Original Research Paper

# Fuzzy Logic Decision Support System for Hypovigilance Detection based on CNN Feature Extractor and WN Classifier

**Ines Teyeb, Ahmed Snoun, Olfa Jemai and Mourad Zaied**

*RTIM: Research Team in Intelligent Machines,*
*University of Gabes, National Engineering School of Gabes (ENIG), Tunisia*

**Abstract:** Fatigue and drowsiness are among the main causes of traffic accidents, just behind excessive speed and alcoholism. This paper deals with the problem of road safety. It attempts to present a driver vigilance monitoring system based on a video approach. This work aims at creating an assistive driving application employing eyes closure duration and head posture estimation as performant signs for alertness control. The proposed system can be summarized in three main steps: Eyes' detection and tracking in a video, eyes' state classification and fusion of both sub-systems based on eyes' blinking and head position. To accomplish the previous tasks, we used the Viola and Jones algorithm for interest area detection thanks to its efficiency in real time applications. For the classification step, we used two novel architectures of transfer learning classifier based on fast wavelet transform and separator wavelet networks, which presents our main contribution of this paper. This novel architecture proves its performance compared to the classic version of the transfer learning based on SVM classifier and to our old classifier based only on fast wavelet networks without a deep learning structure. Different datasets with different classifiers are used to evaluate our new approach. Our second contribution is illustrated by the final system which uses the fuzzy logic and provides five different vigilance levels. Global rates given by experimental results show the effectiveness of our proposed classification system for eyes' state recognition and driver drowsiness detection.

**Keywords:** Transfer Learning, Deep Learning, Wavelet Network, Vigilance, Fuzzy Logic

## Introduction

Driving is a complex activity that involves many tasks: Finding the way, following the road, monitoring the speed, avoiding obstacles, respecting the rules of the road and controlling the vehicle, etc 14. Therefore, It is obvious that this activity requires a very high level of alertness in order to avoid accidents. Unfortunately, accidents related to the hypovigilance do not stop increasing.

Volvo (2013) estimates that the number of deaths on roads of the world is about 1.2 million each year and that 90\% of these accidents are mainly due to drivers' errors. Therefore It is essential to monitoring continuously the driver's vigilance level to ameliorate their ability to maintain safe and efficient driving. This lack of vigilance may take many forms such as, drowsiness, fatigue and

distraction. In this study, we propose a novel approach for driving assistance based on a multimodal system by fusing our cited systems. This new application allows us to detect five levels (alertness, distraction, fatigue, micro-sleep and full sleep) which are different to those cited in the literature.

Our proposed system is composed of these main phases: Location and tracking of face and eyes' regions by the method of Viola and Jones (2001) thanks to its great reliability in object location, eyes' recognition by a novel architecture of transfer learning classifier and fusion task of both control systems to check the driver's alertness level. The focus will be on designing a classification system able to recognize the eyes' state (closed or opened) in real time. Our system uses the deep learning techniques and exactly the transfer learning method. We used a pre-trained Convolutional Neural

Network (CNN) called alexnet to classify the eyes' images. In addition to the alexnet model, we used the Separator Wavelet Network Classifier (SWNC) to minimize the application's computational time. According to the state of the art documentation, the alexnet model is generally used as a feature extractor with SVM or Softmax for classification. Because our system is made under real time constraint, we aim at reducing the processing time. For this reason we propose a novel architecture of the transfer learning. In fact, we keep the alexnet model for feature extraction and we replace the SVM by the Separator Wavelet Network (SWNC) thanks to its reduced time compared to the SVM.

The structure of our paper is as follows: We begin by citing the related works. Then we present the general process of our system. The third part is devoted to explain our first contribution manifested by the novel eyes' classifier based on two novel architectures of transfer learning classifier. The principle of our fusion system allowing us to obtain five vigilance levels that is explained in the fourth part. It is the topic of our second contribution. Results and discussion of our eyes' classification system and hypovigilance detection are presented in the fifth section. Finally, we end up with a conclusion and perspectives.

## Related Works

Various works have been carried out to develop systems for driver condition monitoring in order to emit visual or audible alarms when his behavior is deemed abnormal. They can be divided into three main categories. The first one is based on analysis of physiological signals. It consists on measuring the variation of biomedical signals such as cerebral waves or heart rates using special sensors like the Electroencephalography (EEG) or Electrocardiography (ECG) (Shin *et al.*, 2010). Despite the precision of the cited signals, they do not seem suitable in case of a real driving condition because of the installed materials on the driver's body.

The second category is based on analysis of physical signals. This approach relies mainly on the treatment of the driver's video to measure the level of vigilance reflected by his facial features. It is observed that in the case of hypovigilance, the driver exhibits certain easily observable visual behaviors such as head nodding, prolonged eyes' closure, yawning, fixed gaze, etc (Momin and Abhyankar, 2012). These behavioral signs are analyzed by non-invasive techniques, which use purely visual indicators relating to the vigilance decline. On the practical level, the non-invasive techniques are easier to be exploited under real conditions as they have less constraints than the other techniques during deployment.

The third approach is based on vehicles' behavior control (Klein *et al.*, 1980). Here drowsiness may be detected via different measurements such as pressure on the acceleration pedal, analysis of the movement of the steering wheel and the angle value of the car movement compared to the lane position, etc. When the previously cited signs reach a specific value, it means that there is a low probability which signify that the person is sleepy (Renner and Mehring, 1997). This approach does not seem very efficient because it may depend on the way of driving, the shape and characteristics of the road.

Most of the previous works provided a vigilance classification into two (Liang and Lee, 2015), three (Picot *et al.*, 2012) or four levels (Akrout and Mahdi, 2015). Their systems do not take into account all the hypovigilance levels which are distraction, fatigue and sleeping. We found systems oriented only to fatigue levels' detection (tired, little tired, so tired) (Picot *et al.*, 2012; Akrout and Mahdi, 2015) and others are interested in distraction detection (Céline *et al.*, 2015). These examples of binary or general classification are founded in our previous work (Teyeb *et al.*, 2014a; 2014b; 2015a; 2015b) where two separated vigilance monitoring systems are based successively on eyes' blinking analysis and head pose.

## Overview of Our Hypovigilance Detection System

Through this work, we aim at developing a multi-modal system for vigilance control based on video approach. According to Caschera *et al.* (2007) a multimodal application combines visual information (involving images, text, sketches and so on) with voice, gestures and other modalities to provide flexible and powerful dialogue approaches, enabling users to choose one or more of the multiple interaction modalities. In our system, we combined two visual parameters which are the eyes closure duration and the head movement angle. The general process of our proposed system for vigilance control is illustrated in Fig. 1.

After segmentation of the captured video into frames, the interest areas (head and eyes) are detected and tracked using Viola and Jones algorithm. Our system is based on a multi-variable approach. In fact, it is composed of two sub-systems which are the eyes blinking analysis and the head posture estimation. The first parameter of vigilance monitoring is the head movement angle. If the angle exceeds a specific value, that means that the driver is in an hypovigilant state.

Also the frequency of head movement is an efficient sign for fatigue detection (Teyeb *et al.*, 2014b).

The second parameter is the eyes' closure duration which presents a significant sign for heavy eyelids' detection, where the driver has the desire to close his eyes for a moment because he feels drowsy. If the eyes' closure duration exceeds a predefined time T, that means that the driver is in a drowsiness state (Teyeb *et al.*, 2013; Jemai *et al.*, 2013).
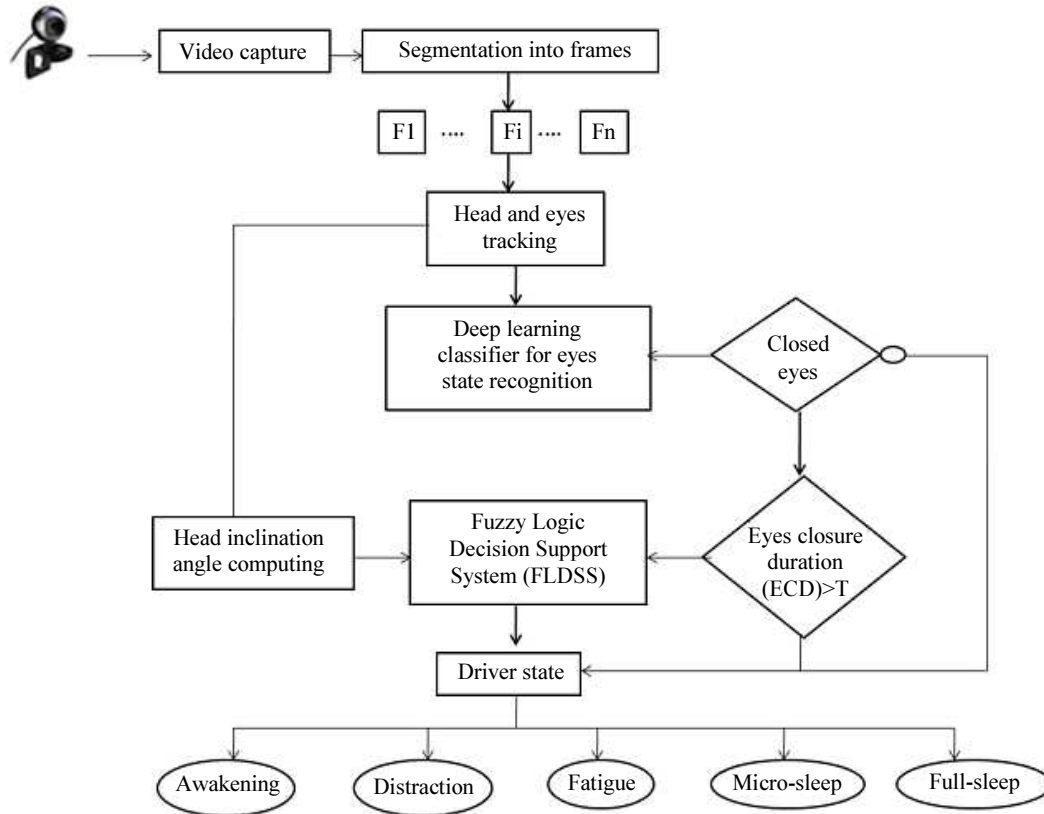
**Fig. 1:** Architecture of our proposed system

Both parameters are merged by a Fuzzy Logic Support Decision System (FLSDS) to recognize the driver state.

## Development of a Multi-Modal Driving Assistance System

In this section, we explain the principle of eyes classification system based on the Convolution Neural Network (CNN) and Wavelet Network Classifier (WNC) used in our system for eyes state recognition

### Eyes Classification System based on CNN Feature Extractor and WN Classifier

### CNN Architecture for Feature Extraction

Deep learning refers to a set of automatic learning methods that are based on the artificial neural network. This new type of learning is used to model the data with high level of abstraction. Indeed, this technique provides a significant and rapid progress in fields of signal analysis, object recognition and computer vision. It is also based on the use of a set of non-linear processing layers for extracting and transforming features. Thus, each layer takes as input the output of the previous one.

Deep learning is characterized by a multi-level learning of details or data representations, called levels of data abstraction. We found several architectures of deep learning. As an example, we cite the Convolution Neural Network (CNN) and the stacked auto-encoder, etc. Due to the difficulty of construction of a big dataset in many systems, researchers has proposed the transfer learning method which can be suitable in case of big or small datasets. In our classification system, we prefer using the transfer learning architecture since the used eyes' dataset is small and it is inefficient to build a convolution neural network from scratch.

As a pre-trained mode, we choose to apply the alexnet CNN for eyes classification (Fig. 2). In the literature there are other models which have a deep architectures like the Visual Geometry Group (Vgg) and the residual network.

But here we choose the alexnet model thanks to its simple architecture which is clearer and simpler than others methods. In fact, the alexnet model is composed of 11 layers (five convolutional layers, three pooling layers and three fully connected layers). However, the Vgg network is composed of 21 layers (thirteen convolutional layers, five pooling layers and three fully connected layers). For the residual network it is characterized by 34 layers (thirty two convolutional layers, one average pooling layer and one fully connected layer).
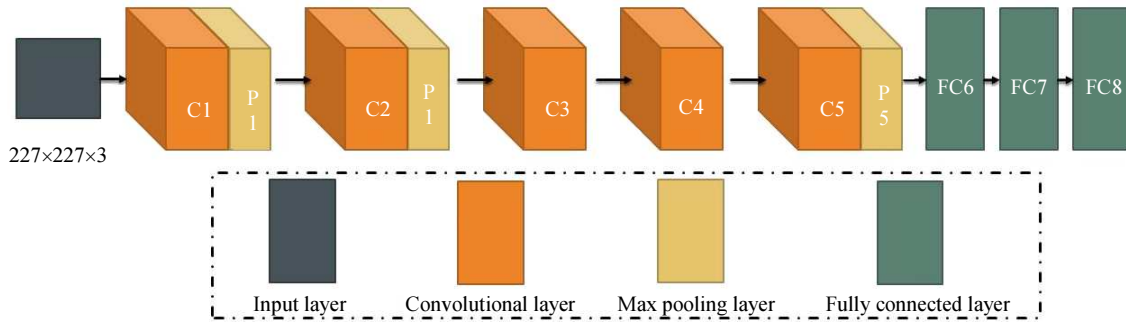
227×227×3

Input layer    Convolutional layer    Max pooling layer    Fully connected layer

**Fig. 2:** Alexnet model architecture

Our objective is to provide an application which takes into account the real time constraint. So, it is the simplest architecture which can provide, the less processing time that system takes. Besides, the alexnet model provides promoting results in terms of eye classification rate and classification of vigilance levels which are tested on many bases. So through this choice we try to satisfy the time/performance constraint. Because we try to minimize the processing time as possible to satisfy the real time criterion.

Like other CNN architectures, alexnet includes a set of independent processing layers which are:

*A Convolution Layer*

It is composed of a set of neurons that are connected to a sub-region of the preceding layer, called receptive field. In fact, the convolution layer $C^i$ (layer $i$ of the network) which is characterized by its number $N$ of convolution maps $M_f^i (j \in 1,…,N)$, also known as feature maps and the size $k$ of the convolution kernels (filters) which are often square.

*A Pooling Layer*

After the convolution layer in a CNN, we usually find a layer of pooling, called also a sub-sampling layer. The image in this layer is split into a series of rectangles of a non-overlapping side pixels. In fact, pooling or sub-sampling is used to reduce the spatial size of an image which makes it possible to reduce the amount of parameters and calculation in the network.

*A Correlation Layer*

The activation layer is located between the convolutional layer and the sub-sampling layer to improve the processing efficiency. This layer applies mathematical functions, called activation functions, to the output signals of the convolution layer

*A Fully Connected Layer*

Connected layer is the last layer of the convolutive neural network. In fact, one or more fully connected layers can be added at the end of the network to ameliorate the classification performance.

This network comprises in total 60 million parameters and approximately 650,000 neurons. The learning of alexnet was done using a part of the ImageNet database (Krizhevsky *et al*., 2012). This part contains about 1.2 million annotated images, composed of 1000 categories.. So, we applied the transfer learning based on the principle of the automatic feature extraction. This technique is based on the exploitation only of the convolutional part of the pre-trained network without using the fully connected layers. Each input image is transformed into a feature vector thanks to the convolutional part of the pre-trained network. These extracted feature vectors will constitute a learning dataset for the classifier. The extracted features are used to train a classifier which will be used to recognize the eyes' states.

The standard architecture of the alexnet mode based on the SVM classifier is cited in Fig. 3. Several types of classifiers can be used to classify the eyes' images using the feature vectors which are generated by the convolutional part of the pre-trained model, such as linear Support Vector Machine (SVM) or softmax.

Indeed, the user has the freedom to replace the SVM classifier. In our system, we have used two classifier based on the Wavelet Network (WN) architecture. The first one is based on the WN classifier learnt by the Fats Wavelet Transform (FWT). The second one use the Separator Wavelet Network Classifier (SWNC).

*Eyes Classification using WN Classifiers*

*Architecture of the Transfer Learning based on FWT Classifier*

The architecture of this system is illustrated by Fig. 4. The combination of wavelet theory and neural networks has led to the development of wavelet networks. Wavelet network presents a version of neural networks which uses wavelets as activation functions.

First of all we must prepare a candidate wavelets and scaling functions library used as activation functions of the wavelet network. The second step consists of computing the coefficients corresponding to the scaling and the wavelets by multiplying each function with its weight.

Finally, the functions of the mentioned library by applying a FWT to the signal f to be learned using a

dual set of scaling and wavelet function filters. Then, we calculated all the possible contributions of activation functions of the library and this network will be built by incrementally adding of a transfer function (wavelet or scaling function) at the hidden layer.

This process will be repeated until reaching a stop criterion. This algorithm is well explained with more details in (Ben Amar and Jemai, 2007; Ejbali *et al.*, 2010; Bouchrika *et al.*, 2012; Said *et al.*, 2009; Zaied *et al.*, 2011; 2008; 2005; Guedri *et al.*, 2011; Ejbeli *et al.*, 2015; Jemai *et al.*, 2011; 2010).

In the test phase, a new coefficients are obtained after projection of each test image on the wavelet networks of the previous stage. Finally, coefficients of both stages are compared by computing Euclidian distances. The learning image having the closest parameters to the test image, gives us to which class the test image belongs.
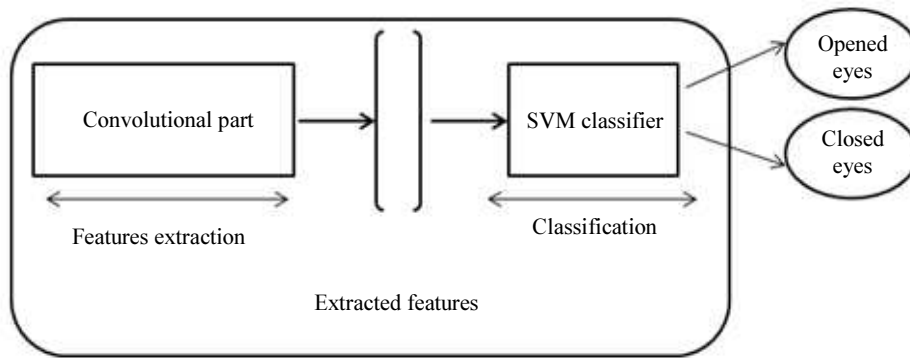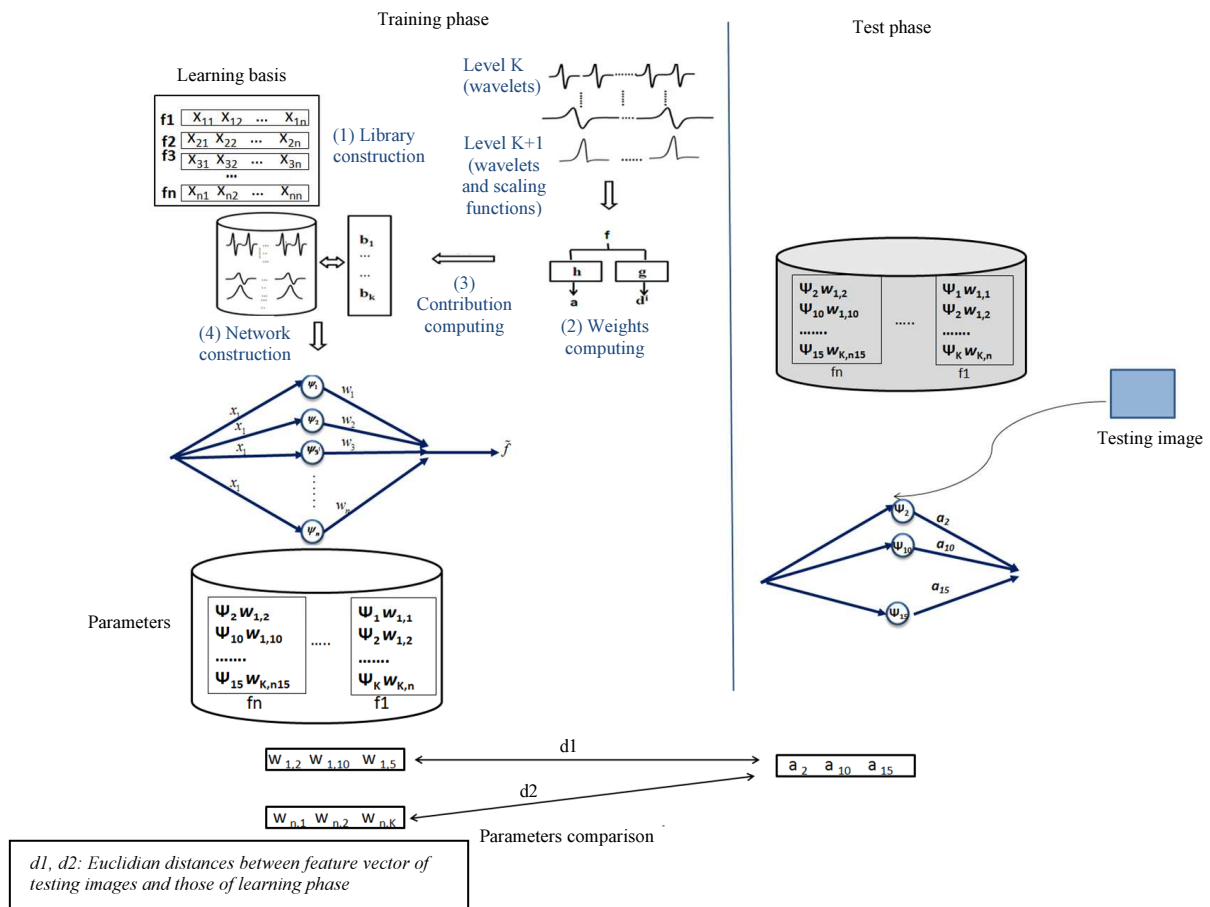


**Fig. 3:** CNN architecture



**Fig. 4:** Principle of wavelet network classifier learnt by fast wavelet transform

*Architecture of the Transfer Learning based on SWN Classifier*

The increase in the training images makes it possible to improve the performance of the wavelet networks based on the fast wavelet transform, but the speed of these networks is reduced, since each learning sample is represented by its own network of wavelets. To solve this problem, we used a classifier based on the advantages of the fast wavelet transform and the adaboost algorithm. It is the separator wavelet network.

The use of the Separator Wavelet Networks Classifier (SWNC) with the alexnet model represents our second contribution. It is a different classification method compared to those of the literature. Indeed, this classifier (Bouchrika *et al.*, 2014) combines the wavelet network learnt by Fast Wavelet Transform (FWT) and the adaboost algorithm advantages (Zhou *et al.*, 2006).

Our proposed classifier uses *n*-1 wavelet networks to represent *n* classes instead of using a wavelet network for each sample in the learning base and since we have two classes (opened eyes and closed eyes) then we will use a single wavelet network which is considered as a separator between both classes.

The SWNC is also based on the adaboost algorithm which allows the selection of the best characteristics used to improve the performance of the classifier.

The training of this classifier is similar in its beginning to the training of the classifier based on the wavelet networks, used in our previous section. In fact, after the extraction of the characteristic vector of each image in the learning base using the last convolutional layer of the Alexnet model, a decomposition of this vector is performed using the Fast Wavelet Transform (FWT). Then, we move to select the best characteristics using the adaboost algorithm which is based on a combination of a set of weak classifiers to construct a single strong classifier to obtain 1 SWNC.

Indeed, to classify an image, we project it on only this SWNC to find new weights' vector. In fact, the formula below is used to predict the class of the test example:

$$D(x) = sigm\left(\sum_{k=1}^{K} \delta_k h_k(x)\right) \qquad (1)$$

With: $D(x)$ is the predicted decision about the test example $x$ which represents the feature vector of the query image generated by the alexnet model, $K$ represents the number of the weights, $h_k$ is the classifier which was trained for a $k$th weight and $\delta_k$ is a threshold calculated, for the $k$th kernel, in the training phase.

If $D(x) > 0$, then the test sample is correctly classified.

Else if $D(x) < 0$, then the test sample is negatively classified.
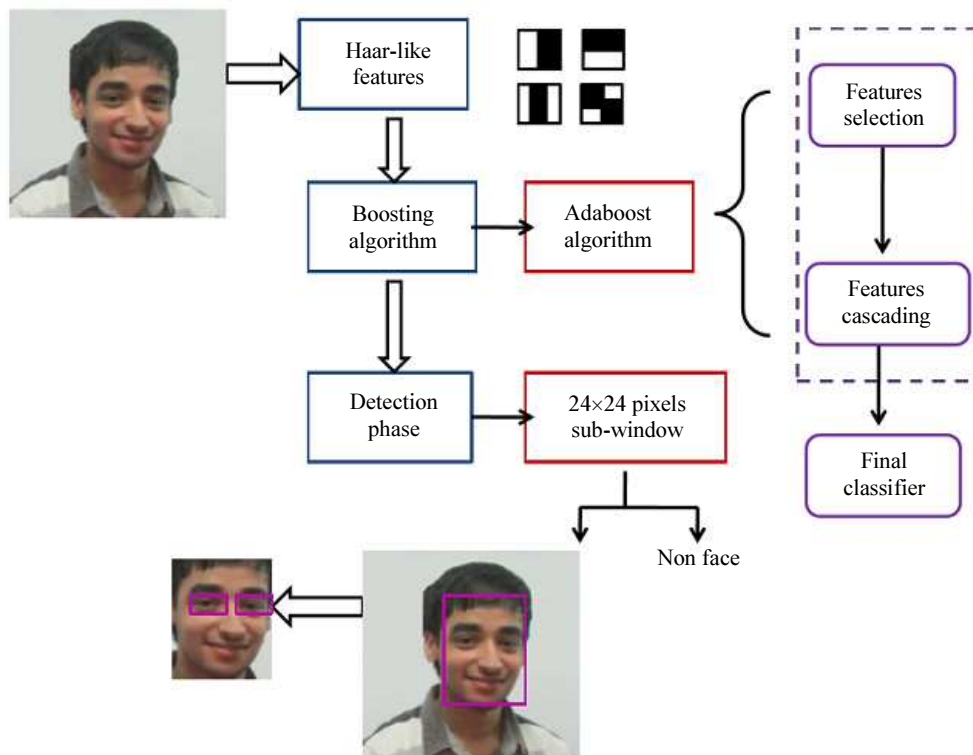


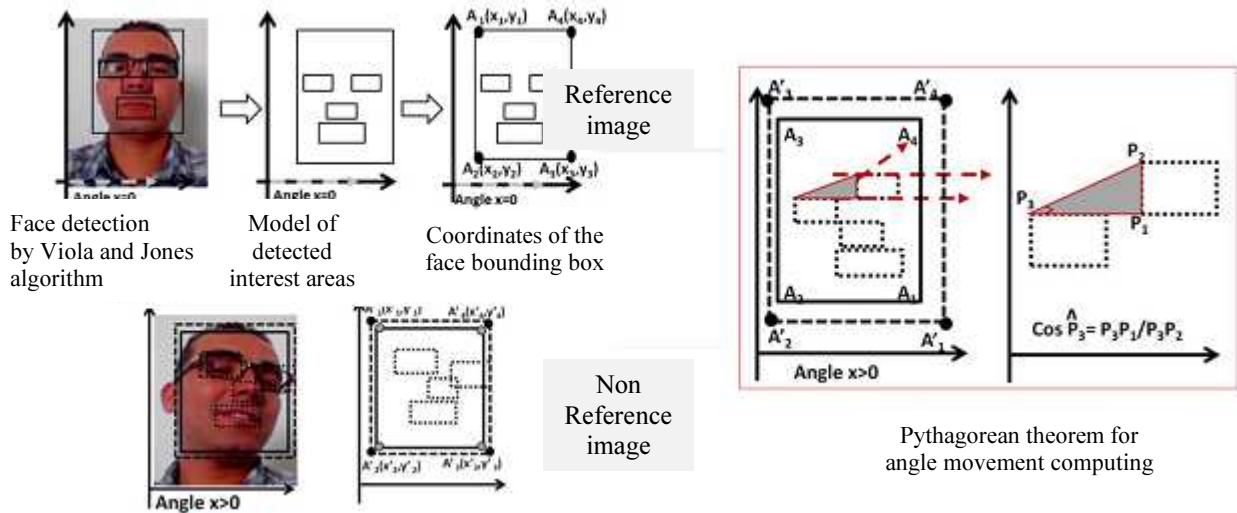**Fig. 5:** Principle of viola and Jones algorithm

**Fig. 6:** Principle of head movement angle computing

## Head Movement Angle Detection System

The head movement angle is measured by applying the Pythagorean theorem in the triangle which is formed by both eyes bounding rectangle after detecting them using the Viola and Jones algorithm (Fig. 5).

After face and eyes detection using Viola and Jones algorithm, we check if the head is in a movement state or not by comparing the extracted coordinates of the corners of bounding rectangle face (Fig 6). The first frame of the captured video is considered a reference image, the following frame is considered a non reference image. If the head is in movement state, we calculate its angle value using the Pythagorean theorem as it was mentioned in Fig. 2. More details can be found in (Teyeb *et al*., 2014b).

Figure 5 illustrates the principle of Viola and Jones algorithm which is used for face and eyes detection.

It is characterized by three steps:

- Haar-like feature: Theses features are computed by the integral image, an image structure which accelerates the computing process
- Boosting algorithm: It is the algorithm of training used by Viola and Jones detector. It is based on combining weak classifiers to obtain a final strong one called the boosted classifier

## Fuzzy Logic Decision Support System for Hypovigilance Detection

As we have already said in the introduction section, our objective is the conception of a multi-variable system for alertness level control through eyes blinking analysis and head position estimation. To achieve this task, we used the fuzzy logic. Its principle is mentioned in Fig. 7.

At this level, our technique can check the position of the head (normal or inclined). But to judge the driver's alertness level, we need other parameters like the inclination degree. If the value of head inclination angle exceeds a predefined value X, we conclude that the driver' s vigilance level is down.

We are now developing another system by adding a third parameter which is the head movement duration to made a more precise constraint for the head movement angle.

In the fuzzification step, we have used the trapezoidal shape in this study.

For the inference task, we will merge both input variables, the eyes' closure duration and the head movement angle to find the alertness level of the output variable. Here we cite the fuzzy rules used for decision making:

- **If** ECD = low **And** A = small **Then** vigilance level = Alertness
- **If** ECD = low **And** A = big **Then** vigilance level = Distraction
- **If** ECD = middle **And** A = small **Then** vigilance level = Fatigue
- **If** ECD = middle **And** A = big **Then** vigilance level = Micro-sleep
- **If** ECD = large **And** A = small **Then** vigilance level = Full-sleep
- **If** ECD = large **And** A = big **Then** vigilance level = Full-sleep

The most widely used method in the literature in the inference task is the max-min method, also known as the "Mamdani" method (Mamdani and Assilian, 1975). This method, like the other methods, goes through three stages, which are:

- Step 1: Activation of the rules
- Step 2: Calculation of the implication

1552

- Step 3: Aggregation

   An example of this method is cited in Fig. 9. The last step is defuzzification task, which converts a fuzzy value obtained from the inference step to a real value. to perform defuzzication, several methods can be used:

- First maximum
- Center of gravity
- Last minimum
- Center maximum
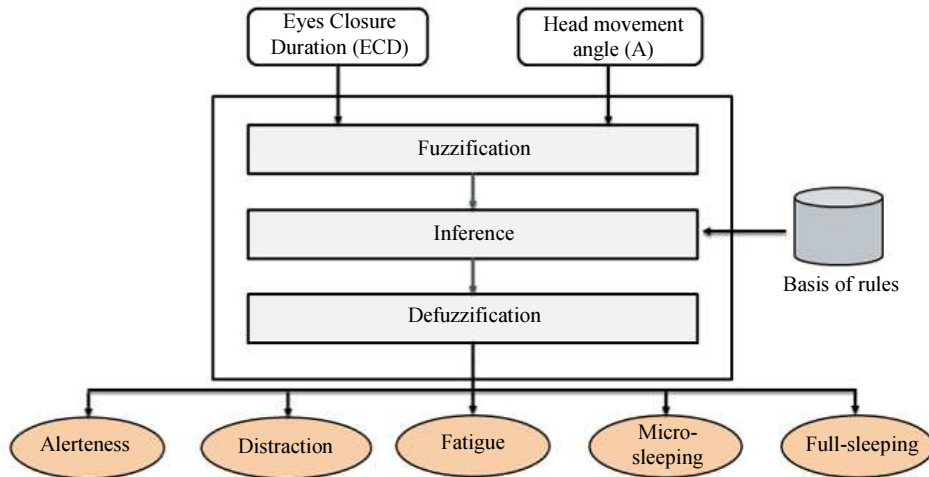- Weighted averages an example is mentioned in Fig. 10



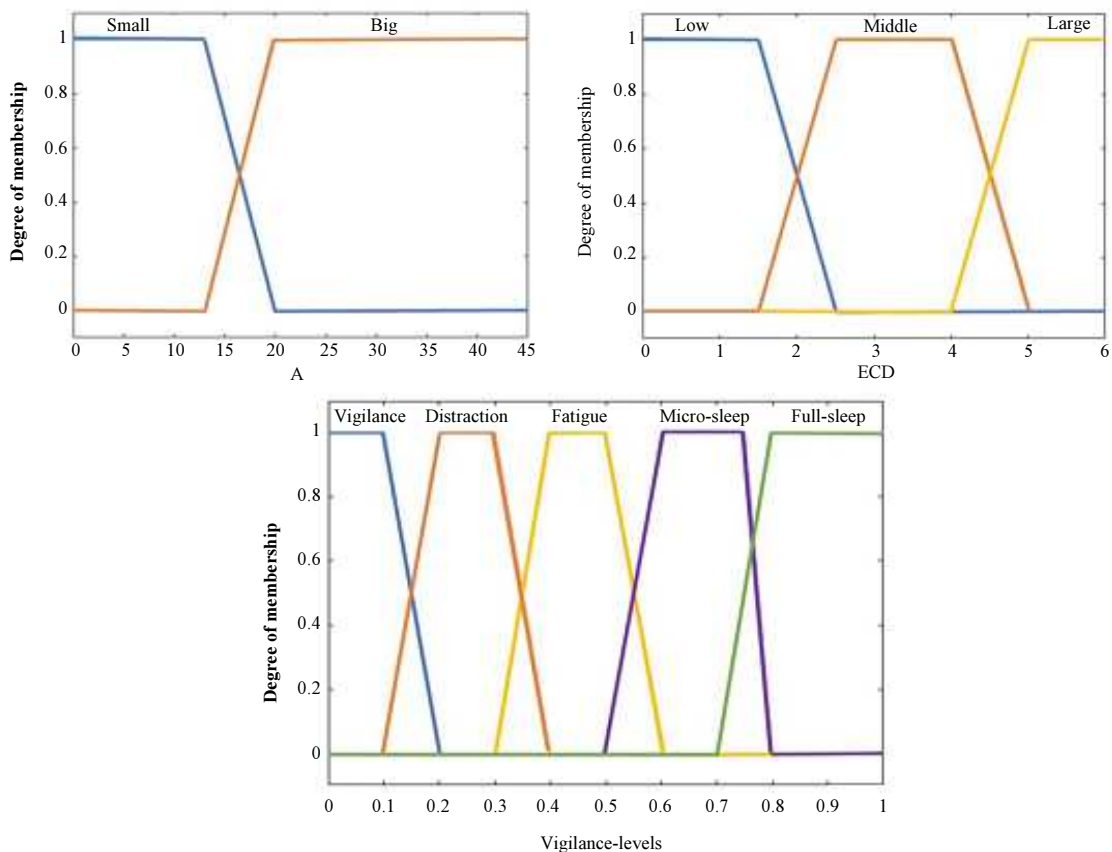**Fig. 7:** Fusion Task principle



**Fig. 8:** Membership function

Figure 8 explain the fuzzification task which is the first step in the fuzzy inference process. This involves a domain transformation where crisp inputs are transformed into fuzzy inputs. In other words, the fuzzification is the process of transforming a real scalar value into a fuzzy value, called also linguistic value. So, we have transformed the real values of the head movement angle into linguistic values which are Small and Big, so for example when the value of the angle is 10°, it means that it is a small angle. And we have done the same operation to our second input which is the eyes closure duration and our output which is the vigilance levels.

The Inference is the process by which fuzzy actions or operations are applied to input variables according to the rules defining the system. In our case, the inference will merge the two input variables, namely, the eyes closure duration and the head movement angle, to find in the output the driver level of vigilance.

For example, as cited in figure 9, if we have as inputs, ECD = 2.5s and the Angle = 20°, so in the first rule where the ECD is low, the 2.5s of the ECD will give us 0.5 as fuzzy value and the 20° will give us 0.2 when the Angle is small. So according to the Mamdani principle, we will take the min of 0.5 and 0.2 since we have the operator "AND" between the two conditions. And, then we do the same operation for the second rule

which will give us 0.5 for the ECD and 1 for the Angle and since we have "AND" between the two conditions, we will take the 0.5. And the last step is the aggregation between the two rules which will be performed using the operator "max".

We used the method based on the center of gravity which is the most used method. For a resulting membership function μR (y), the center of gravity (y *) can be calculated by the following equation:

$$y^* = \frac{\int_0^1 y \mu R(y)\, dy}{\int_0^1 \mu R(y)\, dy} \qquad (2)$$

Its principle is mentioned in Fig. 10 by making the union of fuzzy output subsets and calculating a global center of gravity. The threshold value of angle (A) is fixed experimentally to 16°. We propose to check an approximate value of this parameter. We found 16 as the minimum value in which the driver may be still vigilant and his gaze direction is still well fixed in the road. Because each head movement will be accompanied by a change in the gaze direction of the driver.

Figure 11 mentions examples of our experimental task to fix an approximate value of the head movement angle.
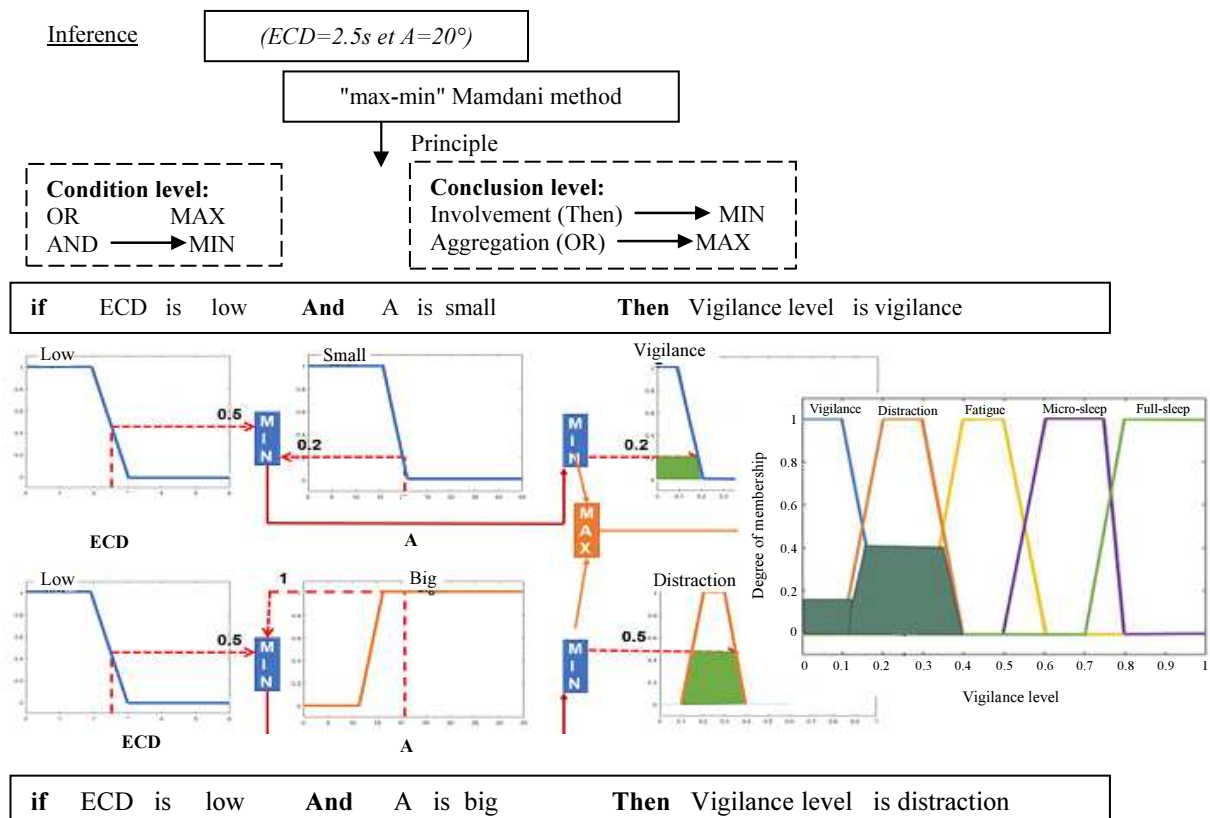


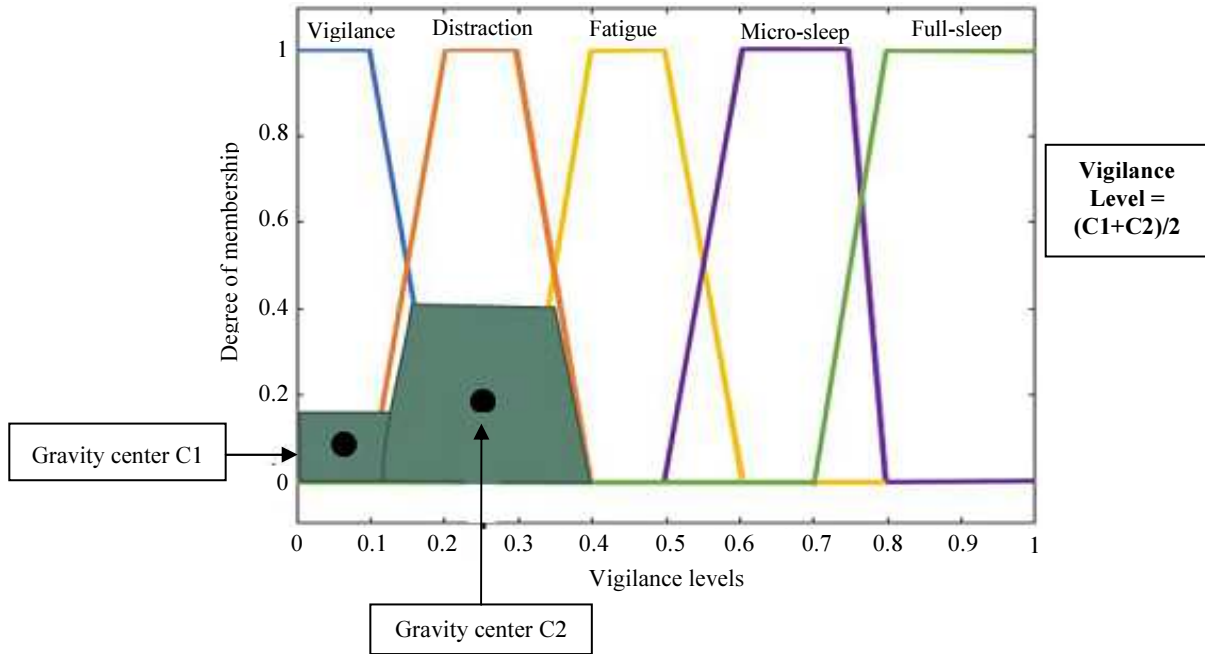**Fig. 9:** Inference example by the Mamdani method

**Fig. 10:** Example of the gravity center method for the deffuzification step



*Normal position*                *Inclined position*

**Fig. 11:** Experimental task to fix the minimum value of head movement angle

Concerning the duration of eye closure, Sarbjit and Nikolaos (1999) considered that the person is in a state of full sleep if he keeps his eyes closed for 5 to 6 sec. But, if this duration is between 2 and 3 sec, the person is in a micro-sleep state. However, Horng *et al.* (2004) affirmed that the driver is drowsy if he closes his eyes for 5 successive frames. According to the research activities of Sharabaty *et al.* (2008), the maximum period of eyes' normal blinking is equal to 0.5 sec. If the closing time exceeds this value, then we are talking about a state of prolonged closure.

At the output of the fuzzy inference, the result is always a fuzzy set. In order to be used in the real world, the fuzzy output needs to be transformed to the crisp

domain by the defuzzifier. This needs the use of a suitable membership functions. In this study, we have used the center of gravity method.

In the inference phase, we used the max-min method, which was called also the Mamdani method. It is well used in the literature thanks to its simplicity and efficiency in the fusion task.

According to these studies, we have categorized the duration of eye closure into three sub-intervals:

- low = [0, 2s]; middle = [2s, 4s]; large = [5s, 6s]
- small (A<= 16°) ; big (A>16°

# Results and Discussion

## Datasets Presentation

Figure 12 summarizes the different bases used in the experimental task to evaluate both systems of eyes state classification and hypovigilance detection.

To test the performance of Viola and Jones algorithm used for face and eyes tracking, we used the Yaw DD base. Once eyes are detected, we pass to the second step which is the eyes state recognition. This task is evaluated in four different basis (BioID, CEW, ZJU and our private one). The last task is the vigilance level classification. Our system is tested on our appropriate video base.

## Experimental Results

## Eyes Tracking by Viola and Jones Algorithm

The position of the camera may influence the quality of eyes' detection and tracking. We check the good eyes' tracking rate using the Viola and Jones algorithm in our appropriate basis and Yaw DD (Abtahi *et al.*, 2014) dataset which is extracted from the labeled faces in the wild (LFW) dataset (https://www.bioid.com/About/BioID-Face-Database; http://parnec.nuaa.edu.cn/xtan/data/ClosedEyeDatabases.html). Composed of 322 videos captured with different camera positions (Dash and mirror position).

When the camera is placed in parallel position to the eyes' axis of the driver, the correct tracking rate is more efficient than the case of dash or mirror position (Fig. 14).
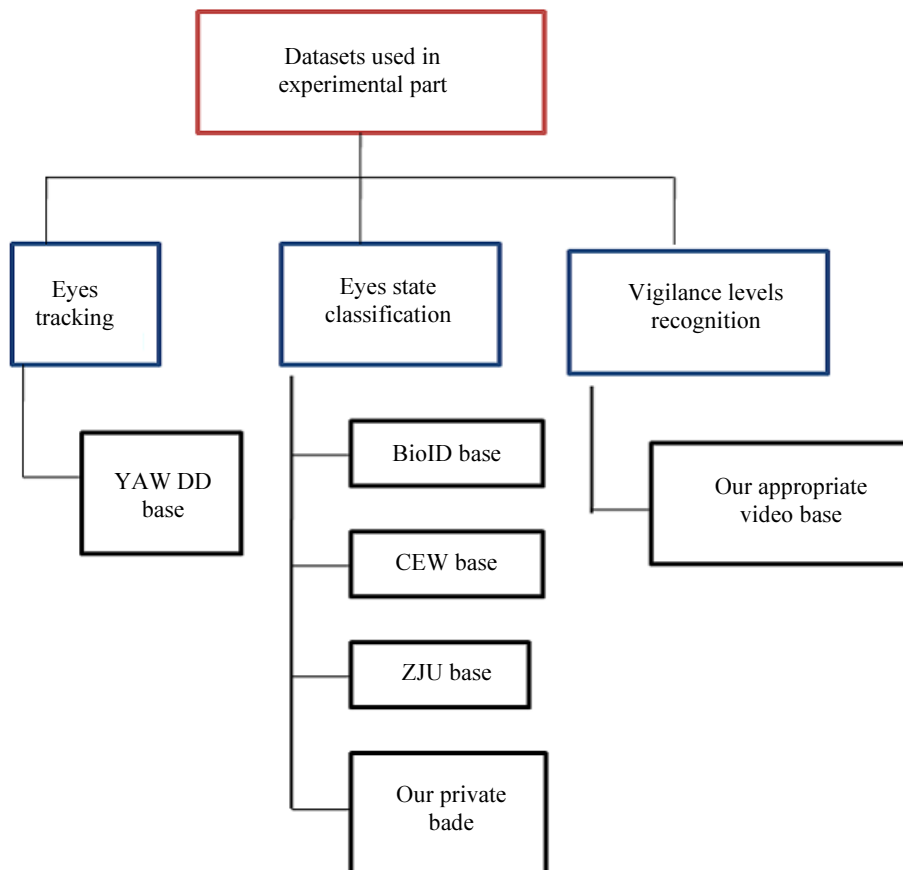


**Fig. 12:** Basis of the experimental part

For performance evaluation of our proposed approach, we used the Correct Detection Rate (CDR) to check the reliability of Viola and Jones technique for eyes detection. The cited metric is calculated as follows:

$$CDR = \frac{Number\ of\ images\ with\ correct\ detection}{size\ of\ the\ dataset} \qquad (3)$$

Table 1 summarizes the results of correct detection rates for each dataset.

In Fig. 13, we mention examples of some typical success and failure cases of Viola and Jones algorithm for eyes detection in both bases CEW ((a) and (b)) and BioID dataset ((c) and (d)).
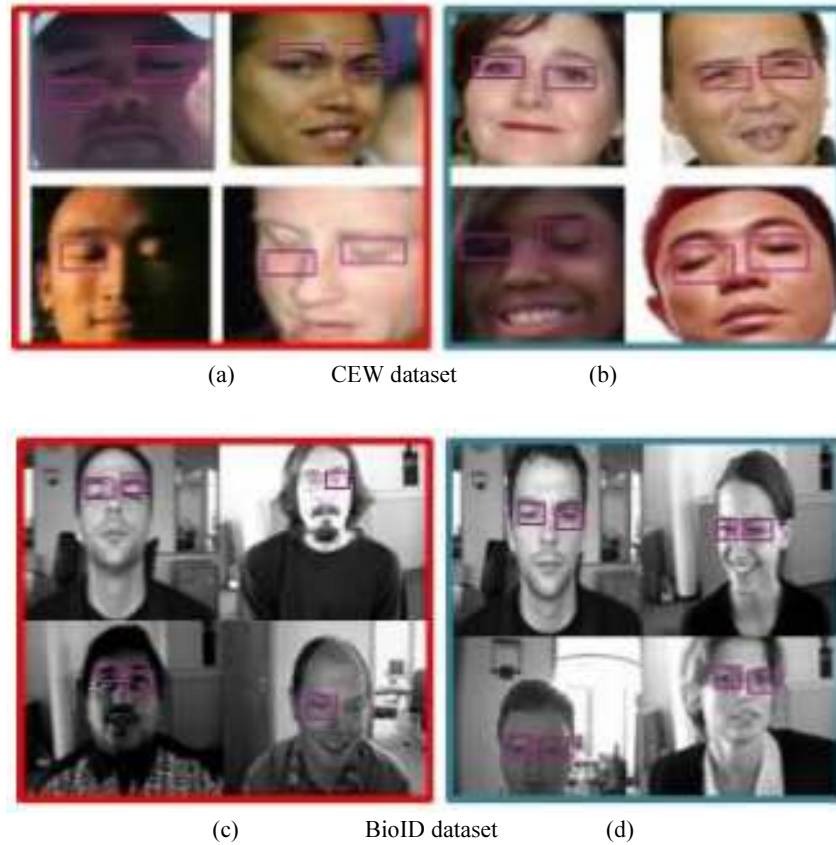


(a)  CEW dataset  (b)

(c)  BioID dataset  (d)

**Fig. 13:** Examples of failure and success cases of eyes location in BioId and CEW dataset



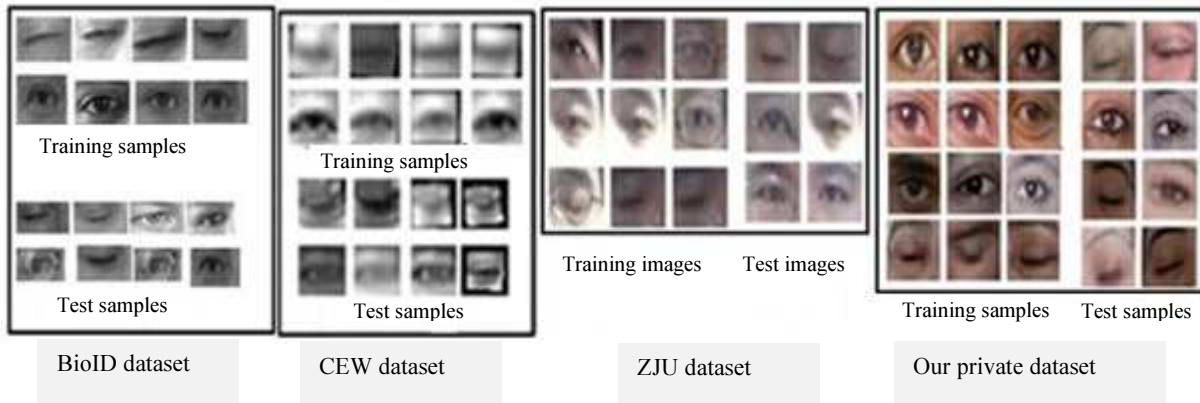**Fig. 14:** Different camera positions in Yaw DD datasets and our appropriate basis

1557

**Fig. 15:** Samples of different basis

**Table 1:** Correct eyes' detection rates on different databases using Viola and Jones algorithm

| Dataset | CDR |
|---|---|
| CEW (opened eyes) | 96% |
| CEW (closed eyes) | 83% |
| BioID | 97.5% |

## Eyes State Classification

Figure 15 mentions samples of different datasets used for performance evaluation of our different eyes classifiers.

### BioId Datasets

This image basis is made up of 1521 gray scale images, characterized by a resolution of 384×286 (Bhler, 1993). It contains a variety of capture conditions. It keeps persons with and without glasses and with different head poses and eyes' states (opened and closed eyes).

### ZJU Datasets

It is collected from the ZJU Eye blink Database (Song *et al.*, 2014). It is composed of 80 video clips produced by 20 participants and every person recorded four clips.

### CEW Basis

The Closed Eyes in the Wild basis is composed of 2243 subjects, where 1192 faces with closed eyes are collected directly from the web and 1231 faces are tried to find examples of eye images of some persons put in the same conditions of driver

### Our Appropriate Datasets

We built our appropriate dataset composed of 90 images with uniform size, among which 2/3 are used for the learning stage and the 1/3 is kept for the test one. Tried to find examples of eye images of some persons put in the same conditions of a driver.

## Results of Classification Rate

We have used the Global Classification Rate (GCR) calculated as follows to measure our classification system performances:

$$GCR = \frac{Good\,classification\,images\,number}{Total\,images\,number} \qquad (4)$$

Here, we compared three different eyes classifiers based on Wavelet Network (WNC) learnt by the Fast Wavelet Network (FWT), the classic architecture of transfer learning based on Alexnet model and SVM (ASVM) and our first proposed classifier based on alexnet Architecture and the Fast Wavelet Network (AFWT) and our second proposed classifier based on Alexnet and the Separator Wavelet Network (ASWNC). This evaluation is released from the previous datasets already cited in the previous parts (BioID, CEW, ZJU and our private basis). Table 2 shows the classification results given by the three systems.

We always notice that the classification by a convolutional neural network pre-trained by the alexnet model performs better than that by the Fast Wavelet Network (FWT). Indeed, this is related to the deep architecture of alexnet which allows it to take into consideration the smallest details of the image to generate a characteristic vector that precisely describes the image.

The second step in the evaluation process is to compare our proposed classification system to other popular classifiers studied in the literature. Diverse classifiers and features such as the nearest neighbor (NN), the support vector machine (SVM) and the adaboost classifiers were used. More details about these approaches were cited in (Song *et al.*, 2014). Results are cited in Table 3.

### Computational Time

We are now interested in comparing their computational time, which is a primordial criterion since

our application is in real-time. We measured the computational time of testing an image from the video sequence with the different methods, adopting the same hardware configuration. Table 4 represents the computational time for each method.

We notice, from these results, that the use of a classifier based on a separator wavelet network with the alexnet model for the eyes classification serves to reduce the application's computational time and that is related to the fact that this method is inspired by the method of boosting to reduce considerably the number of operations used in the test phase. In other words, instead of calculating a distance

between the wavelet network of the query image and that associated with each examples of the base, a separator wavelet network is used between two classes in order to perform only *N*-1 comparisons if *N* is the number of classes, which allows a decrease in the computational time of the algorithms, at the cost, however, of a degradation of the performance of the algorithm. The processing time of the entire system is cited in Table 5.

The global processing time of our system is equal to 0.30 sec when using the alexnet combined with the separator wavelet network and it is equal to 0.42 sec with the alexnet combined with the Fast wavelet transform.

**Table 2:** Results of classification rate using different classifiers on various datasets

| Dataset classifier | BioID | CEW | ZJU | Our basis |
|---|---|---|---|---|
| FWT | 92.27% | 82.88 | 61.75% | 90.91% |
| ASVM | 90.32% | 86.88% | 82.93% | 100% |
| AFWT (our first approach) | 95.09% | 86.88% | 65.67% | 100\% |
| ASWNC (our second approach) | 97.99% | 91.03% | 89.21% | 100% |

**Table 3:** Results of classification rates on CEW and ZJU datasets

| Dataset | CEW | ZJU |
|---|---|---|
| NN | 74.31% | 84.74% |
| SVM | 82.85% | 89.62% |
| Adaboost | 87.09% | 92.06% |
| FWN | 82.88% | 61.75% |
| ASWNC | 91.03% | 89.21% |

**Table 4:** Comparison of the computational time of each classification method

| Method | Computational time |
|---|---|
| FWT | 0.007 s |
| ASVM | 0.045 s |
| AFWT | 0.12 s |
| ASWNC | 0.0015 s |

**Table 5:** Processing time of our driving monitoring system

| Process | Eyes and face detection | Classification | | Fusion task | Global time | |
|---|---|---|---|---|---|---|
| | | ASWNC | AFWT | | with ASWNC | with AFWT |
| Processing time (s/frame) | 0.16 | 0.0015 | 0.12 | 0.14 | 0.30 | 0.42 |



**Fig. 16:** Examples of our private video base

*Hypovigilance Detection System*

To test the efficiency of our hypovigilance detection system, we built our appropriate dataset composed of 45 videos (9 recorded videos for each vigilance level), under different light conditions. The recorded participants are students, workers and volunteers. They are of different sex and aged between 20 to 60 years old.

Our videos are recorded in MP4 format with a resolution of 640×480 and frequency of 30 frames/sec.

Examples are cites in figure 16.

Our built dataset contains five vigilance levels which are alertness state, distraction, fatigue, micro-sleeping and full sleeping. Results of correct hypovigilance detection rates on our dataset using different architectures of transfer learning classifiers are summarized in Fig. 17.

We aim to show the efficiency of our system based on the two proposed techniques which are (Alexnet + WNC and alexent + SWNC) compared to the standard architecture of alexnet based on the SVM classifier.

Results prove that using the Wavelet Network Classifier (WNC) learnt by FWT and the Separator

Wavelet Network Classifier (SWNC) with the alexnet model gives us more precise results. The different classification methods are all based on the alexnet model. The results are generally promising. But, we notice a small degradation in the results given by the system based on the wavelet separator network compared to the alexnet with the SVM classifier and the WN classifier especially in the two important cases; micro-sleep and full-sleep. Here the advantage of the SWNC resides on the minimization of the time response compared to the two other classifiers as shown in Table 5. The time response is a primordial criterion for our system to satisfy the real time constraint. Also, the alexnet model with the WN classifier (FWT) performs better than the standard architecture of the alexnet (SVM) in case of distraction. Generally these both classifiers have a closest performance (Snoun *et al.*, 2017).

Table 6 shows a comparison between our hypovigilance detection system and four other approaches. This comparison argues in favor of our system in term of number of detected vigilance levels.

**Table 6:** Comparison of our system with those of the literature

| Authors | Results | Data | Method | Level |
|---|---|---|---|---|
| Our approach Alexnet + SWNC (2017) | 86% | Video | Fuzzy logic | 5 |
| Our approach Alexnet + FWT (2017) | 91% | Video | Fuzzy logic | 5 |
| Akrout and Mahdi (2015) | 87.75% | Video | Expert system | 4 |
| Céline *et al.* (2015) | 98.4% | Video, audio and mechanic | Bayesian network | 2 |
| Liang and Lee (2015) | 90% | Video and driving performance data | hybrid Bayesian Network | 2 |
| Picot *et al.* (2012) | 81.7% | Video and ECG | Fuzzy logic | 3 |
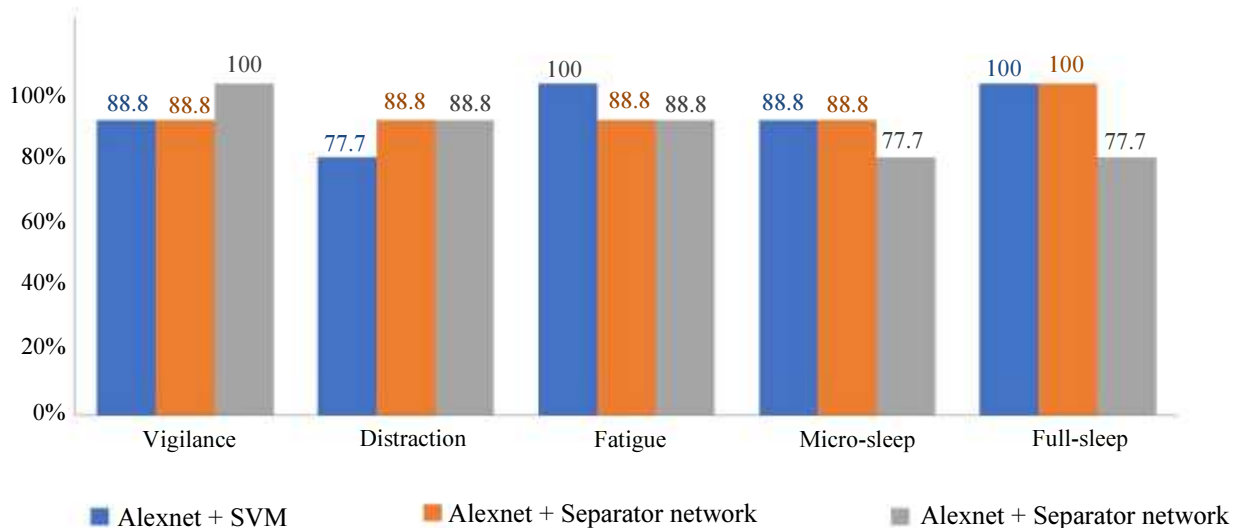


**Fig. 17:** Results of vigilance levels classification rate on our basis using different eyes classifiers

The cited authors in Table 6 do not have the same test bases since each one has specific inputs in his system. Even those who have worked on the same type of data, they do not have the same parameters of analysis. For this reason, they build their own base. In other words, practically there is not a standard basis for testing the hypovigilance approach. Basis of test depends on the parameters and characteristics of the approach developed.
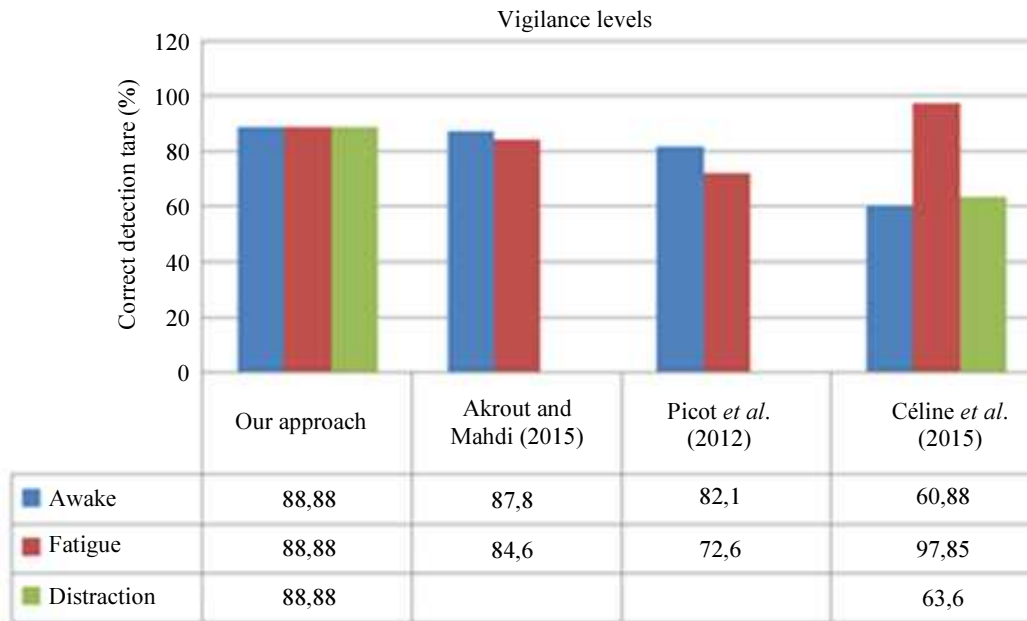
**Vigilance levels**

| | Our approach | Akrout and Mahdi (2015) | Picot *et al.* (2012) | Céline *et al.* (2015) |
|---|---|---|---|---|
| ■ Awake | 88,88 | 87,8 | 82,1 | 60,88 |
| ■ Fatigue | 88,88 | 84,6 | 72,6 | 97,85 |
| ■ Distraction | 88,88 | | | 63,6 |

**Fig. 18:** Comparison of our system with those of the literature per level

Table 6 mentions a comparison in term of global classification rate of vigilance levels. To obtain a more precise comparison, we add Fig. 18 to illustrate a more detailed comparison of vigilance control per level.

According to Fig. 18, our system seems well performant compared to those of other researchers. In fact we have a good detection rate equal to 88,88% for awakening, fatigue and distraction state. These rates are more accurate than the results provided by Picot *et al.* (2012) (82,1% for awakening state and 72,6% for fatigue level).

We have approximately the same rate as Akrout and Mahdi (2015) for awakening state (87,1%) and more precise result in the fatigue level (84,6%). Concerning the system of Céline *et al.* (2015) it is more performant than our system in the detection of fatigue state (97,85%) and less efficient in the other states (awakening 60,88% and Distraction 63,6%).

Our driving assistance system is simulated by using a webcam. The distance between the driver's face and the camera is set by measuring the real distance inside the car between the driver and the dash mirror to be in the real driving conditions. Also, we are now interesting into the phase of the system implementation as an embedded system by using a specific electronic cards that can be useful for different means of transport (car, plane or other). Finaly we aim to extend our driving assistance system by combining it with our smart seat for biomechanical distraction state (Teyeb *et al.*, 2016).

## Conclusion

In this study, we have proposed a multi-modal vigilance monitoring system based on eyes blinking analysis and head posture estimation.

Our first contribution is manifested by both new versions of the eyes state classifiers based on a novel architecture of transfer learning by integrating the fast wavelet transform and the separator wavelet network instead of the SVM classifier with the alexnet model. Different proposed classifiers are evaluated in terms of classification rate and computational time. Results are in favor of the transfer learning classifier based on separator wavelet network.

The second contribution is the conception of the fuzzy decision support system based on fuzzy logic which provide the detection of five vigilance. This classification is different from the work carried out in the literature allowing the detection of a less number of levels.

Our system was tested on various datasets and provides good results in terms of eyes classification rates and vigilance detection levels.

In our future works, we aim at extend our driving control system by adding other inputs like yawning analysis and frequency of head movement.

## Acknowledgement

## Author's Contributions

**Ines Teyeb:** Designed the research plan, organized the study, participated in all experiments, coordinated the data-analysis and participated in the manuscript writing.

**Ahmed Snoun:** Participated in all experiments such as data preparation, selection and testing, coordinated the data-analysis and contributed to the writing of the manuscript.

**Olfa Jemai:** Advise research project, proof reading of the paper and contributed to the writing of the manuscript.

**Mourad Zaied:** Advise research project and designed the research plan and contributed to the paper writing.

## Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and there are no ethical issues involved.

## References

Abtahi, S., M. Omidyeganeh, S. Shirmohammadi and B. Hariri, 2014. A yawning deteCtion dataset. Proceedings of the 5th ACM Multimedia Systems Conference, Mar. 19-19, ACM, Singapore, pp: 24-28. DOI: 10.1145/2557642.2563678

Akrout, B. and W. Mahdi, 2015. Spatio-temporal features for the automatic control of driver drowsiness state and lack of concentration. Mach. Vis. Applic., 26: 1-13.
DOI: 10.1007/s00138-014-0644-z

Ben Amar, C. and O. Jemai, 2007. Wavelet networks approach for image compression. GVIP Special Issue Image Compress.

Bhler, H., 1993. Reglage par logique floue", Presses polytechniques et universitaires romandes. Lausanne.

Bouchrika, T., M. Zaied, O. Jemai and C. Ben Amar, 2012. Ordering computers by hand gestures recognition based on wavelet networks. Proceedings of the 2nd International Conference on Communications, Computing and Control Applications proceedings, Dec. 06-08, Marseilles, France, pp: 1-6. DOI: 10.1109/CCCA.2012.6417911

Bouchrika, T., O. Jemai, M. Zaied and C. Ben Amar, 2014. Cascaded hybrid wavelet network for hand gestures recognition. Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Oct. 05-08, IEEE Xplore Press, San Diego, CA, USA, pp: 1379-1384.
DOI: 10.1109/SMC.2014.6974104

Caschera, M.C., F. Ferri and P. Grifoni, 2007. Multimodal interaction systems: Information and time features. Int. J. Web Grid Serv., 3: 82-99.
DOI: 10.1504/IJWGS.2007.012638

Céline, C., R. Abdullah, K. Mohamed and K. Fakhri, 2015. A multi-modal driver fatigue and distration assessment system. Int. J. Intell. Transport. Syst. Res.

Ejbali, R., M. Zaied and C. Ben Amar, 2010. Wavelet network for recognition system of Arabic word. Int. J. Speech Technol., 13: 163-174.
DOI: 10.1007/s10772-010-9076-y

Ejbeli, R., O. Jemai, M. Zaied and C. Ben Amar, 2015. A speech recognition system using fast learning algorithm and beta wavelet network. Proceedings of the 15th International Conference on Intelligent Systems Design and Applications, Dec. 14-16, IEEE Xplore Press, Marrakech, Maroc, pp: 14-18.
DOI: 10.1109/ISDA.2015.7489241

Guedri, B., M. Zaied and C. Ben Amar, 2011. Indexing and images retrieval by content. Proceedings of the IEEE International Conference on High Performance Computing Simulation, Jul. 4-8, IEEE Xplore Press, Istanbul Turkey, pp: 369-375.
DOI: 10.1109/HPCSim.2011.5999848

Horng, W., C. Chen and Y. Chang, 2004. Driver fatigue detection based on eye tracking and dynamic template matching. Proceeding of the IEEE International Conference on Networking, Sensing and Control, Mar. 21-23, IEEE Xplore Press, Taipei, Taiwan, pp: 7-12.
DOI: 10.1109/ICNSC.2004.1297400

http://parnec.nuaa.edu.cn/xtan/data/ClosedEyeDatabases.html

https://www.bioid.com/About/BioID-Face-Database

Jemai, O., I. Teyeb, T. Bouchrika and C. Ben Amar, 2013. A novel approach for drowsy driver detection using eyes recognition system based on wavelet network. Int. J. Recent Contribut. Eng., 1: 46-52.
DOI: 10.3991/ijes.v1i1.2929

Jemai, O., M. Zaied, C. Ben Amar and A.M. Alimi, 2010. FBWN: An architecture of fast beta wavelet networks for image classification. Proceedings of the IEEE World Congress on Computational Intelligence, International Joint Conference on Neural Networks, Jul. 18-23, IEEE Xplore Press, Barcelona, Espagne, pp: 1953-1960. DOI: 10.1109/IJCNN.2010.5596876

Jemai, O., M. Zaied, C. Ben Amar and A.M. Alimi, 2011. Pyramidal hybrid approach: Wavelet network with OLS algorithm based-image classification. Int. J. Wavelets Multiresolut. Inform. Process., 9: 111-130.
DOI: 10.1142/S0219691311003967

Klein, R., R. Allen and J. Miller, 1980. Relationship between truck ride quality and safety of operations: Methodology development. Technical Report, System Technology.

Krizhevsky, A., I. Sutskever and G.E. Hinton, 2012. Image net classification with deep convolutional neural networks. Adv. Neural Process. Syst. Treatment, 1: 1097-1105.

Mamdani, E.H. and S. Assilian, 1975. An experiment in linguistic synthesis with a fuzzy logic controller. Int. J. Man-Machine Stud., 7: 1-13. DOI: 10.1016/S0020-7373(75)80002-2

Momin, B. and P. Abhyankar, 2012. Current status and future research directions in monitoring vigilance of individual or mass audience in monotonous working environment. Int. J. Soft Comput., 3: 45-53. DOI: 10.5121/ijsc.2012.3204

Picot, A., S. Charbonnier, A. Caplier and N.S. Vu, 2012. Using retina modelling to characterize blinking: Comparison between EOG and video analysis. Mach. Vis. Appl., 23: 1195-1208. DOI: 10.1007/s00138-011-0374-4

Regan, M., 2010. Distraction du conducteur: D´efinition, m´ecanismes, effets et facteurs mod´erateurs. www.ipubli.inserm.fr/bitstream/handle/10608/220/?sequence=21

Renner, G. and S. Mehring, 1997. Lane departure and drowsiness-two major accidents causes-one safety system. Technical Report, Transport Resaerch Laboratory.

Said, S., B. Ben Amor, M. Zaied, C. Ben Amar and M. Daoudi, 2009. Fast and efficient 3D face recognition using wavelet networks. Proceedings of the 16th IEEE International Conference on Image Processing, Nov. 7-10, IEEE Xplore Press, Cairo, Egypt, pp: 4153-4156. DOI: 10.1109/ICIP.2009.5413446

Sarbjit, S. and P. Nikolaos, 1999. Monitoring driver fatigue using facial analysis techniques. Intelligent Transportation Systems, Japan.

Sharabaty, H., B. Jammes and D. Esteve, 2008. Eeg analysis using HHT: One step toward automatic drowsiness scoring. Proceedings of the 22nd International Conference on Advanced Information Networking and Applications – Workshops, Mar. 25-28, IEEE Xplore Press, Okinawa, Japan, pp: 826-831. DOI: 10.1109/WAINA.2008.271

Shin, H., S. Jung, J. Kim and W.Y. Chung, 2010. Real time car driver's condition monitoring system. Proceeding of the IEEE Sensors, Nov. 1-4, IEEE Xplore Press, Kona, HI, USA, pp: 951-954. DOI: 10.1109/ICSENS.2010.5690904

Snoun, A., I. Teyeb, O. Jemai and M. Zaied, 2017. A Multimodal Vigilance Monitoring System Based on Fuzzy Logic Architecture. In: Neural Information Processing, Liu D., S. Xie, Y. Li, D. Zhao and E.S. El-Alfy (Eds.), ICONIP 2017. Lecture Notes in Computer Science, vol 10637. Springer, Cham.

Song, F., X. Tan, X. Liu and S. Chen, 2014. Eyes closeness detection from still images with multi-scale histograms of principal oriented gradients. Patt. Recognit., 47: 2825-2838. DOI: 10.1016/j.patcog.2014.03.024

Teyeb, I., O. Jemai, M. Zaied and C. Ben Amar, 2014a. A novel approach for drowsy driver detection using head posture estimation and eyes recognition system based on Wavelet Network. Proceedings of the 5th International Conference on Information, Intelligence, Systems and Applications, Jul. 07-09, Chania, Greece, pp: 379 384. DOI: 10.1109/IISA.2014.6878809

Teyeb, I., O. Jemai, M. Zaied and C. Ben Amar, 2014b. A drowsy driver detection system based on a new method of head posture estimation. Proceedings of the 15th International Conference on Intelligent Data Engineering and Automated Learning, Sept. 10-12, Salamanca, Espagne, pp: 362-369. DOI: 10.1007/978-3-319-10840-7_44

Teyeb, I., O. Jemai, M. Zaied and C. Ben Amar, 2015a. A multi level system design for vigilance measurement based on head posture estimation and eyes blinking. Proceedings of the SPIE 9875, 8th International Conference on Machine Vision, (CMV' 15).

Teyeb, I., O. Jemai, M. Zaied and C. Ben Amar, 2015b. Vigilance measurement system through analysis of visual and emotional drivers signs using wavelet networks. Proceedings of The 15th International Conference on Intelligent Systems Design and Applications, Dec. 14-16, IEEE Xplore Press, Marrakech, Maroc, pp: 140-147. DOI: 10.1109/ISDA.2015.7489215

Teyeb, I., O. Jemai, T. Bouchrika and C. Ben Amar, 2013. Detecting driver drowsiness using eyes recognition system based on wavelet network. Proceedings of the 5th International Conference on Web and Information Technologies, May 09-12, Hammamet, Tunisia, pp: 245-254.

Teyeb, I., O. Jemai, M. Zaied and C. Ben Amar, 2016. Towards a smart car seat design for drowsiness detection based on pressure distribution of the driver's body. Proceedings of the International Conference on Software Engineering Advances (ICSEA 2016), Rome-Italy, pp: 217-222.

Viola, P. and M. Jones, 2001. Robust real-time object detection. Int. J. Comput. Vis., 57: 137-154. DOI: 10.1023/B:VISI.0000013087.49260.f

Volvo, 2013. European accident research and safety report, volvo trucks. http://pnt.volvo.com/pntclient/loadAttachment.aspx?id=27116

Zaied, M., C. Ben Amar and A.M. Alimi, 2005. Beta wavelet networks for face recognition. J. Dec. Syst., 14: 109-122. DOI: 10.3166/jds.14.109-122

Zaied, M., O. Jemai and C. Ben Amar, 2008. Training of the Beta wavelet networks by the frames theory: Application to face recognition. Proceedings of the 1st Workshops on Image Processing Theory, Tools and Applications, Nov. 23-26, IEEE Xplore Press, Sousse, Tunisia, pp: 165-170. DOI: 10.1109/IPTA.2008.4743756

Zaied, M., S. Said, O. Jemai and C. Ben Amar, 2011. A novel approach for face recognition based on fast learning algorithm and wavelet network theory. Int. J. Wavelets, Multiresolut. Inform. Process., 9: 923-945. DOI: 10.1142/S0219691311004389

Zhou, M., H. Wei and S. Maybank, 2006. Gabor wavelets and adaboost in feature selection for face verification. Proceedings of the Workshop in Conjuction with Applications of Computer Vision, (ACV' 06), Graz, Austria, pp: 101-109.