Sleep Deprivation Detection for Real-Time Driver Monitoring Using Deep Learning

Miguel García-García[1,2(✉)], Alice Caplier[2], and Michele Rombaut[2]

[1] Innov+, Batiment 503, Centre Universitaire d'Orsay, 91400 Orsay, France
miguel.garcia@innov-plus.com
[2] Univ. Grenoble Alpes, CNRS,
Grenoble INP, GIPSA-lab, 38000 Grenoble, France
{alice.caplier,michele.rombaut}@gipsa-lab.fr

Abstract. We propose a non-invasive method to detect sleep deprivation by evaluating a short video sequence of a subject. Computer Vision techniques are used to crop the face from every frame and classify it (within a Deep Learning framework) into two classes: "rested" or "sleep deprived". The system has been trained on a database of subjects recorded under severe sleep deprivation conditions. A prototype has been implemented in a low-cost Android device proving its viability for real-time driver monitoring applications. Tests on real world data have been carried out and show encouraging performances but also reveal the need of larger datasets for training.

Keywords: Mobilenet · Road safety · Driver drowsiness
Sleep deprivation

1 Introduction

Sleep deprivation is the condition of not having enough sleep. Its physical impact have been studied [1], as well as its effects on cognitive [2] and driving performance [3]. It is estimated that having less than 8 h of sleep is analogue to drinking a certain amount of alcohol [4].

Drowsy driving is a major threat to road safety. In 2014, 846 fatalities related to drowsy drivers were recorded in the United States [5], but it is believed that this number is systematically underestimated and the real cipher may be near 6,000 fatal crashes each year [6]. Therefore, drowsiness detection is an important challenge for the automotive industry, which proposes several options either for alerting the driver in real time, for offering coaching sessions to correct risky behaviors, or for handing over the control to an autonomous vehicle.

There are two main categories of real-time drowsiness detectors: vehicle-focused or driver-focused. Vehicle-focused detectors try to infer drowsiness from a deterioration on driving performance by monitoring the behavior of vehicle.

A. Campilho et al. (Eds.): ICIAR 2018, LNCS 10882, pp. 435–442, 2018.
https://doi.org/10.1007/978-3-319-93000-8_49

They are limited to certain specific conditions, such as highway driving. Driver-focused detectors may infer drowsiness from psychophysiological parameters of the subject (electroencephalogram, electrooculogram, electromyogram, skin conductivity) but need the usage of invasive captors, which can be uncomfortable. Computer Vision based methods are getting popular within the industry as they can evaluate in real-time certain driver parameters without invasive instruments.

Sleep deprivation is not the only cause for drowsy driving, but it leads almost inevitably to it. Hence, being able to detect if a driver has not had enough sleep may be a good way to improve the performance of Advanced Driver Assistance Systems (ADAS) and therefore to prevent road accidents. We propose here a method to estimate sleep deprivation based on Computer Vision and Deep Learning techniques. We prove its feasibility by building a prototype on a low cost device which evaluates driver status in real time.

2 Related Work

Recent studies have detailed the effects of sleep deprivation on facial appearance [7], specifically particular patterns on eyes, mouth and skin after a prolonged wake of 31 h. Classical approaches for drowsiness detection use one or several of these features, such as blinking frequency, eye closure or yawning frequency [8]. Other studies use the Facial Action Coding System [9] to perform the classification, obtaining very subject-dependent results [10]. These methods do not work with all the possible information facial features can offer.

Another approaches have tried to use the information of the whole facial area to infer fatigue. Several studies have been carried out over NTHU Driver Drowsiness Detection dataset [11], which provides data from 8 subjects performing a set of facial expressions simulating different scenarios [12,13]. Best results are achieved by Lyu et al. [14], who use a Long-term Multi-granularity Deep Framework for a performance of 90.05%. However, the use of simulated data does not seem appropriated for sleep deprivation detection, as some of the effects revealed by [7] are impossible to imitate consciously (skin tone, circles around eyes, etc.).

Dwivedi et al. [15] used a private dataset of 30 male subjects playing a video game after midnight at different fatigue levels and analyzed the whole face with a shallow convolutional neural network, obtaining an accuracy score of 78%. Drowsiness was induced by increasing fatigue, not by sleep deprivation.

Our goal will be the detection of sleep deprivation using information of the whole face and non-simulated data with objective ground truth.

3 Proposed System

3.1 Face Classification

The goal of the system is to classify images within two classes: “rested” and “sleep deprived”. Faces are extracted from the videos using OpenCV Haar Cascades [16], cropping the central 80% area to avoid border effects and normalizing.

Although we considered using FaceNet [17] (a state of the art face recognition model) for face classification, we ultimately chose MobileNets [18] as we intended to embed it on a smartphone-based system. MobileNets are a class of highly efficient non-linear models created specifically for mobile and embedded vision applications. They are based on 3×3 depthwise separable convolutions, which need 8 to 9 times less computation than standard convolutions with only a small loss in performance. Standard convolutions perform a filtering and then combine the outputs in one step in order to produce a new representation. Depthwise separable convolutions split the operation into two steps: depthwise convolutions apply a single filter per each input channel and their outputs are linearly combined by 1×1 pointwise convolutions. The entire architecture has 28 layers.

Two hyper-parameters may be tuned in order to construct smaller and less computationally expensive networks. The width multiplier ($\alpha \in (0, 1]$) thins the network uniformly at each layer, scaling the computational cost and the number of parameters by α^2. The resolution multiplier ($\rho \in (0, 1]$) changes the resolution of the input image scaling the computational cost by ρ^2. A 1.0 MobileNet-224 model ($\alpha = 1, \rho = 1$) was chosen, as it is the one that offers the best performance in ImageNet classification [18].

The output of MobileNet for a frame n is $P_{sd}(n)$, which represents an estimation of the probability of the frame to belong to "sleep deprived" class. If $P_{sd}(n) > 0.5$, the subject in the frame n is classified as "sleep deprived".

3.2 Video Classification

The proposed system is conceived to perform image classification. However, sleep deprivation is a behavior with a certain persistence over time. Thus, in a short video sequence, every frame should belong to the same class. This consistence allows for an improvement of the performance if the decision is made over the entire sequence by reducing the impact of outliers in the classification results.

In order to maximize the weight of the frames classified with higher confidence we applied a logistic function [19] with empirical parameters $L = 1, k = 20, x_0 = 0,5$, and averaged the result over the entire sequence. Let n be a frame from a sequence of N frames, and $P_{sd}(n)$ the probability that the MobileNet step assigns to a facial image to be drowsy; the subject in the video sequence is classified as "sleep deprived" if $\overline{P_{SD}} > 0.5$ in Eq. 1.

$$\overline{P_{SD}} = \frac{1}{N} \sum_{n=0}^{N-1} \frac{1}{1 + e^{-20[P_{sd}(n)-0.5]}} \tag{1}$$

4 Experimental Settings

4.1 Datasets

The ULg Multi-modality Drowsiness Database (DROZY). Drozy [20] provides multi-modal data of 14 healthy people (3 males and 11 females aged

22.7 ± 2.3, without any alcohol dependency, drug addition or sleep disorder) who were asked to perform three 10-min psycho-motor vigilance tests (PVT) under conditions of increasing sleep deprivation induced by prolonged wake: first day at 10:00 AM (normal sleep patterns), second day at 3:30 AM (20 h of sleep deprivation) and second day at 11:00 AM (28 h of sleep deprivation). 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 of a yellow stimulus counter within the box. Videos are recorded with an artificial near-infrared illumination using a Microsoft Kinect v2 sensor which provides frames of size 512×424 pixels at 30 frames per second for a duration of 10 min. Figure 1 shows examples of NIR frames.

Fig. 1. Examples of NIR frames from DROZY database.

Toucango Dataset. As part of the Toucango project [21], we conducted a naturalistic data acquisition campaign between March 2015 and February 2017 with the participation of professional drivers [22]. We equipped operational medium size vans and buses with portable devices carrying a near-infrared camera mounted in the dashboard. Videos were recorded at 30 frames per second and with a size of 640×480 pixels under artificial near-infrared illumination. Trips were filmed at day and night, with natural changing lightning conditions. As per the agreement between the companies involved, videos and images must remain confidential and thus samples cannot be shown here.

For the particular purpose of sleep deprivation detection, we selected video sequences from 5 drivers at two separate times: at daytime at the beginning of a work shift ("rested") and just before dawn at the end of a work shift ("sleep deprived"). Subjects are professional drivers in operation, so sleep deprivation conditions are not as extreme as in the DROZY dataset.

4.2 Implementation for Real-Time Classification on Android

Training and offline tests were performed on a laptop equipped with an Intel Core i7-6500U CPU (3.1 GHz), 8 GB RAM and a GPU Nvidia GeForce 920M.

In order to evaluate is feasibility for an usage in real time we implemented the system in Android. It was embedded in a smartphone Lenovo K6 Note, equipped

with a Qualcomm Snapdragon 430 MSM8937 chipset (Octa-core 1.4 GHz Cortex-A53 CPU, Adreno 505 GPU, 3 GB RAM) running a standard version of Android 7.0. In order to replicate the conditions of the training dataset, we used a ELP-USB100W04H-RL35 1.0 Megapixel camera illuminated by 8 near-infrared LEDs, connected and powered via an On-The-Go USB cable.

We used the Android Face API to crop and normalize faces, and a Tensorflow Lite framework to compile the MobileNet model previously trained on a laptop. The prototype records the subject's face during 10 s and make the decision over all the frames processed in the time window. In our experimental set-up, we achieved a frame rate of around 5–10 frames per second. Improvements of this performance would be possible with optimized code and the usage of reduced MobileNet networks (with a consequent decrease in accuracy).

5 Experimental Results

DROZY was selected for training purposes as it provides objective ground truth on sleep deprivation. We trained the model using data from the first and second PVT only (respectively labeled as "rested" and "sleep deprived") of 11 subjects randomly selected (every subject but numbers 3, 8 and 14). We applied knowledge transfer [23] to a 1.0 MobileNet-224 pre-trained on Inception network. (by freezing all but the last layer, for a total of 2004 free parameters).

5.1 Intraindividual Model

First we evaluated the intraindividual performance by training a model on each subject. Data was split into training (80%), validation (10%) and test (10%) subsets. Accuracy scores are very high (more than 99% both for image and video classification) but the models are fit to one subject and are not translatable.

5.2 Interindividual Model

For evaluating interindividual performance, we randomly selected 1000 frames per subject and class for a total of 22,000 images, which were divided into two subsets: 85% for training and 15% for validation. The remaining frames, which were not included in the training were reserved for test purposes. On 236,981 frames, image classification got an accuracy score of 90.48%. Video classification improved this result, with an accuracy score of 93.48% over 966 sequences.

Evaluating Videos from the Third PVT. The samples of the class "sleep deprived" used for training were all from the second PVT, the subjects being awake for 20 h. Data from the third PVT (more than 28 h of sleep deprivation) was not seen by the model during the training phase, and should be classified as "sleep deprived". The model trained on the first two PVTs was tested on the videos from the third PVT (for the same 11 subjects). The videos were cropped in the same way as we did for the other two classes, obtaining 100,561 images and 485 video sequences. When applying classification, accuracy score was 77.82% on images and 91.13% on video sequences.

Extending to Subjects Not Previously Seen by the Model. The next test consisted in classifying images and videos from the three subjects who were not included in the training phase (subjects 3, 8 and 14) and who had not been seen by the model yet. Videos from the first PVT are labeled as "rested", and videos from the second and third PVTs are labeled as "sleep deprived". The resulting dataset had 124,289 images and 486 video sequences. Accuracy score reached 82.9% on images and improved to 86.21% on videos.

The final accuracy score over the entire DROZY dataset is 85.68% on 461,831 images and 91.07% on 1,937 videos. Confusion matrices in Fig. 2 show that the system is really good in avoiding false positives and gets to identify most of the sleep deprived sequences.

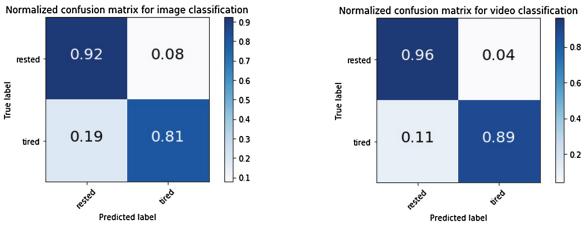

Fig. 2. Confusion matrices for image (left) and video classification (right) on DROZY

Applying the Model to Real World Data. Finally, we evaluated the application of the system to the classification of sequences from real drivers, recorded under naturalistic conditions with Toucango devices. We used the model previously trained on the DROZY dataset. applying the system to this data, the system got an accuracy score of 68.7% for image classification on 9473 frames, and an accuracy score of 72.73% for video classification on 44 sequences.

Confusion matrices in Fig. 3 show that the system performs well when classifying sequences of rested drivers but struggles with tired drivers. This is coherent with the limitations of this new dataset, described in Sect. 4.1; as the conditions of sleep deprivations are less severe, the system only classifies a portion of them as "sleep deprived".

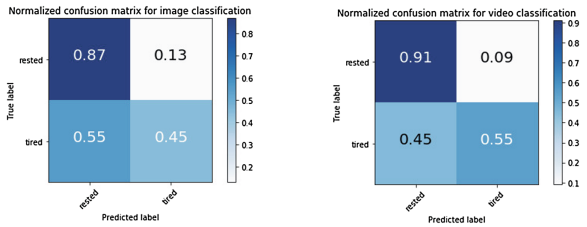

Fig. 3. Confusion matrices for image (left) and video classification (right) on Toucango

Facial features for sleep deprivation differ to those for fatigue. Therefore, in order to improve results on the Toucango dataset, we should look toward the retraining of the system using more videos from drowsy drivers specifically.

6 Discussion, Applications and Future Work

We propose a new way to evaluate fitness-to-drive by detecting sleep deprivation. The decision is made by classifying every frame from a video into two classes ("rested" and "sleep deprived") and averaging the results over the length of the sequence. Intraindividual accuracy is very good, and interindividual performance observed on several experiments over a controlled dataset are satisfactory. Early applications to real world data are also encouraging.

We imagine two separate applications for its usage in Advanced Driver Assistance Systems (ADAS) in order to evaluate subject's fitness to drive (FTD). First one is implementing a FTD test that needs to be passed by the driver as a condition to start the trip. The subject would have to record himself with the ADAS (or with a smartphone). The system would estimate then if he is well rested or if he is too tired to drive. Another possible application is real-time drowsiness detection. In this case, this estimation should be combined among other well-known parameters (eye closure, blinking frequency, yawning, etc.). The feasibility of the system have been demonstrated with our Android prototype. A system like this can be easily implemented in a low cost device and work in real-time, which makes possible its deployment in a commercial ADAS.

However, the system has relevant limitations. The database used for training is too small and sleep deprivation conditions are too extreme (more than 20 h). In addition, the dataset lacks of subject diversity regarding ethnicity, age, facial hair, physical characteristics or glasses usage.

Future work should try to assess these limitations by retraining the system on a much larger dataset including a wider variety of sleep deprivation conditions but also more subject diversity. It should also evaluate the usage of 3D Convolutional Neural Networks in order to include spatio-temporal features.

References

1. Taheri, S., et al.: Short sleep duration is associated with reduced leptin, elevated ghrelin, and increased body mass index. PLoS Med. **1**(3), e62 (2004). Ed. Philippe Froguel. PMC
2. Durmer, J.S., Dinges, D.F.: Neurocognitive consequences of sleep deprivation. Semin. Neurol. **25**(1), 117–129 (2005). Copyright 2005 by Thieme Medical Publishers Inc, 333 Seventh Avenue, New York, NY 10001, USA
3. Peters, R.D.: Effects of partial and total sleep deprivation on driving performance. US Department of Transportation, February 1999
4. Metzgar, C.: Moderate sleep deprivation produces impairments in cognitive and motor performance equivalent to legally prescribed levels of alcohol intoxication. Prof. Saf. **46**(1), 17 (2001)

5. Drowsy Driving NHTSA reports. https://www.nhtsa.gov/risky-driving/drowsy-driving. Accessed 02 June 2017
6. Masten, S.V., Stutts, J.C., Martell, C.A.: Predicting daytime and nighttime drowsy driving crashes based on crash characteristic models. In: 50th Annual Proceedings, Association for the Advancement of Automotive Medicine, Chicago, October 2006
7. Sundelin, T., et al.: Cues of fatigue: effects of sleep deprivation on facial appearance. Sleep **36**(9), 1355–1360 (2013)
8. Sahayadhas, A., Sundaraj, K., Murugappan, M.: Detecting driver drowsiness based on sensors: a review. Sensors **12**(12), 16937–16953 (2012)
9. Ekman, P.: Facial action coding system (FACS). A human face (2002)
10. Vural, E., Cetin, M., Ercil, A., Littlewort, G., Bartlett, M., Movellan, J.: Drowsy driver detection through facial movement analysis. In: Lew, M., Sebe, N., Huang, T.S., Bakker, E.M. (eds.) HCI 2007. LNCS, vol. 4796, pp. 6–18. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-75773-3_2
11. Weng, C.-H., Lai, Y.-H., Lai, S.-H.: Driver drowsiness detection via a hierarchical temporal deep belief network. In: Chen, C.-S., Lu, J., Ma, K.-K. (eds.) ACCV 2016. LNCS, vol. 10118, pp. 117–133. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-54526-4_9
12. Shih, T.-H., Hsu, C.-T.: MSTN: multistage spatial-temporal network for driver drowsiness detection. In: Chen, C.-S., Lu, J., Ma, K.-K. (eds.) ACCV 2016. LNCS, vol. 10118, pp. 146–153. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-54526-4_11
13. Huynh, X.-P., Park, S.-M., Kim, Y.-G.: Detection of driver drowsiness using 3D deep neural network and semi-supervised gradient boosting machine. In: Chen, C.-S., Lu, J., Ma, K.-K. (eds.) ACCV 2016. LNCS, vol. 10118, pp. 134–145. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-54526-4_10
14. Lyu, J., Zejian Y., Dapeng C.: Long-term multi-granularity deep framework for driver drowsiness detection. arXiv:1801.02325 (2018)
15. Dwivedi, K., Biswaranjan, K., Sethi, A.: Drowsy driver detection using representation learning. In: 2014 IEEE International Advance Computing Conference (IACC) , 21–22 February 2014
16. Bradski, G., Adrian, K.: OpenCV. Dr. Dobb's journal of software tools 3 (2000)
17. Schroff, F., Dmitry, K., James, P.: Facenet: a unified embedding for face recognition and clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2015)
18. Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H.: Efficient convolutional neural networks for mobile vision applications. CoRR, Mobilenets (2017)
19. Reed, L.J., Berkson, J.: The application of the logistic function to experimental data. J. Phys. Chem. **33**(5), 760–779 (1929)
20. Massoz, Q., Langohr, T., Francois, C., Verly, J.G.: The ULG Multimodality Drowsiness Database (called DROZY) and Examples of Use, WACV (2016)
21. http://www.innov-plus.com/en/toucango/
22. García-García, M., Caplier, A., Rombaut, M.: Driver head movements while using a smartphone in a naturalistic context. In: 6th International Symposium on Naturalistic Driving Research, The Hague, Netherlands, vol. 8, Jun 2017
23. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Trans. Knowl. Data Eng. **22**(10), 1345–1359 (2010)