Driver Fatigue Detection based on Eye State Recognition

Fang Zhang
School of Electronics and Information Engineering
Tianjin Key Laboratory of Optoelectronic Detection Technology and Systems
Tianjin Polytechnic University
Tianjin, China
hhzhangfang@126.com

Jingjing Su
School of Electronics and Information Engineering
Tianjin Key Laboratory of Optoelectronic Detection Technology and Systems
Tianjin Polytechnic University
Tianjin, China
sujingjing0609@gmail.com

Lei Geng*
School of Electronics and Information Engineering
Tianjin Key Laboratory of Optoelectronic Detection Technology and Systems
Tianjin Polytechnic University
Tianjin, China
genglei@tjpu.edu.cn

Zhitao Xiao
School of Electronics and Information Engineering
Tianjin Key Laboratory of Optoelectronic Detection Technology and Systems
Tianjin Polytechnic University
Tianjin, China
xiaozhitao@tjpu.edu.cn

***Abstract*—Driving fatigue is a main factor caused the traffic accidents. Our faces contain a lot of useful information, we can use the state of eyes to detect the fatigue, but the eye state would be affected by wearing sunglasses. In this paper, to solve above problems and make the algorithm keep the accuracy and real-time at the same time, we use the infrared videos for detecting and propose an eye state recognition method based on convolution neural network (CNN), eventually calculating percentage of eyelid closure over the pupil over time (PERCLOS), blink frequency to detect the fatigue. The experimental results show that the proposed method has high recognition accuracy of state of eyes when wearing glasses and can detect the fatigue effectively.**

Keywords- Fatigue detection; Eye state recognition; CNN; PERCLOS

I. INTRODUCTION

With the rapid growth of traffic accidents, it has brought serious losses to the state and personal property. Research shows that the fatigue driving is one of the main causes of traffic accidents. In recent years, many countries and governments already pay attention to the driving safety problems. Researches for driving fatigue detection have the vital significance.

Fatigue detection methods based on computer vision is a non-intrusive way. The facial features can be calculated by analyzing the changes of facial expression, such as blinking, eye closure duration, yawning and so on.

The calculation of blink frequency depends on the reliable recognition of driver's eye state which is one of the key functionalities of fatigue detection. Eye closure is widely used to calculate the fatigue parameters as PERCLOS and blink. The recognition of eye state is an important step in fatigue detection. The changing of illumination and wearing sunglasses will affect the states detection of eyes.

The method uses the vertical and horizontal projection intensity of the image to detect the state of eyes; it will make the eye state detection inaccurate when wearing sunglasses. The set of methods involves relying on the traditional classification, such as AdaBoost classifier [1], LBP features and SVM classifier [2], and so on. In these methods, select the appropriate feature or not will be the key factor that affects the result of classification. The method based on CNN [3] has higher recognition accuracy. CNN can extract features adaptively and make the features have a better ability of characterize, CNN has better features expressive, avoiding the manual feature selection.

In our paper, we design a driver face image acquisition system based on infrared videos and propose a recognition method about the state of eyes based on CNN. As shown in Figure 1, the method mainly includes three parts: extract the area of eye, state of eye recognition, fatigue detection.

The rest of the paper is organized as follows: Eye area is located based on face detection and facial landmarks detection in next section, Section 3 shows the algorithm of the CNN. Calculate the fatigue parameters with the state of eyes in section4. Section 5 shows the experimental results for driver fatigue detection. Finally conclusion is in section 5.

978-1-5090-4993-6/17 $31.00 © 2017 IEEE
DOI 10.1109/CMVIT.2017.25

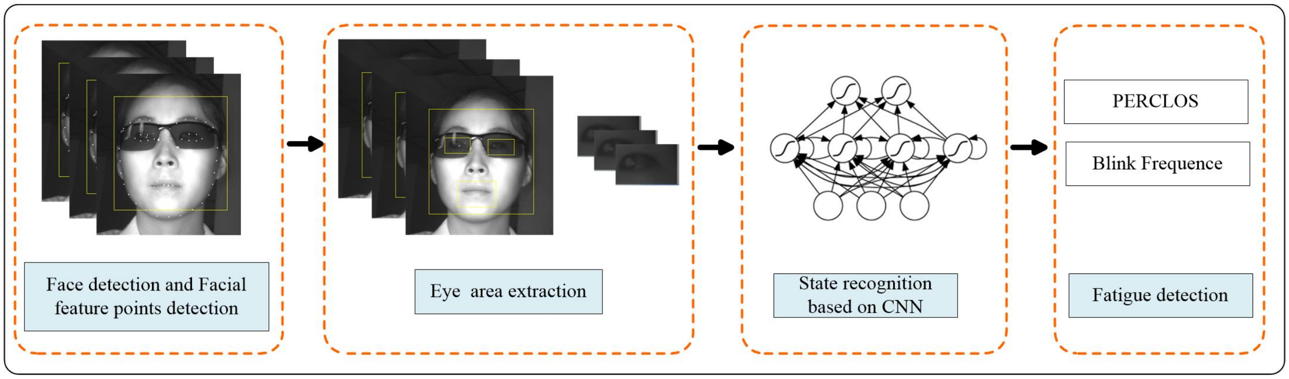

Figure 1. Algorithm block diagram

II. EXTRACTION THE AREA OF EYE

We use the AdaBoost algorithm to detect the faces, and as a classical analytical method, regression analysis can predict and easily build mapping information from input to output. It can be directly used for the extraction of facial landmarks [4]. Ren et al. detected facial landmarks by cascade regression, with a set of local binary features (LBF) which learned by a locality principle [5]. Each facial landmark was learned by a series of highly discriminative local binary features. The linear regression was composed by the obtained local binary features as the final output. The shape regression method predicted the facial shape S by cascaded refined by a shape increment ΔS stage-by-stage. A shape increment at stage ΔS^t is given in (1):

$$\Delta S^t = W^t \Phi^t(\mathrm{I}, \mathrm{S}^{t-1}) \tag{1}$$

Where I is the input image, S^{t-1} is the shape from the previous stage, Φ_t is a feature mapping function, and W^t is a linear regression matrix. The feature mapping function is composed of a set of local feature mapping functions $\Phi^t = [\phi_1^t, \phi_2^t, ..., \phi_L^t]$ (L is the number of landmarks), each ϕ_l^t is learned by independently regressing l th landmark.

As shown in Figure 2 is the detection result of face and facial landmarks.

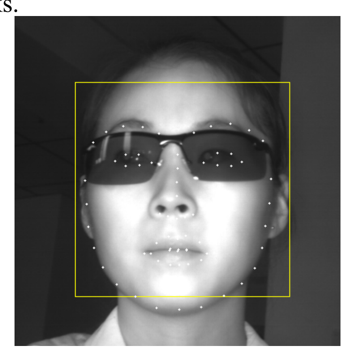

Figure 2. The detection result

With highly sparse of LBF feature, the speed of feature extraction and regression is extremely rapid, in this paper, we extract the area of eyes based on the facial landmarks as shown in (2). Figure 3 is the extraction result of eye area.

$$\begin{cases} W\_e = 1.6 * \mathrm{X}\_e \\ H\_e = 2 * Y\_e \end{cases} \tag{2}$$

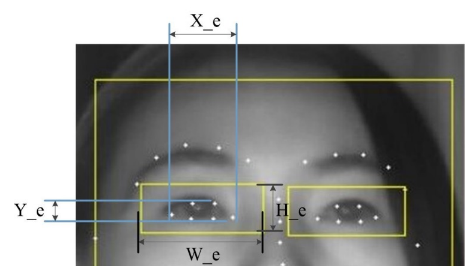

Figure 3. The extraction result of the eye area

III. CONVOLUTIONAL NEURAL NETWORK

The structure of CNN includes convolution layer, pooling layer and fully connected layer. To reduce the number of neurons and the weights, Convolution by local receptive fields, parameter sharing and pooling neural network were used to optimize the structure. CNN autonomously learn multiple levels of representations of the data through their layered structure which enables complex features to be learned [6]. The network structure is described as Figure 4.

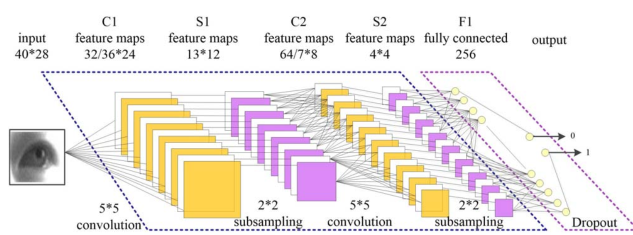

Figure 4. structure of the convolution neural network

A. Convolutional Layer

Convolution layer is the core structure of CNN. As shown in Fig. 3, two convolutional layers are used in our proposed network. Each convolution layer contains some meaningful features and computes the convolution of the input feature with a fixed-size kernel as described in (3) to get a feature map.

$$x_j^l = f(\sum_{i\in M_j} x_i^{l-1} * k_{ij}^l + b_j^l) \quad (3)$$

Where x_j^l is the j th feature map at layer l , $f(\bullet)$ represents activation function, k_{ij}^l is the weight connecting neuron i at layer $l-1$ to neuron j at layer l . The convolution kernel is determined by these trainable weights k_{ij}^l . b_j^l is the j th bias of the feature map at layer l , M_j represents the collection of the input feature map.

The convolution layer makes it possible to combine different local structures and present more useful details within a region. Figure 5 is the feature map that input image computed by different convolution kernels at layer C1. It is clearly seen in Figure 5 that the different convolution kernels generate different feature maps.

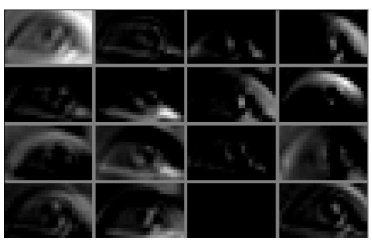

Figure 5. The feature map of convolution layer

B. Pooling Layer

The common algorithm of pooling includes max pooling, average pooling, etc. It has the advantages of invariance to transformations, better robustness to noise, and more compact [7]. Max pooling is more robustness when there exists ambient noises. So max pooling is adopted here instead of other pooling methods. The formula for the pooling is (4).

$$x_j^l = f\left(\beta_j^l down\left(x_j^{l-1}\right) + b_j^l\right) \quad (4)$$

Where $down(\bullet)$ represents sampling function, β indicates a weight coefficient, b is the bias of the output features, $f(\bullet)$ is the activation function of the neurons.

We use two networks, one net used the pooling layer and another not use. Table 1 shows the results (the iterative 100000 times). The result shows the performance of the pooling that reduces the amount of calculation.

TABLE 1 THE TEST RESULT OF POOLING LAYER

Network Structure	Training time/min	Accuracy
with pooling	42	98.7%
no pooling	106	95.2%

C. Fully Connected Layer

Fully connected layer will generate more parameters. Since the feature dimension has been reduced by the layers of convolution, greatly reduces the amount of computation. The output of each neuron is described in (5).

$$h_{W,b}(x) = f(\mathrm{W}^T x + b) \quad (5)$$

Where x is the input neurons, $h_{W,b}(x)$ is the output of neurons, W is the weight of connecting neurons, b is the bias.

D. Activation Functions

Sigmoid function and tanh function are commonly used non-linear activation functions, but these functions exist the gradient vanishing, to overcome this problem, we use the ReLU function (Rectified linear unit) [8] which is defined as (6)：

$$\mathrm{ReLU}(x) = \max(0, x) \quad (6)$$

The network can get sparse expression after the ReLU function, with the advantage of unilateral suppression. Its performance is usually better than the other activation function. Figure 6 is the RELU function.

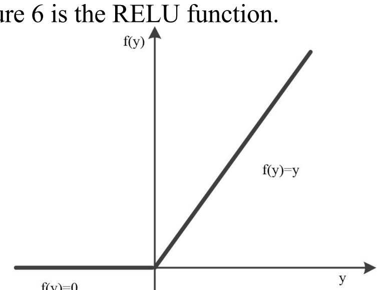

Figure 6. The RELU function

IV. FATIGUE DETECTION

We test a video to recognize the state of eyes detected by CNN, the score result of eye state is shown in Figure 7. It is degree of the eyes open, when the score closes to the ‘1’, the state of eye is more likely to open.

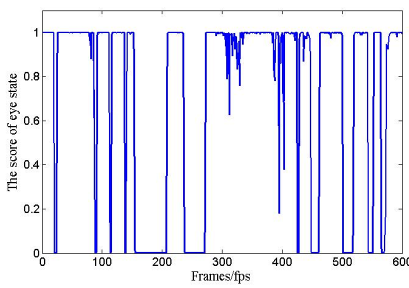

Figure 7. The eye state score result

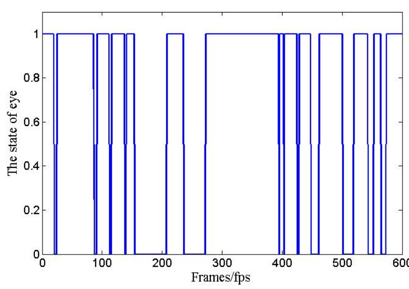

Figure 8. The eye state result

The Figure 8 is the actually output of CNN classification results. '1' represents eye-open, '0' represents eye-close.

A. PERCLOS

PERCLOS (percentage of eyelid closure over the pupil over time) is a parameter that is used to detect driver fatigue [9]. It is calculated as (7):

$$f_{PERCLOS} = \frac{n_{\text{close}}}{N_{total}} \times 100\% \tag{7}$$

Where the total number of eye-close frames is n_{close}, N_{total} is the total number of frames over a period time.

B. Blink frequency

According to the survey, when driver is awake, its eyes blink frequency is about 15 ~ 30 times/min, 0.25 ~ 0.3 seconds every time. When the blink frequency is less than 7 times/min, the driver may be fatigue.

With the result of eye state, we can count the numbers that eye-open changes to eye-close, then compute the blink frequency over a period of time.

$$f_{BFreq} = n / N \tag{8}$$

Where n is the eye-open to eye-close number, the N is the total number of frames over a period time.

V. EXPERIMENTS

A. Experimental Platform

VS2012 and opencv2.4.9 are used to carry out all the operations, running on a Win7 system with Intel (R) Core (TM) i7-6700HQ, CPUs (2.60GHz), 8GB memory.

B. Database and Data Preparation

In order to overcome the influence of light on image, make the system work at any time of the day and night and detect the eyes when the driver is wearing sunglass, the image acquisition system used active infrared light (850nm) fill light illumination and combined with 850nm narrowband filter to get the images under the infrared illumination.

We build a facial video database (IRF database) under the infrared illumination, include four categories (simple, glass, night vision goggles, sunglass). Figure 9 is facial images collected by acquisition system based on infrared illumination. In Figure 9, the first line is the color images collected under the normal illumination, and the second line is images captured under the infrared illumination.

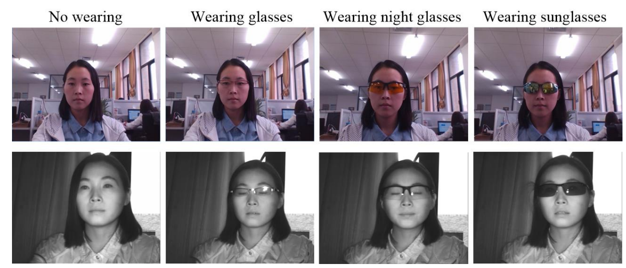

Figure 9. Sample images of the database

IRF Database: There is 160 video clips in IRF database from 20 individuals, eight clips for each one divided into two parts of train and test.

ZJU Database: The database is established by Wu et al., including 7 females and 13 males which captured under natural light conditions [10].

We select images in each blinking process, including eye images of open and closed. The eye images are then divided into two separate sets for the purpose of training and testing.

Select 7732 images as training samples, 5 pictures of eye open, 2585 pictures of eye open, the image is converted into a gray scale image (40×28 pixels). Parts of the training samples are shown in Figure 10.

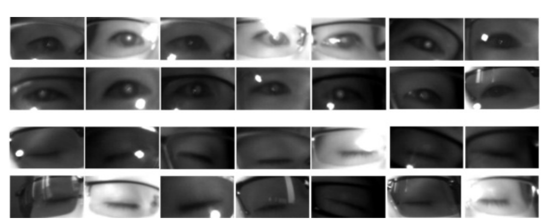
Figure 10. Parts of the training samples

C. *Experimental results and analysis*

The eye state recognition rate of the network with the increase in the number of iterations when training the samples, the result is shown in Figure 11.

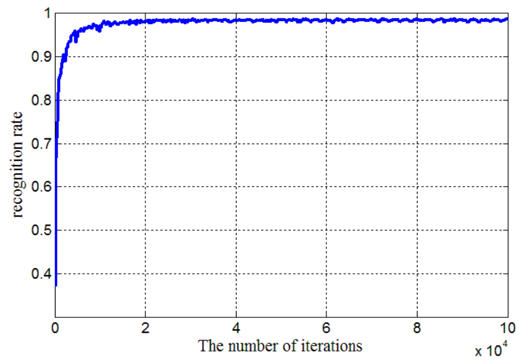

Figure 11. The result of recognition rate

The test result for the task of eye state recognition is shown in Table 2 with above databases.

TABLE 2 THE TEST RESULT OF EYE STATE

Database	State	Test number	Accuracy
IRF	open	2588	99.69%
(no glasses)	close	1550	99.74%
IRF	open	2495	98.04%
(with glasses)	close	930	96.02%
ZJU	open	2475	98.83%
(no glasses)	close	638	99.22%
ZJU	open	3202	96.69%
(with glasses)	close	1053	98.96%

As experimental results shown, our proposed method can classify excellently for the task of eye state recognition, despite of wearing glasses or sunglasses.

Through statistical 9 test videos includes 6844 frames of 320 * 240 images, computing the average time-consuming of the method include each module and overall time. Table 3 is the time-consuming result. The method complies with the requirement of real-time.

TABLE3 THE RESULT OF TIME CONSUMING

	Eye area extraction	State recognition	Fatigue detection	Total
test 1	43.61ms	4.87ms	0.0012ms	61.25ms
test 2	41.29ms	5.73ms	0.0012ms	58.29ms
test 3	40.91ms	5.64ms	0.0011ms	56.40ms
test 4	43ms	5.87ms	0.0011ms	55.52ms
test 5	36.14ms	5.52ms	0.0013ms	62.31ms
test 6	40.23ms	4.94ms	0.0010ms	48.93ms
test 7	40.82ms	6.23ms	0.0013ms	56.37ms
test 8	38.79ms	5.71ms	0.0011ms	65.15ms
test 9	38.97ms	5.91ms	0.0011ms	63.26ms
Time/ms	40.42ms	5.60ms	0.0012ms	58.61ms

Figure 12 shows PERCLOS result, when the driver is fatigue, the PERCLOS value is bigger than normal. In this paper, the PERCLOS threshold is set to 0.25, then combine with the parameter of blink frequency to detect the fatigue.

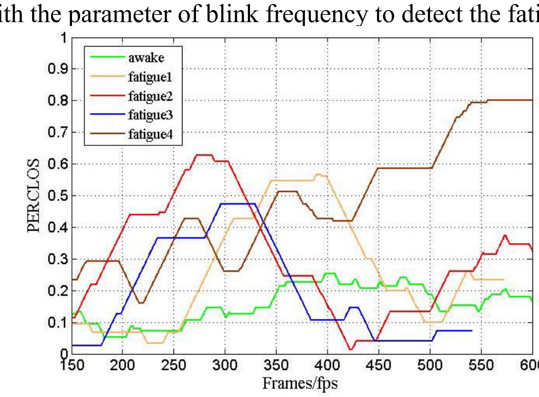

Figure 12. The PERCLOS value

Figure 12 shows the PERCLOS result, when the driver is fatigue, the PERCLOS value is bigger than normal. Walter W. Wierwille gives the PERCLOS threshold 0.075 and 0.15 as the standards to distinguish the states of awake, suspected fatigue, fatigue [11].

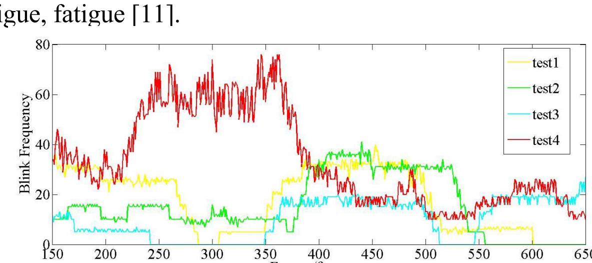

Figure 13. The result of blink frequency

Figure 13 is the result of blink frequency, the value of blink frequency is floating obviously, not appropriate as the main fatigue parameter, but if the value is too low, it is likely to be fatigue. When driver is in fatigue, the blink frequency is lower than 10, so in this paper we set the threshold of blink frequency to $t_{bf} = 10$.

In this paper, we combine the PERCLOS and blink frequency as the fatigue parameters. The PERCLOS thresholds are set to $t_{p1} = 0.25$ and t_{p2}=0.15 , then combine with the parameter of blink frequency to detect the fatigue. When PERCLOS is higher than 0.25, the driver state is judged to be fatigue; when PERCLOS is less than 0.2 but bigger than 0.15, if blink frequency less than 10, the driver state is judged to be fatigue; otherwise awake. It is described as (9):

$$\begin{cases} fatigue & f_{PER} > t_{p1} \cup t_{p2} < f_{PER} < t_{p1} \,||\, f_{BFreq} < t_{bf} \\ awake & others \end{cases} \quad (9)$$

In order to verify the validity of the method, we test the videos with the above principle, the experiment results is shown in Table 4.

TABLE4 THE RESULT OF FATIGUE DETECTION

	Actually occurred	Correct detection	Error detection	Correct detection accuracy	Correct detection accuracy
No wearing glasses	167	158	9	95.81%	4.19%
Wearing glasses	152	139	13	91.45%	8.55%

The experimental results show that the fatigue detection method can effectively detect the fatigue. However, when wearing glasses, the detection of eyes state are disturbed, in the future work, we can use the others fatigue information, such as the yawning, head posture, gazing direction, and so on.

VI. CONCLUSION

In this paper, we propose a method for driver fatigue detection based on eye state recognition. The method of eye state recognition provides high accuracy. By using the state of eyes can calculate the PERCLOS and blink frequency parameter, we can detect the driver fatigue early enough to avoid the accidents. Experimental results show that our method yields a much more robust and accurate state recognition. The method can work in condition of wearing glasses. In the future, we intend to combine other fatigue parameters to our system and optimize the parameters of the current model to improve the real-time performance and detection accuracy.

ACKNOWLEDGMENT

This work is supported by National Natural Science Foundation of China under grant No. 61302127, the key technologies R & D program of Tianjin under grant No.14ZCZDGX00033, and research fund for the doctoral program of higher education of China under grant No. 20131201110001.

REFERENCES

[1] Shi-Feng, X. U., & Zeng, Y. (2007). Eyes state detection based on adaboost algorithm. Computer Simulation, 24(7), 214-217.

[2] Yao Sheng,Li Xiaohua,Zhang Weihua,Zhou Jiliu (2015) . Eyes state detection method based on LBP. Application Research of Computers, 32(6):1897-1901.

[3] Lecun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., & Hubbard, W., et al. (1989). Backpropagation applied to handwritten zip code recognition. Neural Computation, 1(4), 541-551.

[4] LI Yue-Long ,JIN Yan ,WANG Jian-Ming , XIAO Zhi-Tao , GENG Lei (2015). Face Feature Points Extraction: A Review, Chinese Journal of Computers, Vol.38: Online Publishing No.137.

[5] Ren, S., Cao, X., Wei, Y., & Sun, J. (2014). Face alignment at 3000 fps via regressing local binary features. IEEE Transactions on Image Processing, 25(3), 1685-1692.

[6] Dieleman, S., Willett, K. W., & Dambre, J. (2015). Rotation-invariant convolutional neural networks for galaxy morphology prediction. Monthly Notices of the Royal Astronomical Society, 450(2), 1441-1459.

[7] Boureau, Y. L., Ponce, J., & Lecun, Y. (2010). A Theoretical Analysis of Feature Pooling in Visual Recognition. International Conference on Machine Learning (Vol.32, pp.459-459).

[8] Nair, V., & Hinton, G. E. (2010). Rectified Linear Units Improve Restricted Boltzmann Machines Vinod Nair. International Conference on Machine Learning (pp.807-814).

[9] Wierwille, W. W., Ellsworth, L. A., Wreggit, S. S., Fairbanks, R. J., & Kim, C. L. (1994). Research on vehicle-based driver status/performance monitoring:development, validation, and refinement of algorithms for detection of driver drowsiness. National Highway Traffic Safety Administration, Final Report: DOTHS 808 247.

[10] Pan, G., Sun, L., Wu, Z., & Lao, S. (2007). Eyeblink-based Anti-Spoofing in Face Recognition from a Generic Webcamera. IEEE International Conference on Computer Vision (pp.1-8).

[11] Knipling, R. R., & Wierwille, W. W. (1994). Vehicle-based drowsy driver detection: Current status and future prospects. Moving Toward Deployment. Proceedings of the IVHS America Annual Meeting. 2 Volumes.