ROBUST REAL-TIME DRIVER DROWSINESS DETECTION SYSTEM FOR HETEROGENEOUS LIGHTNING CONDITIONS

Leire Varona, Juan Diego Ortega, Peter Leškovský, and Marcos Nieto

Vicomtech-IK4, Paseo Mikeletegi 57, Donostia-San Sebastián, Spain

{lvarona, jdortega, pleskovsky, mnieto}@vicomtech.org

Abstract: Driver Fatigue Detection (DFW) has become the most active research area in the field of Advanced Driver Assistance Systems (ADAS). Its robust performance under real driving conditions requires the use of Infra-Red illumination. Direct and indirect IR illumination systems represent the current options for manufacturers and OEM, according to comfort and ergonomics. Such variety of illumination directly affects the performance of computer vision algorithms, which tend to be sensitive to the visual appearance of faces and eyes in the images. In this work we present a robust DFW strategy that filters images and works seamlessly for direct and indirect IR. Extensive tests in a wide range of scenarios (day and night-time, presence of glasses, laboratory and field videos) show the promising results of this system.

Keywords: ADAS, computer vision, fatigue detection.

1 INTRODUCTION

The increasing number of vehicle users in the last 10 years had as a consequence a direct increase in the number of vehicle accidents. That is the main reason why Advanced Driver Assistance Systems (ADAS) peaked in the last few years. The principal function of these systems is providing safety and easiness to the driver, avoiding distractions to remain focused on the road.

One of the major facts of influence in vehicle crashes is fatigue, which represents 20% of the annual registered accidents in the world, according to World Health Organization (WHO). Driver Fatigue Warning systems (DFW) emerged to solve this problem, with the main function to detect signs of fatigue while driving, so that it is possible to activate a warning alarm and suggests to rest.

Considering the different sensors observing the car motion or the driver itself, machine learning and computer vision methods have the largest influence on ADAS. Most of the current out-of-box systems use one or more cameras with extra lighting to track the driver´s face and analyze the fatigue level. The main problem for a camera based systems is the illumination, as it is different during day-time and night-time, rainy or sunny day. Especially in sunny days, strong contrast and moving shadows hinder successful detection and analysis. At night, artificial illumination must be used to render the driver´s face visible.

To solve this problem, this paper proposes an efficient DFW system capable of detecting fatigue in professional and non-professional drivers, in both, day and night conditions, using a single camera, situated in the frontal interior part of the vehicle, continuously monitoring the driver and using a non-intrusive lighting method consisting in infrared lighting LEDs which illuminate the driver during night as well as day-time.

2 RELATED WORK

In recent years, several approaches have been tested for real-time driver fatigue warning systems. Only few of them have been implemented in real conditions, as most of them have been research studies. Based on the subject of behavior analysis, the DFW systems can be grouped into following three solutions.

2.1 Vehicle motion analyzing based methods

This approach is based on analyzing driving characteristics, such as jerky motion of the steering wheel, which may suppose that the driver is correcting the direction of the vehicle inadequately because of micro-sleeping sequences. These systems are highly limited by the drivers driving style, like sporty, novel or professionally [1] and [2].

2.2 Physiologically based methods

This set of DFW is based on physiological signals, such as pulse rate or Electroencephalography (EEG). However, even though these methods are very precise, it is not possible to develop them as most of them are either intrusive methods or they require wearing special helmets (EEG) while driving [3]. Therefore, these systems are not comfortable for the driver and they may even negatively interfere with their driving.

2.3 Computer vision based methods

Finally, DFW methods based on computer vision, suppose a clear advantage to these kinds of systems because of being non-intrusive and of reduced size. It is possible to divide them into two subsets.

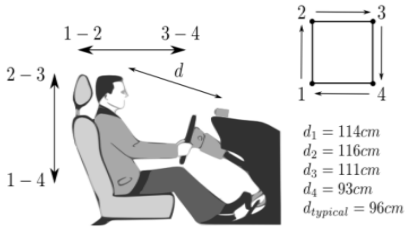
The first subset focuses on the mode driving is executed, based on the road images captured from the car during driving. These are for example Lane Departure Warning Systems or Collision Avoidance Systems [4]. Similarly as it was mentioned in section 2.1, it is highly dependent on the different driver’s driving style, like sporty, novel or professional. However, this system combined with the second subset of computer vision based DFW, might strengthen the whole purpose of the DFW.

The second subset analyses on the visible driver's physical signs [5]. Some of these methods are based on head decays to determine drowsiness, others focuses on yawning, or, like the one presented in this paper, are based on blink duration, considering the time in which the driver has his eyes closed.

To develop a computer vision based DFW it is necessary the use of cameras, either to monitor the road or the driver. Even though there are several methods based on stereo cameras, the use of them implies bigger size of the system, which is not appropriate to install in a vehicle because it might discomfort the driver. In this paper we focused on a solution based on one single camera using infrared illumination (IR). The use of IR increases illumination on night-light conditions without disturbing the driver.

3 DFW OVERVIEW

To develop our computer vision based DFW it was required the employment of a camera. In this case it has been used two different cameras, both low cost and equipped with IR illumination. One uses direct IR illumination and the other indirect. The camera is located in the interior frontal zone of the vehicle, monitoring directly the driver. Figure 1 shows the distance and location of the camera inside the vehicle.



(a)

Figure 1: Camera location inside the vehicle

The main purpose of the DFW implemented is to determine In this procedures are included a face detector based on classification to determine the face position in the image, eye detectors based on classification to determine the position of the eyes in the image and eye state detector to determine if the eye is open or closed. These three main procedures are the most important blocks of the algorithm and are described below:

- **Face Detector**: The face detector used in the algorithm is a Cascade Classifier trained and proportioned by OpenCV. It is based on the Viola & Jones model [6]. The face detector is applied so that it is possible to identify driver absence in the vehicle, as well as the position of the face in the image. This implies a decrease in both computational cost and processing time of the system. The face's Region Of Interest (ROI) will be used as the searching zone for the eye detectors, instead of searching in the whole image, and as it has been mentioned before, it is strictly necessary that the system can achieve a real-time performance, and this means 25 fps of total processing time.
- **Eye Detector**: The eye detector used in the algorithm is a Cascade Classifier trained and proportioned by OpenCV. It uses Haar features to describe the image. The eye detector is applied so that it is possible to identify the position in the image.
- **Eye State Detector**: The eye state detector uses several image processing methods, such as Gaussian filters and the median of intensity of the image, to determine if the eye is open or closed. If closed eyes are detected during a specific time the alarm is activated.

Figure 2 shows the scheme of the algorithm designed, which consists in the detection and tracking of the driver's eyes and face in the image, so that it can be detected the duration of every blink. Thus, if the system detects signs of fatigue in the driver, by any of the characteristics named, it can activate the alarm to alert the driver that a break is needed.

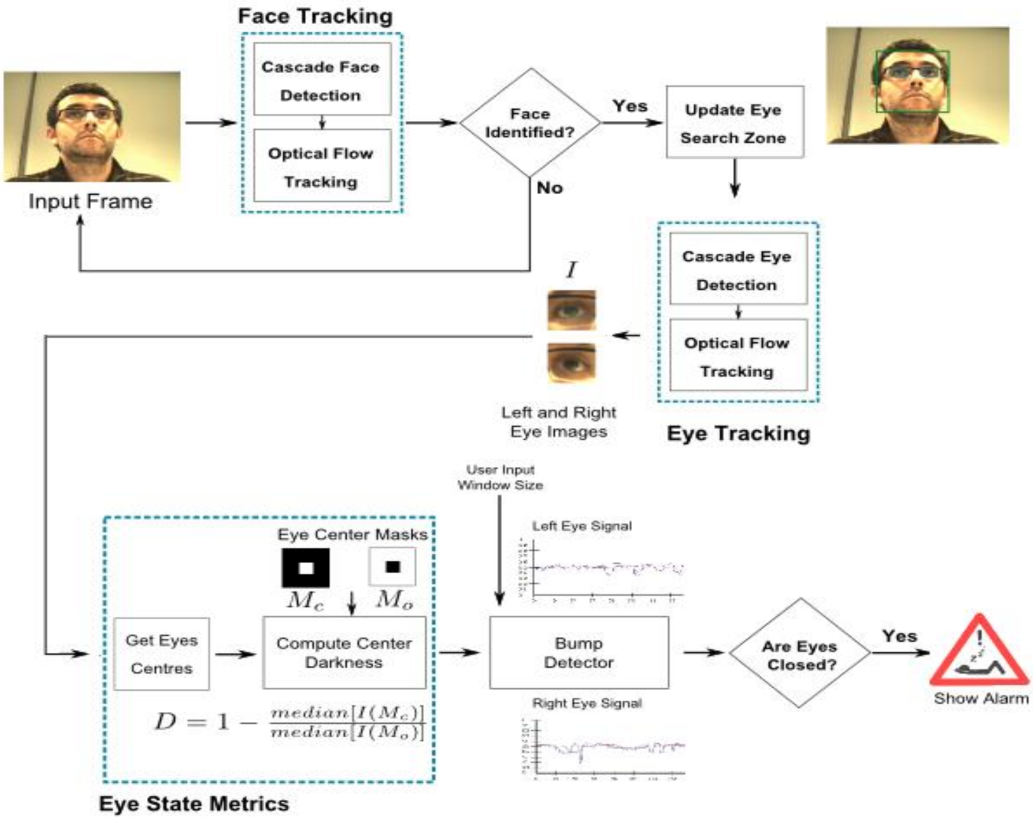


Figure 2: DFW algorithm

3.1 Performance:

The main functionality of the system is to determine the blink duration, so it is important to distinguish between open and close eyes. The use of infrared light in this case is necessary because at night-time light conditions the distinction may be difficult without blinding the driver. There are several ways to place the infrared LEDs on the camera; the direct and indirect are the most common.

Direct infrared light is one whose direction of incidence is parallel to the axis of the camera. One of the cameras used on the development of this project use direct infrared LEDs around the camera. This kind of illumination has as a consequence the bright pupil effect, at night-time, due to the infrared direct incidence on the eye. Figure 3 illustrates an example of this characteristic.

However, this problem can be solved with the so called indirect infrared light, because it incides obliquely on the eye. Due to this characteristic, the bright pupil effect disappears from the image, making them black as the ones on daytime.

(a) (b)

Figure 3: IR illumination at night-time (a) Direct, (b) Indirect

In a former approach of the system [4], the algorithm was designed to use SVM-HOG detectors for the eyes detection. HOG features are supposed to be stable to different lightning conditions, as it is based on the histogram of gradients in the images. However, the bright pupil effect in the images captured at night, made the HOG to determine opposite gradients in both day and night pictures. The solution given to this problem was to train different classifiers for day and night, and switch them automatically as a function of the lightning conditions.

Moreover, the HOG features of the closed eyes could be easily mixed with any other part of the person, such as its mouth, eyebrows and even the open eyes, so the employment of one classifier was infeasible. As a result, in the former approach of the system, it was necessary the employment of two different classifiers, one for each eye state. The need of training a classifier for each light condition besides open and closed eyes has as a result 4 different classifiers. This implies a higher total processing time of the algorithm.

To simplify the algorithm and reduce the total processing time, but maintaining the quality of the detections and tracking of the driver's eyes, some restrictions has been taken into account, among which include:

- Detected eyes in profile faces won't be taken into account, due to the absence of one of the two eyes in the image and because of the hypothesis that a driver who moves the head is awake and therefore there are no fatigue signs.
- There must be a reasonable distance between each eye, that is, there cannot be superposing detections.
- It is strictly necessary to have face detection before applying the eyes detector because if there is no face detected in the images, it could mean the absence of the driver in the vehicle. These restrictions reduce considerably fake alarms and detections in the image.

3.2 Fatigue Detection algorithm

In order to detect open-close changes in driver’s eyes, the face and eyes are detected using a combination of a cascade classifier and data association tracking algorithm. When the eye regions are extracted, approximate positions of the eye centers are computed as follows: a 2D Gaussian centered mask is applied to the left and right eye images. This emphasizes the central zone of the images as it is more likely that the center of the eye could be founded within the center of the image ROI (Figure 4). Additionally, this helps to prevent that other dark points, such as the eyebrows, are selected as eye centers, which represented one of the major problems in the SVM-HOG approach.

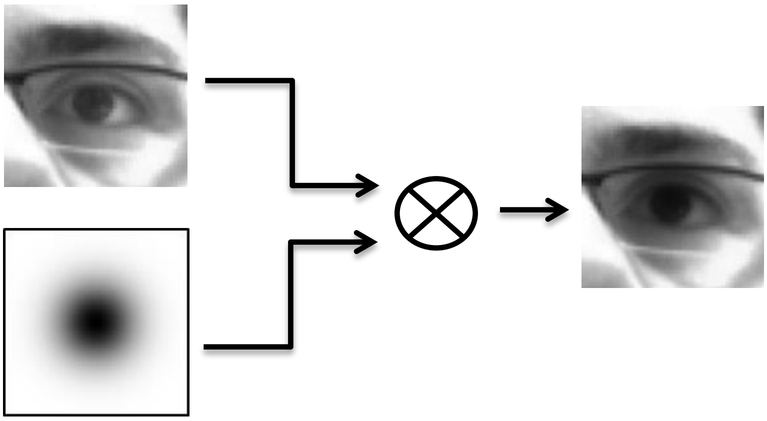


Figure 4: Darkened pupil method

Then the darkest pixel is selected as the center of the eye, since in the image the pupil corresponds to the darkest point, due to the use of indirect illumination in the image acquisition stage. The eye center allows defining two masks with which the eye measurements are perform in order to detect changes between open and close eye states.

(a) (b)

Figure 5: Mask (a) Center Mask M_c, (b) Surrounding Mask M_o

Note that the sizes of the masks are equal to the size of the eye region, and the rectangular center of the mask is placed according to the eye center detection method described before. Therefore the center of the mask does not need to match the center of the picture, but the area in which the pupil is. Finally the median of the image intensities are computed applying the two masks shown in Figure 5.

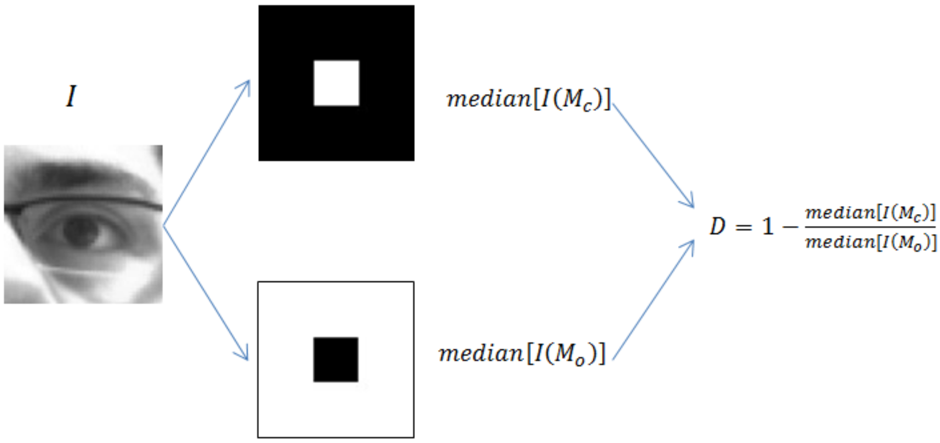


Figure 6: D value calculating method

The value D, shown in Figure 6, is the ratio between the median darkness of the center of the eye and the median of the surroundings. This value is motivated by the fact that the pupil and center of the eye usually is darker than the rest of the eyeball and skin. Therefore the value D is close to 1 when the central region of the mask is darker than the surroundings. This value could discriminate between open and close eyes because in open eye state the eye is darker and occupies almost all the central region of the mask, whereas in closed eye the darker region is reduced and only the eyelid is visible, which is brighter.

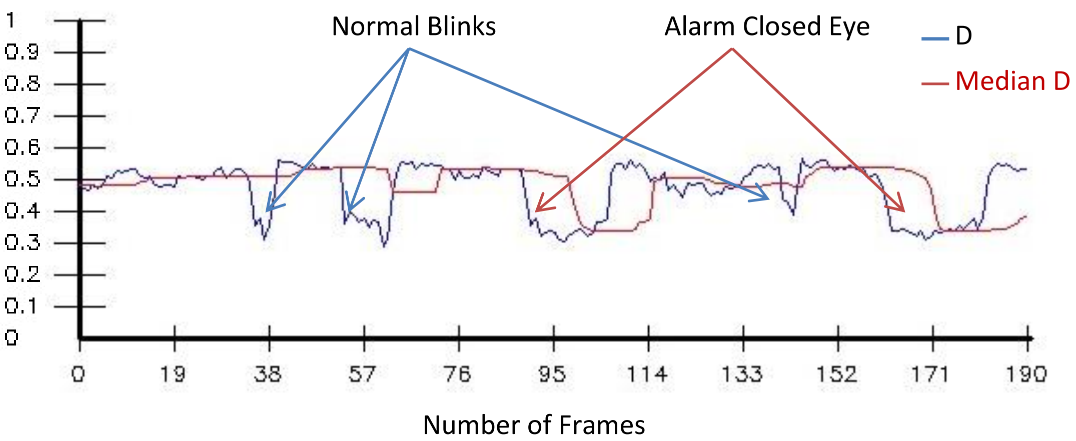
(a) (b)

Figure 7: (a) Open eye and (b) Close eye

This method allows the system to perform correctly under every kind of lighting condition. The indirect illumination facilitates the correct performance of the system, proving it to be more efficient, robust and faster than the former approach.

Besides, the SVM-HOG approach is very limited to different types of eyes, as blue or almond eyes. However, with our approach this limitation is reduced.

It is possible to plot this signal and compute its median. To compute the median of the darkness value, a temporal window has to be defined by the user as input. A bump detector is applied to detect extended periods of closed eyes.



When there is a long period of closed eyes the level of D value will decrease and an alarm is shown when the median of this signal also decreases some value (defined by the user). On the other hand, short periods of closed eyes (blinks) will not affect the median and hence no alarm will be shown. The use of the median of the value D helps to reduce the noise in the signal.

4 RESULTS

To test the algorithm implemented it is necessary to have a dataset as large as possible to prove that the system is stable to different conditions. The dataset used to test the system consist of several videos captured by Datik - Irizar Group and Vicomtech-IK4. The dataset includes laboratory videos and videos captured inside a bus with professional drivers in different situations. All of them have been separated into categories, so that it is easier to prove that the system is stable on different situations.

The dataset used for testing the algorithm is sectioned as follows:

- Direct or indirect IR illumination.
- Lightning conditions: day or night.
- Driving or laboratory video.
- Glasses, sunglasses, or none.
- Camera type.
- IR filter used or none.

This separation helps identifying the stability of the system developed. As it was mentioned in the former section, the algorithm was tested by SVM-HOG and Cascade Classifier method for the eye detection.

From the dataset it has been generated a group of manual annotations to evaluate the system's performance. Since the manual annotation requires plenty time, it has been selected a representative video from every subcategory from the dataset.

The result of the manual annotation is a data file containing every interval of time in which the subject has the eyes opened and closed. All of the events are defined by 3 attributes:

- Type: in this case correspond to Open or Close.
- Frame Start: frame in which the event has started.
- Frame End: frame in which the event finalizes.


```xml
<videoContentDescription>
    <video_data>
        <name>C:/Proyects/FireflyMVFMVU03MTC_xvid.avi</name>
        <frame_rate>30</frame_rate>
        <num_frames>1069</num_frames>
        <width>640</width>
        <height>480</height>
    </video_data>
    <object_list/>
    <event_list>
        <event id="20000" type="Open" frameStart="0" frameEnd="10" annotate="Event n. 1" message="Nothing relevant."/>
        <event id="20001" type="Close" frameStart="10" frameEnd="15" annotate="Event n. 2" message="Nothing relevant."/>
        <event id="20002" type="Open" frameStart="15" frameEnd="34" annotate="Event n. 3" message="Nothing relevant."/>
        <event id="20003" type="Close" frameStart="35" frameEnd="40" annotate="Event n. 4" message="Nothing relevant."/>
        <event id="20068" type="Open" frameStart="1065" frameEnd="1069" annotate="Event n. 69" message="Nothing relevant."/>
    </event_list>
</videoContentDescription>
```


Figure 8: Example of the result of manual annotation

When the algorithm is executed, the results obtained are saved in the same format as the manual annotation, so that it is possible to evaluate them. The evaluation is made by computing several values such as Recall, Precision, Fmeasure, True Positive, False Positive and False Negative rate. The evaluation has been performed considering two types of analysis:

- Event level analysis: this type of evaluation gives results of TP, FP and FN at event level. That is, in a sequence of events it will be denominated TP to event correctly detected, FP no existent detected events, and FN non detected events.

- Frame level analysis: this type of evaluation gives results of TP, FP and FN for each image. That is, every image will be considered as an event itself. Due to this type of evaluation, we can get more reliable results for the algorithm, because in the event level analysis, each frame incorrectly detected can suppose different events and the result is negatively affected when an isolated frame won't affect to the driver's fatigue detection.

The videos selected from the dataset to represent each category are described in the Table 1. There has been set an ID so that in the following sections it is easier to know which video is mentioned. The cameras with which the testing has been made are: PointGrey FireFly MV equipped with direct IR light and Irisbond Chameleon equipped with indirect IR light.

Table 1: Dataset for evaluation

PointGrey FireFly MV: Direct IR illumination							
ID	**Category**				**Hour**	**Frames**	**Events**
ID1	IR	Day	Driving	No glasses	17:41:39	1127	69
ID2	IR	Day	No driving	No glasses	14:24:28	1736	109
ID3	IR	Night	Driving	No glasses	01:12:05	1951	123
ID4	IR	Night	No driving	Glasses	00:24:16	3088	129
ID5	IR	Night	No driving	No glasses	00:28:38	3234	133
ID6	No IR	Day	No driving	No glasses	14:58:32	478	57
Chameleon: Indirect IR illumination							
ID	**Category**				**Hour**	**Frames**	**Events**
ID7	IR	Night	No driving	Glasses	18:00:45	488	17
ID8	IR	Night	No driving	Glasses	18:01:41	255	16
ID9	IR	Night	No driving	No glasses	17:59:12	150	6
ID10	No IR	Night	No driving	Glasses	18:30:41	3267	43

4.1 Event level analysis

Table 2 shows the results obtained in the event level analysis from the videos. The values shown in the table correspond with Recall (R), Precision (P), Fmeasure (F), True Positive (TP), False Positive (FP) and False Negative (FN) rate.

Table 2: Event Level Analysis

	R	**P**	**F**	**TP**	**FP**	**FN**
ID1	0.623	0.597	0.610	43	29	26
ID2	0.624	0.747	0.680	68	23	41
ID3	0.707	0.604	0.652	87	57	36
ID4	0.814	0.905	0.857	105	11	24
ID5	0.880	0.959	0.918	117	5	16
ID6	0.579	0.688	0.629	33	15	24
ID7	0.647	0.733	0.688	11	4	6
ID8	0.875	0.875	0.875	14	2	2

ID9	0.667	1.000	0.800	4	0	2
ID10	0.721	0.62	0.667	31	19	12

Note that in the manual annotation, it has been taken into account frames in which the driver is looking at both sides, not just frontally, even though the system is not implemented to detect eyes in profile faces. However, due to the face detection and tracking, when the eyes are not detected and there is still face detection and tracking, it is possible to say that the driver is not fatigue, in fact is awake enough to pay attention to the road, and probably looking at the rearview mirrors. Additionally, if there is no face detected, it is possible to affirm that the driver is not inside the vehicle.

Overall, table 2 shows that the system gives better results with indirect IR illumination in the event level analysis. However, for the other light conditions the performance is also stable and acceptable, all of them tested in different conditions of ambient lighting and with sunglasses, normal glasses, and without glasses.

4.2 Frame level analysis

Table 3 shows the results obtained in the event level analysis from the videos. The values shown in the table correspond with Recall (R), Precision (P), Fmeasure (F), True Positive (TP), False Positive (FP) and False Negative (FN) rate.

Table 3: Frame Level Analysis

	R	P	F	TP	FP	FN
ID1	0.732	0.949	0.827	825	44	302
ID2	0.760	0.960	0.849	1320	55	416
ID3	0.848	0.922	0.884	1654	139	297
ID4	0.834	0.940	0.884	2575	163	513
ID5	0.876	0.957	0.915	2834	128	400
ID6	0.860	0.901	0.880	411	45	67
ID7	0.795	0.958	0.869	388	17	100
ID8	0.718	0.901	0.799	183	20	72
ID9	0.793	0.915	0.850	119	11	31
ID10	0.848	0.971	0.906	2771	82	496

The results obtained in the frame level analysis are better than the once obtained in the event level analysis. That is because the event level analysis is highly affected with misclassified frames, as an isolated misclassified frame causes two separated events and one bad detected. This failure in the detection does not affect to the fatigue detection of the system, in fact a single misclassified frame won't even be notice in the overall performance.

The frame level analysis can give a more realistic vision of the performance of the system, which, as it is shown, works better with indirect IR illumination, but there is not a significant change.

5 CONCLUSIONS

In this paper we have introduced a vision-based Driver Fatigue Warning system (DFW). It has been devised to work seamlessly for the two most used IR schemas: direct and indirect illumination. The system has been implemented and tested showing real-time performance in laboratory and field tests, in a wide range of scenarios, including day and night-time, the presence of glasses, sunglasses, and a variety of drivers. One of the conclusions of our study is that preferably, indirect IR shall be used, for its best performance compared with direct IR. However, ergonomy, size and weight of the sensors might not always allow installing indirect IR equipment in real cars. The next steps of this project include the migration of this algorithm to an embedded system using FPGA capabilities to increase the frame rate at which it can operate (currently near 20 fps in a low-cost ARM-based platform).

ACKNOWLEDGEMENT(s)

The works described in this paper have been partially supported by the program ETORGAI 2013-2015 of the Basque Government under project IAB13. This work has been possible thanks to the cooperation with Datik - Irizar Group for their support in the installation, integration and testing stages of the project.

References

[1] **Mohan Manubhai Trivedi, Tarak Gandhi, & and Joel McCall,** *Looking-in and looking out of a vehicle: Computer-vision-based enhanced vehicle safety.* Intelligent Transportation Systems, IEEE Transactions on, 8(1), 108-120.

[2] **Takei, Y. Furukawa, Y.**, *Estimate of driver's fatigue through steering motion. In: Man and Cybernetics*, 2005 IEEE International Conference. (Volume: 2, On page(s): 1765-1770 Vol.2)

[3] **K. Hong and Chung,** *Electroencephalographic study of drowsiness in simulated driving with deprivation.* International Journal of Industrial Ergonomics. (Volume 35, Issue 4, April 2005, Pages 307-320)

[4] **Marcos Nieto, Oihana Otaegui, Gorka Velez, Juan Diego Ortega and Andoni Cortes,** *On creating vision-based advanced driver assistance systems,* IET Intelligent Transport Systems, 2014, DOI: 10.1049/iet-its.2013.0167

[5] **Flores, M. J., Armingol, J. M., & Escalera, A.**, *Real-time drowsiness detection system for an intelligent vehicle.* In: Intelligent Vehicles Symposium, 2008 IEEE (pp. 637-642). IEEE.(2008, June).

[6] **Viola, P., Jones, M.,** *Rapid object detection using a boosted cascade of simple features.* In: Proc. Computer Society Conf. on Computer Vision and Pattern Recognition, vol. 1, pp.511-518. IEEE (2001)

[7] **Jo, J., Jung, H. G., Park, K. R., Kim, J., & Lee, S. J.** *Vision-based method for detecting driver drowsiness and distraction in driver monitoring system*. Optical Engineering, 50(12) (2011).

[8] **G. Velez, O. Senderos, M. Nieto, O. Otaegui, and G. Van Cutsem**, *Implementation of a computer vision based advanced driver assistance system in Tizen IVI,* in Proc. 21st ITS World Congress, Detroit, U.S., 7-11 September 2014.