to predict RTs.
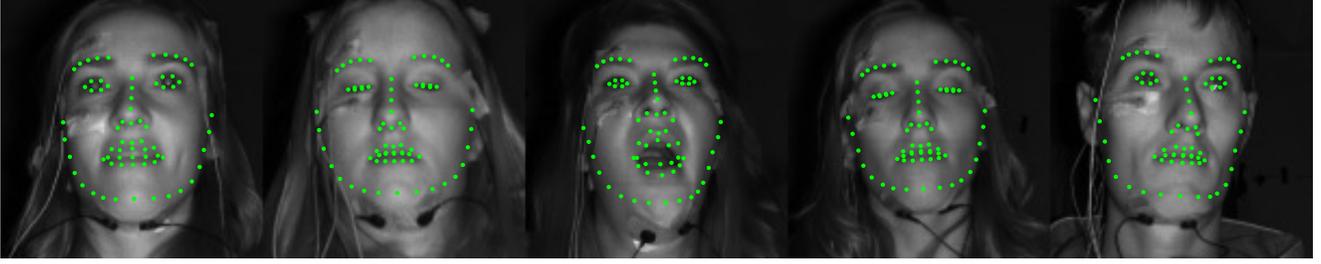
Figure 4. Examples of automatic annotations for five frames.

5. Examples of use

5.1. Problem

We wish to design and evaluate a drowsiness monitoring system that can, at each instant, predict how fast a subject can react based on the behavior of his/her eyes at that instant.

5.2. Solution and systems

It seems natural to quantify the subject reaction speed in terms of the RTs and the behavior of his/her eyes in terms of some ocular parameters. We decided to base all ocular parameters (defined below) on the distance between the upper and lower eyelids for each eye, and more precisely on the average of the left and right distances. The database provides all necessary eyelid landmarks and related, time-synchronized RTs.

The processing steps are as follows: (1) compute normalized eye opening, (2) compute ocular parameters, (3) perform regression or classification (leading to two distinct systems). In a system confronted with new, unseen images, the first step would be to locate the face and track the 68 face landmarks (or at least the eyelid landmarks), exactly as described above. Details of the three steps follow.

5.3. Compute normalized eye opening

The eyelids of each eye are described by 6 landmarks: 2 at the corners, 2 on the upper eyelid, and 2 on the lower one; the 4 eyelid landmarks are grouped into a left pair and a right pair. The eye opening is defined as the average of the distances between the landmarks in each of the 4 eyelid pairs corresponding to both eyes. We denote this opening by $d[n]$, where n is the discrete time, or frame, index. We treat quantities such as $d[n]$ as discrete-time signals [12].

Since the maximum opening of the eye changes with time, *e.g.* with the gaze direction, it proves useful to divide $d[n]$ by the maximum opening around time n, denoted by $b[n]$, which results in the normalized eye opening signal $s[n]$. The baseline signal $b[n]$ is computed recursively according to

$$b[n] = \big(1 - \alpha[n]\big)b[n-1] + \alpha[n]d[n], \qquad (1)$$

where $\alpha[n]$ is a smoothing factor defined as

$$
\begin{aligned}
\alpha[n] = \alpha_0 &* \exp\{-\alpha_d\big(d[n] - d[n-1]\big)^2\} \\
&* \exp\{-\alpha_a\big[d[n] - b[n-1]\big]_+\} \\
&* \exp\{-\alpha_b\big[b[n-1] - d[n]\big]_+\} \\
&* H\big(d[n] - \alpha_m d_{median}[n]\big),
\end{aligned}
\qquad (2)
$$

where $[x]_+$ is x if $x \geq 0$, and 0 otherwise, the Heaviside step function $H(x)$ is 1 if $x \geq 0$, and 0 otherwise, and $d_{median}[n]$ is the median value of $\{d[i]\colon \forall i \in [1, n]\}$. The values of α_0, α_d, α_a, α_b, and α_m are empirically set to 0.4, 15, 0.5, 2, and 0.7, respectively.

The smoothing factor $\alpha[n]$ is designed to favor eyelids idleness (*i.e.* small values of derivative) but not in the case of nearly closed eyelids (*i.e.* $d[n]$ below its median). It is also designed to be flexible enough to allow the baseline signal $b[n]$ to quickly adapt when the gaze direction changes.

5.4. Compute ocular parameters

We consider here 15 ocular parameters, all computed from the values of the normalized eye opening signal $s[n]$ in a contiguous window W ending at the present time index. (The length of W is specified below.) Ten parameters are related to a histogram of the values of $s[n]$ in W, and five are related to blinks.

The 10 histogram(-related) parameters are the 10 proportions of elements in each the 10 successive bins of a 10-bins histogram, which all values of $s[n]$ in W are arranged in, with all values above 1 placed in the last bin. (The 10 proportions sum to 1.)

The 5 blink parameters are obtained as follows. The blinks are extracted from the first derivative signal of $s[n]$ (computed from the time stamps) using fixed, experimentally determined thresholds. The blink parameters are the average blink duration, average eyelid closure duration, average closed eyelid duration, average eyelid reopening duration, and number of microsleeps (defined as blinks with durations ≥ 500 ms).

5.5. Perform regression or classification

We consider two distinct systems, one using regression, and the other classification. Regression produces, from the pre-stimulus ocular parameters, the post-stimulus mean RT (in milliseconds), where the mean is computed over the one-minute period after the stimulus. Classification determines, also from the pre-stimulus ocular parameters, whether the actual post-stimulus mean RT is above some fixed threshold, where the mean is computed as above. We set the threshold to 500 ms since an RT above this value is considered as a lapse (error of omission). The "positive" class corresponds to RT $\geq$ 500 ms, and the "negative" class to RT < 500 ms.

For regression, we use the Epsilon-Support Vector Regression (epsilon-SVR) model, and for classification, the Support Vector Machine (SVM) classifier. In both cases, we tried two distinct kernels, and we evaluated the (generalization) performance with leave-one-subject-out cross-validation.

5.6. Results

Both for regression and classification, we start from the same feature matrix, consisting of 3064 examples (1 per stimulus) of 105 features (for 15 ocular parameters extracted from 7 pre-stimulus windows). The 7 windows have lengths of $\{30, 35, 40, 45, 50, 55, 60\}$ seconds, respectively, and they all end at the time of the related stimulus. All 105 features are linearly normalized to the $[0, 1]$ range. For the classification, out of the 3064 examples, 448 are labeled "positive" and 2616 "negative".

Regression: Using a linear kernel, we obtain an RMSE of 113.2973 ms, and a Pearson correlation coefficient (PCC) of 0.6021, with $C = 4$ and $\epsilon = 90$ as parameters. Using a radial basis function (RBF) kernel, we obtain an RMSE of 105.8370 ms, and a PCC of 0.6681, with $C = 50$, $\epsilon = 45$ and $\gamma = 0.4$ as parameters. Figure 5 illustrates the performance of the regression via a scatter plot.

Classification: Using a linear kernel, we obtain a specificity (true negative rate) of 78.36%, a sensitivity (true positive rate) of 79.02%, and an accuracy of 78.46%, with $C = 0.34$ as parameter. Using an RBF kernel, we obtain a specificity of 86.24%, a sensitivity of 77.68%, and an accuracy of 84.99%, with $C = 1500$ and $\gamma = 0.15$ as parameters.

6. Conclusions

We introduced the new, multimodality database DROZY designed to allow researchers to develop innovative drowsiness monitoring systems and related experiments by using
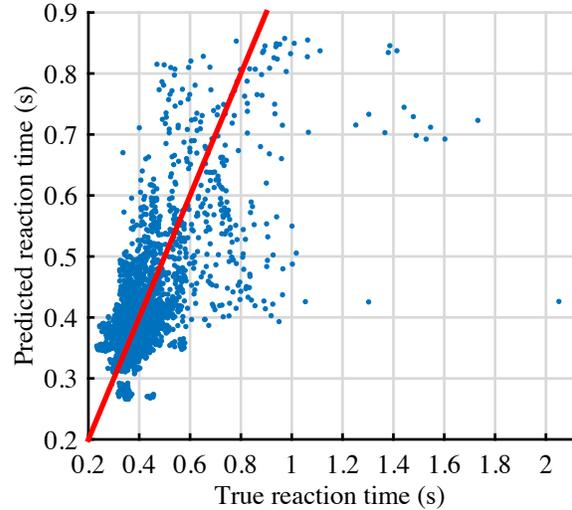
Figure 5. Scatter plot where each point shows both the true reaction time (RT) and the predicted RT for the use of regression, an RBF kernel, and leave-one-subject-out cross-validation. The straight line segment corresponds to the perfect regressor.

controlled data that is often difficult and expensive to obtain. In addition, we presented a pair of related, innovative drowsiness monitoring systems developed by using the database, and thus serving also as prototypical examples of its use.

The database was built using multimodality data collected from 14 subjects who performed three successive psychomotor vigilance tests (PVTs) in conditions of increasing sleep deprivation. For each subject, and for each PVT, the database contains the following perfectly time-synchronized raw data: Karolinska Sleepiness Scale (KSS) data, PVT data (including reaction times), polysomnography (PSG) signals, and near-infrared (NIR) intensity and range images of the face. In addition, it contains the 2D and 3D positions (*i.e.* annotations) of 68 face landmarks for all frames.

The presented drowsiness monitoring systems draw conclusions about a subject's speed of reaction based on 3D images of his/her face. More specifically, these systems can, based on ocular parameters derived from the NIR images of the face, either predict the reactions times (by regression) or determine whether the true reaction time is above some fixed threshold (by binary classification). These two systems have the following performances: an RMSE of 106 ms with a Pearson correlation coefficient of 0.67 for the regression, and a probability of correct detection of 85% for the classification.

These examples of use demonstrate the usefulness of the database. We are looking forward to other researchers using the database in some other innovative ways. Of course, the ultimate goal is that research and development based on the

DROZY database contribute some day to save thousands of human lives on the road!

Acknowledgements

References

[1] T. Åkerstedt and M. Gillberg. Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience*, 52(1-2):29–37, 1990.

[2] M. Basner and D. F. Dinges. Maximizing sensitivity of the psychomotor vigilance test (PVT) to sleep loss. *Sleep*, 34(5):581–591, 2011.

[3] D. F. Dinges and R. Grace. PERCLOS: A valid psychophysiological measure of alertness as assessed by psychomotor vigilance. *US Department of Transportation, Federal Highway Administration, Publication Number FHWA-MCRT-98-006*, 1998.

[4] J. Dorrian, N. L. Rogers, and D. F. Dinges. Psychomotor vigilance performance: Neurocognitive assay sensitive to sleep loss. *Sleep Deprivation: Clinical Issues, Pharmacology and Sleep Loss Effects*, pages 39–70, 2005.

[5] C. François, J. Wertz, M. Kirkove, and J. G. Verly. Evaluation of the performance of an experimental somnolence quantification system in terms of reaction times and lapses. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5820–5823. IEEE, 2014.

[6] M. Gillberg, G. Kecklund, and T. Åkerstedt. Relations between performance and subjective ratings of sleepiness during a night awake. *Sleep: Journal of Sleep Research & Sleep Medicine*, 1994.

[7] M. Gillberg, G. Kecklund, and T. Åkerstedt. Sleepiness and performance of professional drivers in a truck simulator - comparisons between day and night driving. *Journal of Sleep Research*, 5(1):12–15, 1996.

[8] J. A. Horne and S. D. Baulk. Awareness of sleepiness when driving. *Psychophysiology*, 41(1):161–165, 2004.

[9] M. W. Johns, A. Tucker, R. Chapman, K. Crowley, and N. Michael. Monitoring eye and eyelid movements by infrared reflectance oculography to measure drowsiness in drivers. *Somnologie-Schlafforschung und Schlafmedizin*, 11(4):234–242, 2007.

[10] K. Kaida, M. Takahashi, T. Åkerstedt, A. Nakata, Y. Otsuka, T. Haratani, and K. Fukasawa. Validation of the Karolinska Sleepiness Scale against performance and EEG variables. *Clinical Neurophysiology*, 117(7):1574–1581, 2006.

[11] National Highway Traffic Safety Administration. Traffic safety facts crash stats: drowsy driving. Washington, DC: US Department of Transportation, National Highway Traffic Safety Administration; 2011, 2012.

[12] A. V. Oppenheim and R. W. Schafer. *Discrete-Time Signal Processing*. Prentice Hall Press, Upper Saddle River, NJ, USA, 3rd edition, 2009.

[13] A. Picot, S. Charbonnier, and A. Caplier. On-line automatic detection of driver drowsiness using a single electroencephalographic channel. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 3864–3867. IEEE, 2008.

[14] A. Rechtschaffen and A. Kales. A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. 1968.

[15] J. M. Saragih, S. Lucey, and J. F. Cohn. Deformable model fitting by regularized landmark mean-shift. *International Journal of Computer Vision*, 91(2):200–215, 2011.

[16] R. Schleicher, N. Galley, S. Briest, and L. Galley. Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired? *Ergonomics*, 51(7):982–1010, 2008.

[17] E. A. Schmidt, M. Schrauf, M. Simon, M. Fritzsche, A. Buchner, and W. E. Kincses. Drivers misjudgement of vigilance state during prolonged monotonous daytime driving. *Accident Analysis & Prevention*, 41(5):1087–1093, 2009.

[18] B. C. Tefft. Prevalence of motor vehicle crashes involving drowsy drivers, United States, 2009–2013. *Rep. Prepared for the American Automobile Association (AAA) Foundation for Traffic Safety, Washington, DC*, 2014.

[19] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. *Computer Vision and Pattern Recognition*, 1:511–518, 2001.