Advanced Motion-Tracking System with Multi-Layers Deep Learning Framework for Innovative Car-Driver Drowsiness Monitoring

Francesca Trenta[1], Sabrina Conoci[2], Francesco Rundo[2], Sebastiano Battiato[1]
[1] IPLAB – Department of Mathematics and Computer Science, University of Catania, 95125 Catania, Italy
[2] STMicroelectronics ADG – Central R&D, 95121 Catania, Italy

Abstract— **Recently, the ability to monitor driver drowsiness has attracted a great deal of attention in the automotive industry, in order to prevent the risk due to an inadequate driver psycho-physical state. Specifically, the research effort has focused on the study of the physiological signals to assess the attention level. The main idea consists in verifying the drowsiness level through analyzing the Heart Rate Variability (HRV). The HRV allows to understand the activity of the autonomic nervous system that regulates a series of unconscious and involuntary activities (e.g. the heartbeat, the blood pressure). The HRV is traditionally obtained from electrocardiography (ECG) even though the photoplethysmography (PPG) signal has been proposed as valid alternative to ECG in order to overcome some limitations derived from it. For the above reasons, we analyzed the skin micro-movements and changes in facial color due to blood circulation quite indistinguishable with naked eye in order to extract facial landmarks and to reconstruct PPG signal. The results we obtained by validation confirmed the correlation between the PPG signal detected by sensors and the reconstructed PPG signal from facial landmarks.**

I. INTRODUCTION

Our research investigation is focalized on the analysis of drowsiness level through cardiac activity. The HRV permits to measure the variance of intervals between heartbeats therefore it represents a valid indicator of cardiovascular system activity to obtain information about different physiological states of a human being. As already stated, the PPG signal is often used to find a correlation between the heart rate and the level of attention of a subject. Specifically, the pulse transit time is correlated with the changes in the blood pressure and the stress level. In addition, it has been noted that the blood pulse causes the color changing on different spots of the skin. According to this, the proposed approach uses a low frame rate video camera to capture a video sequence of a subject to analyze facial micro-movements in order to evaluate pulse transit time and reconstruct PPG signal using facial landmarks. The results confirmed that the proposed method has the potential to be used in automotive field.

The remainder of the paper is structured as follows. In Section II we propose a detailed description about PPG signal and the devices to measure it. In Section III we present the prior art. An overview about our proposed pipeline and Long Short-Term Memory neural networks is discussed in Section IV. Section V explains how the experiments are performed. Section VI presents the results. Finally, in Section VII we discuss about advantages of our method and future works.

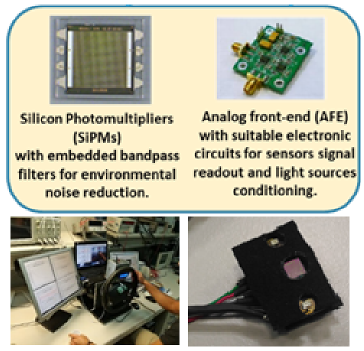

Fig. 1. SiPM detector filter and PPG probe

II. THE PPG SIGNAL

PhotoPlethysmoGraphy (PPG) is a non-invasive technique which is becoming increasingly popular not only to provide information on the arterial stiffness and elasticity [1] but also to measure the blood volume changes in the microvascular bed of human tissue [2],[3]. During the cardiac cycle, the heart pumps blood to the organs and tissues of the body. This pressure pulse causes a volumetric change of the heart that it can be measured by illuminating a patch of the skin with a light-emitting diode (LED) source and evaluating the amount of light either transmitted or reflected to a photodiode [4]. Each cardiac cycle appears as a peak in the resulting signal. The PPG signal has been considered for use in the automotive field e.g. in order to gain useful information on the behaviour and drowsiness of the driver in various situations which may occur in a vehicle [5]. Although PPG signal represents an effective solution in order to measure the HRV and to assess the physiological state of a subject, PPG sampling pipeline shows some noise or signal artifacts (e.g. signal distortion, sensors issue, random noises) also after a careful filtering of the raw signal. For the above reason, a PPG compliant waveforms recognition mechanism is needed in order to improve the robustness of the data processing performed from PhotoPlethysmoGraphy data. In the following section we described a robust PPG signal sampling pipeline.

A. The PPG Detection System

The PPG probes used for the measurements reported in this paper and shown in Figure 1 are based on the use of large area n-on-p Silicon Photomultipliers (SiPMs) [4]. The SiPMs

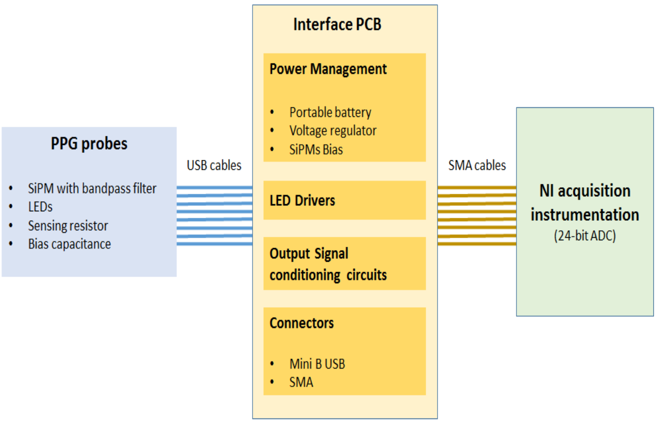

Fig. 2. PPG Signal Sampling pipeline: System Architecture

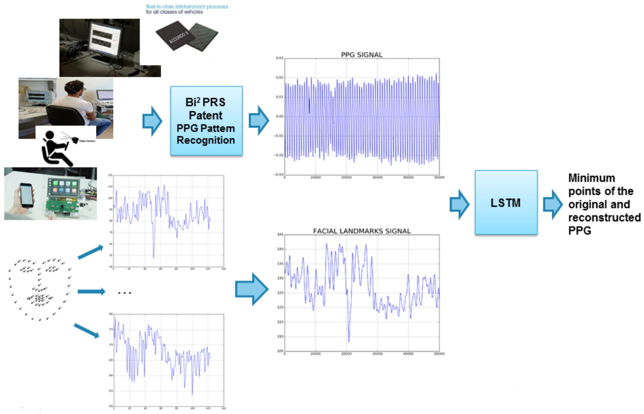

Fig. 3. The LSTM Pipeline

have a total area of 4.0x4.5 mm^2 and 4871 square microcells with 60 m pitch. The devices have a geometrical fill factor of 67.4% and are packaged in a surface mount housing (SMD) with 5.1x5.1 mm^2 total area [1],[4]. A Pixelteq dichroic bandpass filter with a pass band centered at 542 nm with a Full Width at Half Maximum (FWHM) of 70 nm and an optical transmission higher than 90% in the pass band range was glued on the SMD package by using a Loctite 352TM adhesive. With the dichroic filter at 3V-OV the SiPM has a maximum detection efficiency of about 29.4% at 565 nm and a PDE of about 27.4% at 540 nm (central wavelength in the filter pass band). Furthermore it has been shown how the dichroic filter reduces the absorption of environmental light of more than 60% in the linear operation range of the detector operating in Geiger mode above its breakdown voltage ($\sim$ 27V). OSRAM LT M673 LEDs in SMD package emitting at 529 nm and based on InGaN technology have been used as optical light sources [4]. The LEDs have an area of 2.3x1.5 mm^2, viewing angle of 120°, spectral bandwidth of 33 nm and typical power emission of a few mW in the standard operation range. In Figure 2 is reported the block scheme of the system architecture.

III. RELATED WORKS

In the literature there are several approaches about the drowsiness detection systems based on PPG signals. In [6], the authors proposed a wireless base-station system in order to record the driver physiological signals. The purpose consists in obtaining information about fatigue level using HRV derived from PPG signal. In [7], a new method to evaluate car driver condition using PPG signal was investigated. Specifically, the authors utilized the Low and High Frequency of PPG detected at fingers and earlobes. The authors of [8] present a system that combines the wavelet packet transform (WPT) in order to extract features of PPG signal and a functional-link-based fuzzy neural network (FLFNN) to find the peaks. In [9] the authors provide an innovative method for HRV detection from PPG signal. Specifically, the authors present an algorithm for peak detection in corrupted PPG signals by using sensors placed on a smartwatch. The purpose consists in obtaining a reliable peak detection in order to extract relevant information about HRV from a corrupted PPG signal derived from the movements of the wrist. In [10], the authors reached effective results by using a low power wireless PPG sensor, the innovative approach consists in a working memory test to correlate HRV, extracted by a new sensor. The described methods confirm the effectiveness to determine the driver drowsiness by using PPG signal. Specifically, the drowsiness state can be determined through the Heart Rate Variability (HRV) that represents the frequency (Fourier Spectrum) of the PPG minimum points. As reported, one of the main advantage of using PPG signal is characterized by the fact that it is not necessary to put both hands on the steering wheel to acquired it, instead of ECG signal. Despite the approaches are very promising, they present some limitations. In particular, the described methods are based on highly complex systems (PPG/ECG sensors on steering wheel etc.). Moreover, they require that the car driver has to put his hands over the car steering or in other parts of the car in which the sensors are placed. Consequently, in order to prove the robustness of our algorithm, it is necessary to address the case when the car driver does not put his hands on the PPG sensor over the steering wheel. For this reason, we implemented a method to reconstruct PPG signal from facial landmarks in order to analyze the driver condition though HRV signal, as described in the following sections.

IV. THE PROPOSED PIPELINE

The proposed method is based on a mixed Long Short-Term Memory (LSTM) – Convolutional Neural Network (CNN) system [11]. In Figure 3 and 4 we present the LSTM and CNN pipeline, respectively. Our purpose is to obtain information about the car driver drowsiness by calculating HRV frequency domain of the driver's heart rate time-series. As discussed, PPG signal represents a reliable solution in order to measure the HRV. The PPG signal can be detected by using a LED that illuminates the skin and a photodetector that measures the intensity of the transmitted or reflected light in order to evaluate the changes in the blood volume.

Differently from other approaches, our method consists in detecting and extracting facial landmarks with the aid of Computer Vision techniques to reconstruct PPG signal, instead of detecting it through specific devices. The main idea is based on the concept of Video Magnification [12] that can reveal face tiny movements due to blood circulation and difficult to observe with the naked eye. In order to take information about driver condition, it is useful to estimate the blood pressure that causes these movements by analyzing HRV signal through PPG detection. For this purpose, we decided to extract facial landmarks from images of the face of car driver. The landmarks points are very useful for facial analysis task because we can obtain a lot of information about fundamental facial components (e.g nose, mouth, eyes, etc.) [13]. We used these points to calculate the landmarks pixel intensities and their variations for each video frame, that we previously recorded, in order to define the landmarks time-series as input for LSTM pipeline. In fact, the aim of LSTM neural networks is to detect dependencies in sequential data (time-series), in particular their ability to understand long term dependencies makes them very powerful for sequence learning tasks. Furthermore, we implemented a CNN model in order to classify accurately the driver physiological state (wake or drowsy) by using face images in order to validate the LSTM results, as described in the following sections.

A. *Mathematical Background of LSTM*

Long Short-Term Memory (LSTM) neural networks are a particular type of Recurrent Neural Network (RNN) that show good performance in time-series prediction. They represent an improvement of the classic RNN model because they are able to evaluate the hidden nonlinear correlations between input data [14]. A classical LSTM main unit is composed by a cell, an input/ output/ forget gate. The cell is able to store (recall) values over arbitrary time intervals while above gates are able to manage the input/output data flow of the cell. Firstly, the input of LSTM is a vector that represents the old memory and it passes through the forget layer that decides what information can throw away from the cell state using equation (1). Then the model decides what new information will be stored in the cell state using equations (2) and (3) to merge the new and the old memory.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \tag{1}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \tag{2}$$

$$\widetilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C), \tag{3}$$

The next step updates the old cell state C_{t-1} into the new cell state, called C_t, as reported by equation (4).

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t, \tag{4}$$

Finally, the LSTM output is generated by merging the previous output, the input and the bias vector according to the following equations (5) and (6).

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \tag{5}$$

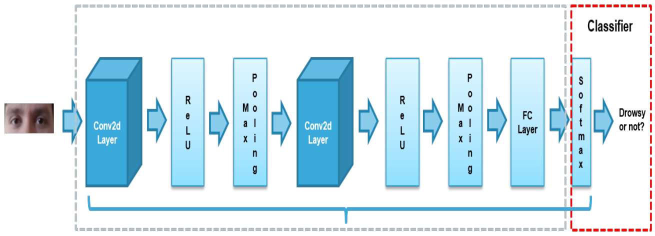

Fig. 4. The CNN pipeline

$$h_t = o_t * \tanh(C_t), \tag{6}$$

where:

- W_f, W_i, W_C, W_o represent the LSTM weights
- b_f, b_i, b_C, b_o are the bias
- C_t is the cell state
- σ is the sigmoid function

V. EXPERIMENTS

In this paper, we provide two datasets to perform our LSTM–CNN pipeline. One of datasets consists of PPG signal samples, and the other consists of images of the face of car driver. The data acquisition was conducted in accordance with the Helsinki Declaration of 1975. All patients provided written informed consent before enrollment. The study was approved by the local Ethical Committee of University of Catania. Then, we proceeded to collect 71 patients/drivers having different genders (male and female), ages (between 20 and 70 years old) and pathologies (we collected healthy subjects and sick ones with different issues such as cardiac problems, hypertension, diabetes, etc.). We performed sampling of PPG signals in two different Drowsiness status:

- Wakeful: Under physiologists suggestions, we emulate a full wakeful scenario confirmed by simultaneously ECG signal sampling which showed Beta waves meaning high wakeful brain activity;
- Drowsy: Under physiologists suggestions, we emulate a full drowsy scenario confirmed by simultaneously ECG signal sampling which showed Alpha waves meaning clear drowsiness of the monitored subject;

We performed 5 minutes of PPG/ECG sampling for each scenario. At the same time, we proceeded to record a video of the face using a Full-HD video-camera with 25fps in order to validate our method based on the PPG reconstruction by using the facial landmarks, as described in the following section. For each experiment, we used Python as the software framework, running on a PC i5 quad-core. Moreover, the datasets for these experiments are further subdivided into a dataset for actual training (70%) and a dataset for validation (15%) and testing (15%).

A. *CNN Block Description*

We decided to implement ad-hoc CNN model in order to track and learn facial expressions of the car driver. The final purpose is to improve the drowsiness detection robustness of the proposed pipeline. After acquiring the PPG signal for a preliminary system calibration and for real time continuous

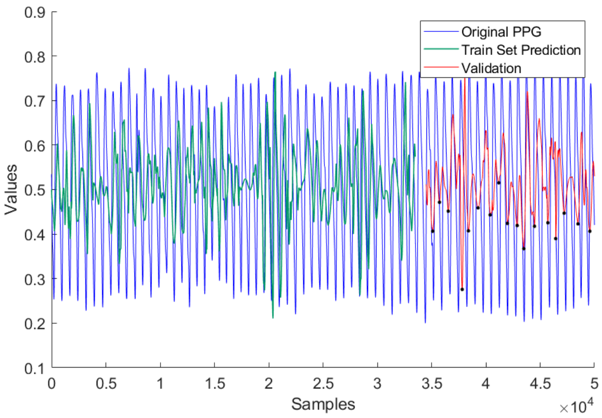

Fig. 5. Validation set (red) minimum points detection

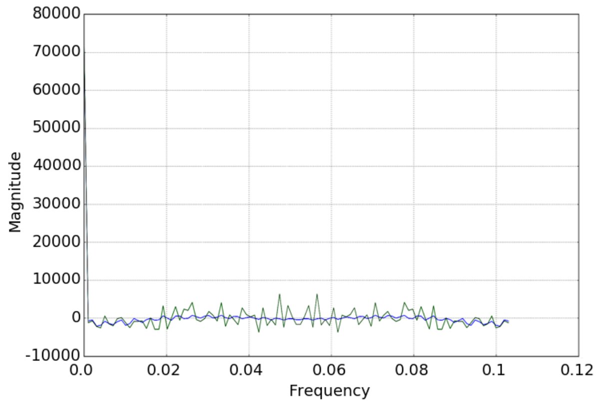

Fig. 6. Correlation between FFT Spectrum of the original PPG minimum points (Blue) vs reconstructed PPG minimum points (Green).

learning, we used the classification results to validate LSTM prediction. The proposed CNN architecture is based on LeNet architecture [15]. It consists of 2 convolutional layers, as shown in Figure 4. The training of the model is performed using a batch of size 32, an initial learning rate of 0.001. Moreover, we built our model along Adam optmizer. We used 32 neurons in the hidden layers and 2 output neurons, as the target dataset has 2 classes. The experimental results reported that our model classifies accurately the images. In particular, we reached an accuracy of 80%, but some validation results reported an accuracy close to 90%.

B. LSTM Block Description

In order to develop our LSTM pipeline, we recorded a video of the face of a subject with a low frame rate video camera in order to demonstrate that good results can be obtained without the aid of sophisticated devices. We extracted video frames and detected facial landmarks which can be found around the nose, the mouth and the eyes by using the detector included in dlib library that is an implementation of [13]. We choose original PPG signal sampled by car sensor, placed on the steering wheel, as target data and facial landmarks time-serie as input data. All values were scaled in the range (0.2, 0.8), with classical MinMaxScaler algorithm. The training is performed using 256 neurons, a batch of size 128 and an initial learning rate of 0.001. Moreover, to prevent overfitting, we used a dropout ratio of 0.2. In the final step, we extracted the minimum points of the PPG original signal and the PPG reconstructed signal (from LSTM) in order to verify the robustness of reconstructed PPG. For this purpose we calculated the distance of these points and compared the distance of the original PPG minimum points to the distance of the reconstructed PPG minimum points. In Figure 5, the detected minimum points are shown.

VI. RESULTS

In order to prove the effectiveness of the proposed method, we calculated the correlation between Fast Fourier Transform (FFT) spectrum of the original PPG minimum points and FFT spectrum of the reconstructed PPG minimum points. In Figure 6 we provided the result. Moreover, we have tested our proposed approach by using the reconstructed PPG time-series for computing HRV dynamic in order to evaluate the actual drowsiness of the tested subject (car driver) included in our dataset. The results showed very promising performance of the proposed approach as we were able to discriminate drowsy subject from wakeful ones with accuracy near to 100% which coincides with the average performance obtained by similar pipeline reported in scientific literature [6],[16].

VII. CONCLUSIONS

We proposed a sensor-less method in order to take information about drowsiness state of a car driver. The experimental results confirm that exists a correlation between PPG obtained by using sensors and the PPG reconstructed by using facial landmarks. Compared to other approaches based on the analysis of bio-data to analyze HRV, our proposed method does not require sophisticated and expensive devices to record PPG signal, we can construct it from facial landmarks. In fact, we present a valid alternative to obtain it and to overcome the principal limitation of other works (e.g. the driver has to put his hands over the steering wheel to record PPG signal). Our future works will focalize on collecting a large amount of face image. In particular, our goal is to collect images of subjects with different characteristics in terms of gender, age, ethnicity, etc. in order to further validate the proposed method. Some further results will be reported at conference time.

VIII. ACKNOWLEDGMENTS

The research activity leading to the results shown in this paper was partially funded from the National Funded Programme 2014-2020 under grant agreement n. 1733, (ADAS+ Project).

References

[1] V. Vinciguerra, E. Ambra, L. Maddiona, S. Oliveri, M.F. Romeo, M. Mazzillo, F. Rundo, G. Fallica, "Progresses towards a processing pipeline in photoplethysmogram (PPG) based on SiPMs", *In European Conference on Circuit Theory and Design (ECCTD)*, Catania, Italy, 46 September 2017, pp. 15.

[2] F. Rundo, S. Conoci, A. Ortis, S. Battiato, An Advanced Bio-Inspired PhotoPlethysmoGraphy (PPG) and ECG Pattern Recognition System for Medical Assessment, *Sensors*, vol. 18(2), 2018, pp. 405.

[3] F. Rundo, A. Ortis, S. Battiato, S. Conoci, Advanced Bio-Inspired System for Noninvasive Cuff-Less Blood Pressure Estimation from Physiological Signal Analysis, *Computation*, vol. 6(3), 2018, pp. 46

[4] M. Mazzillo, L. Maddiona, F. Rundo, A. Sciuto, S. Libertino, S.A. Lombardo, G. Fallica, Characterization of SiPMs With NIR Long-Pass Interferential and Plastic Filters, *IEEE Photonics Journal*, vol. 10, 2018, pp. 1-12.

[5] F. Rundo, F. Trenta, S. Conoci, S. Battiato, Image Processing Method, Corresponding System, Vehicle And Computer Program Product, IT Patent Nr. 102019000000133, 7 January 2019.

[6] H-S. Shin, S-J. Jung, J-J. Kim, W-Y. Chung, Real time car drivers condition monitoring system, *Sensors*, 2010, pp. 951-954.

[7] S. Koh, B.R. Cho, J.-I. Lee, S.-O. Kwon, S. Lee, J.B. Lim, S.B. Lee, H.-D. Kweon, "Driver drowsiness detection via PPG biosignals by using multimodal head support", *in 4th International Conference on Control, Decision and Information Technologies*, Barcelona, Spain, 2017, pp. 383-388.

[8] N. N. Sari, Y. Huang, "A Two-Stage Intelligent Model to Extract Features from PPG for Drowsiness Detection", *in 2016 International Conference on System Science and Engineering (ICSSE)*, Puli, Taiwan, 7-9 July 2016, pp. 1-2.

[9] T. Bhowmik, J. Dey, V. N. Tiwari, "A novel method for accurate estimation of HRV from smartwatch PPG signals", *in 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Seogwipo, South Korea, 11-15 July 2017, pp. 109-112.

[10] Y. J. Xu, F. Ding, Z. Wu, J. Wang, Q. Ma, K. Chon, E. Clancy, M. Qin, Y. Mendelson, N. Fu, S. Assad, S. Jarvis, X. Huang, "Drowsiness control center by photoplythesmogram", *in 38th Annual Northeast Bioengineering Conference (NEBEC)*, Philadelphia, PA, USA, 16-18 March 2012, pp. 430-431.

[11] Y. LeCun, Y. Bengio, G. Hinton, Deep Learning, *nature*, vol. 521(7553), 2015, pp. 436.

[12] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, W.T. Freeman, Eulerian Video Magnification for Revealing Subtle Changes in the World, *ACM Transactions on Graphics*, vol. 31(4), 2012.

[13] V. Kazemi, J. Sullivan, "One millisecond face alignment with an ensemble of regression trees", *in 2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, 23-28 June 2014, pp. 1867-1874.

[14] S. Hochreiter, J. Schmidhuber, Long Short-Term Memory, *Neural Computation*, vol. 9(8), 1997, pp. 1735-1780.

[15] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of IEEE*, vol. 86(11), 1998, pp. 2278-2324.

[16] Y.-P. Huang, N. N. Sari ; T.-T. Lee. "Early Detection of Driver Drowsiness by WPT and FLFNN Models", *in IEEE Proceedings of 2016 IEEE International Conference on Systems, Man, and Cybernetics, SMC*, Budapest, Hungary, October 9-12, 2016, pp. 463-468.