

University of Wollongong
Research Online

Faculty of Engineering and Information
Sciences - Papers: Part B

Faculty of Engineering and Information
Sciences

2019

Eye state recognition method for drivers with glasses

Lei Geng

Tianjin Polytechnic University

Haibing Yin

Tianjin Polytechnic University

Zhitao Xiao

Tianjin Polytechnic University, zhitaoxiao@tom.com

Jiangtao Xi

Tianjin Polytechnic University, University of Wollongong, jiangtao@uow.edu.au

Follow this and additional works at: <https://ro.uow.edu.au/eispapers1>



Part of the [Engineering Commons](#), and the [Science and Technology Studies Commons](#)

Recommended Citation

Geng, Lei; Yin, Haibing; Xiao, Zhitao; and Xi, Jiangtao, "Eye state recognition method for drivers with glasses" (2019). *Faculty of Engineering and Information Sciences - Papers: Part B*. 3173.
<https://ro.uow.edu.au/eispapers1/3173>

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

Eye state recognition method for drivers with glasses

Abstract

Eye state recognition is a key step in fatigue detection method. However, factors such as occlusion of different types of glasses and changes in lighting conditions may have some impact on eye state recognition. In order to solve these problems, a driver's eye state recognition method based on deep learning is proposed. Firstly, the driver's face images are acquired using an infrared acquisition device. Secondly the multi-task cascaded convolution neural networks are used to detect the face bounding box and feature points of the driver's face image, and then the eye regions are extracted. Finally the Convolution Neural Network (CNN) is adopted to identify the open and closed state of the eyes. Experimental result shows that the proposed method can accurately identify the state of eyes and help to calculate the fatigue parameters of drivers.

Disciplines

Engineering | Science and Technology Studies

Publication Details

L. Geng, H. Yin, Z. Xiao & J. Xi, "Eye state recognition method for drivers with glasses," *Journal of Physics: Conference Series*, vol. 1213, pp. 052049-1-052049-7, 2019.

PAPER • OPEN ACCESS

Eye state recognition method for drivers with glasses

To cite this article: Lei Geng *et al* 2019 *J. Phys.: Conf. Ser.* **1213** 052049

View the [article online](#) for updates and enhancements.



IOP | ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

Eye state recognition method for drivers with glasses

Lei Geng^{1,2}, Haibing Yin^{1,2}, Zhitao Xiao^{1,2*} and Jiangtao Xi^{1,2,3}

¹ School of Electronics and Information Engineering, Tianjin Polytechnic University, Tianjin, Tianjin, 300387, China

² Tianjin Key Laboratory of Optoelectronic Detection Technology and System, Tianjin Polytechnic University, Tianjin, Tianjin, 300387, China

³ School of Computer and Telecommunications Engineering, University of Wollongong, Wollongong, New South Wales, NSW 2522, Australia

*Corresponding author's e-mail: xiaozhitao@tjpu.edu.cn

Abstract. Eye state recognition is a key step in fatigue detection method. However, factors such as occlusion of different types of glasses and changes in lighting conditions may have some impact on eye state recognition. In order to solve these problems, a driver's eye state recognition method based on deep learning is proposed. Firstly, the driver's face images are acquired using an infrared acquisition device. Secondly the multi-task cascaded convolution neural networks are used to detect the face bounding box and feature points of the driver's face image, and then the eye regions are extracted. Finally the Convolution Neural Network (CNN) is adopted to identify the open and closed state of the eyes. Experimental result shows that the proposed method can accurately identify the state of eyes and help to calculate the fatigue parameters of drivers.

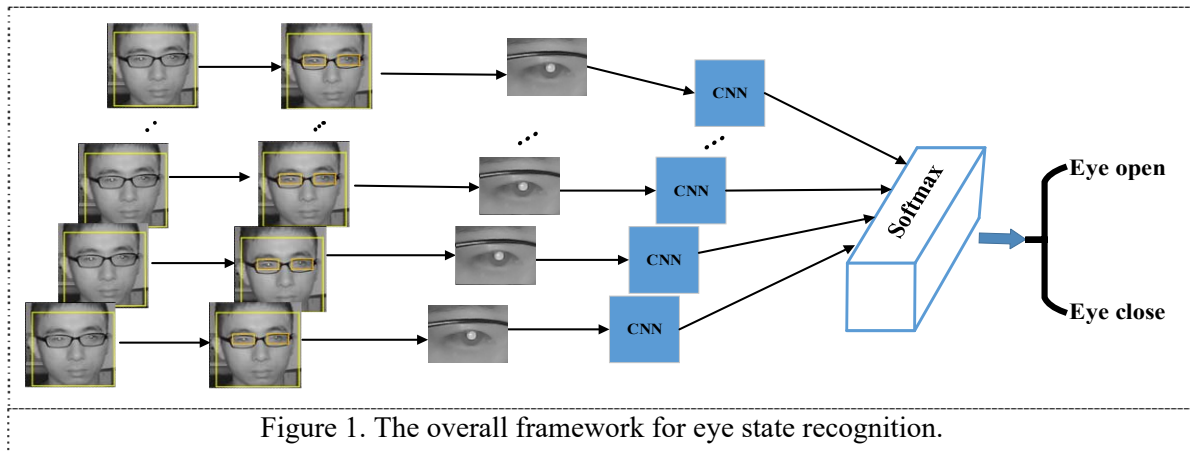
1. Introduction

With the continuous improvement of living conditions and the increasing number of cars, traffic accidents occurs more frequently. In order to avoid traffic accidents caused by fatigue driving, fatigue judgment can be made based on the driver's facial image. The state of the eye changes when a person is tired. Therefore, the driver's eye state recognition technology is of great significance for preventing traffic accidents. There are many ways to identify the state of the eye. The template matching method were used to determine the state of the eye [1]. Since the position of the iris in the eyelid is not fixed, it is easy to cause false detection. Multi-template matching results in low detection efficiency and poor real-time performance. The gray-scaled projection curve of the iris region of the eye was used to judge the state of the eye [2], it had a good detection effect when the iris information is perfect, but it had strict requirements on illumination. On the basis of face detection, according to the good clustering of eye white in YCbCr space, the Gaussian eye white segmentation model was established [3], and the white area of the eye was used as the eye opening and closing index, the algorithm had a low complexity, but it was sensitive to changes in illumination. An eye opening and closing detection method based on the combination of LBP and SVM was proposed [4]. Although this method has a high detection rate, it has certain limitations when the driver wears sunglasses and the posture changes. For the feature in each case, the traditional classifier needs to manually select the appropriate features. Whether the feature selection is appropriate, this will become a key factor that constrains the effect of



the classifier. Features can be adaptively extracted by using CNN, which makes the classification effect more superior. The method is now widely used.

In recent years, the deep learning algorithm represented by CNN which has powerful feature extraction capabilities and robustness, has been successfully applied to visual tasks such as target categorization, object segmentation, target detection and so on. This paper locates the face image through the multi-task cascaded convolutional neural networks (MTCNN) [5], and then the eye regions are extracted, and finally the eye state through the CNN are recognized. The overall framework for eye state recognition based on CNN is shown in Figure 1.



2. Design and effect analysis of image acquisition system

2.1. Infrared video acquisition system

A high quality face image is a necessary condition for recognizing the driver's eye state, it is essential to acquire high quality face images. During the driving process, due to changes in lighting conditions and face posture, occlusion of different types of glasses, and the complexity of the imaging environment, the difficulty in eye state recognition is increased. Aiming at the above problems, a driver's facial image acquisition system based on infrared illumination is designed in the paper, which is suitable for different lighting conditions and glasses occlusion. The image acquisition system is based on an infrared camera, which is matched with an infrared light source and a filter to optimize image acquisition. When the lighting conditions are poor, infrared light can enhance the image quality and also have good transparency to the glasses due to filters have the function of filtering and selecting spectral lines [6]. The infrared video acquisition system is shown in Figure 2.

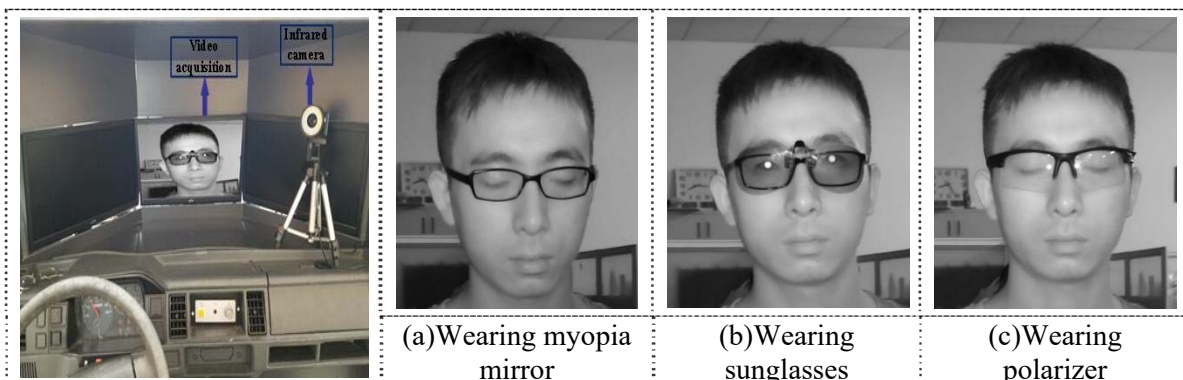


Figure 2. Infrared video acquisition system.

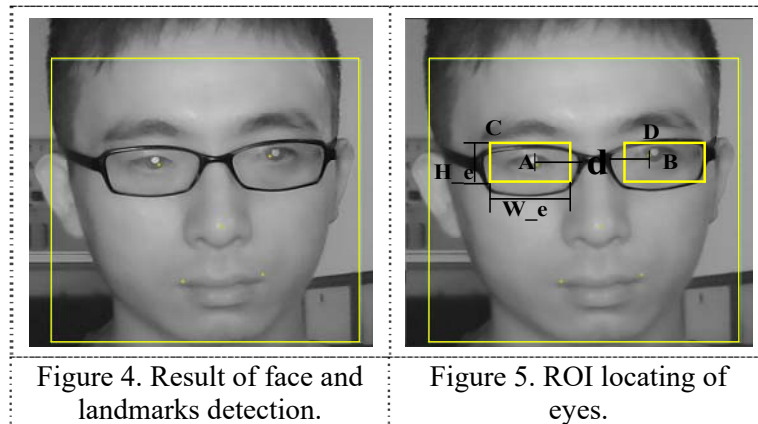
Figure 3. Data samples under different conditions.

2.2. Experimental platform and dataset preparation

The experimental environment is Windows 7 operating system, Intel Core i7-6700, CPUs (3.40GHz), and 16GB of memory. We collected infrared face videos of 25 people were under test wearing myopia mirror, wearing polarizer and wearing sunglasses (including 15 male testers and 10 female testers) as datasets through an infrared acquisition system and named the data set "TJPU-FDD". Figure 3 is the infrared face images of the people under three conditions. In each case, the frame rate of the video can reach 30 fps. A total of 100 videos with a length of about 120 seconds are collected, and the video image size is 1024×768.

2.3. The eye region extraction

Face detection and feature point location are important components of the eye state recognition method. Locating the eye region accurately is the main research goal. The MTCNN method firstly locates the face bounding box and facial landmarks, then unifies the face detection and feature point positioning into the multi-task framework, finally three sub-networks and non-maximum suppression are used to complete sample mining during training. This method can improve the accuracy of detection. The face and feature point detection results are shown in Figure 4.



The center position of the eye refers to the feature point coordinates of the pupil position of both eyes in the MTCNN. The coordinates of the region of interest (ROI) can be obtained by combining the feature point coordinates with the position constraints. The eye region solving rule is as shown in the following equations:

$$\begin{cases} d = x_B - x_A \\ W_e = 0.5 * d \\ H_e = 0.5 * W_e \\ x_C = x_A - 0.5 * W_e \\ y_C = y_A - 0.5 * H_e \\ x_D = x_B - 0.5 * W_e \\ y_D = y_B - 0.5 * H_e \end{cases} \quad (1)$$

where (x_A, y_A) and (x_B, y_B) denote the coordinates of eye feature point A, B, respectively, and the horizontal distance between point A and point B is d . The two points C and D are the upper left vertex of the eye area, and their coordinates are (x_C, y_C) and (x_D, y_D) , respectively. W_e and H_e refer to the width and height of the extracted eye region, respectively. The eye region is positioned as shown in Figure 5.

2.4. Results of eyes ROI location

In this paper, in order to verify the effectiveness of the eye region extraction algorithm, positioning experiments were performed on TJPU-FDD datasets in three conditions (myopia mirror, polarizer, and sunglasses). Then, the positioning accuracy of the eye region is calculated. Table 1 shows the results of eye region localization under different conditions. It can be seen from the experimental results that the proposed method has higher accuracy under different illumination and occlusion conditions, and its average accuracy can reach 97.88%.

Table 1. Result of eyes ROI location.

Dataset	Model	Eye location	Test numbers	Accurate numbers	Accuracy (%)
TJPU-FDD	Myopia mirror	Left	5320	5230	98.30
		Right	5320	5238	98.45
TJPU-FDD	Sunglasses	Left	5608	5445	97.09
		Right	5608	5455	97.27
TJPU-FDD	Polarizer	Left	4983	4885	98.03
		Right	4983	4889	98.11

3. Eye state recognition network

As a kind method of deep learning, CNN has become a research hotspot in the field of image processing. A two-dimensional image is directly input to CNN, and then automatically learns the image features and the implicit relationship inside the data, which ultimately avoids complex preprocessing of the image and has a high classification recognition rate. In the paper, the network structure of CNN includes two convolution layers, two pooling layers, and a fully connected layer. The ESR-Net structure of eye state recognition in this paper is shown in Figure 6.

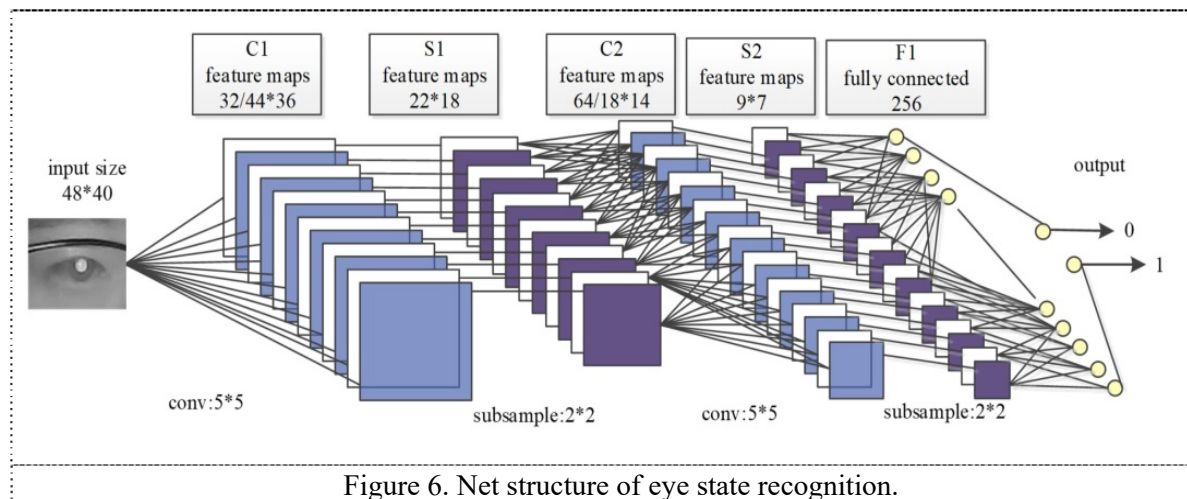


Figure 6. Net structure of eye state recognition.

3.1. Convolutional layer

The convolution layer is the core part of the CNN, which is used for feature extraction and feature mapping by using the convolution kernel, and then the local features in different positions of the image are extracted by using local connection and weight sharing. For a neuron, the extracted local features in the same location of different feature maps. Each one of the convolution layers contains a plurality of feature planes. Firstly, the image of the previous layer is convolved by the convolution kernel, and then the bias is added to obtain the feature map of the current layer. Different convolution kernels extract features in different positions of the input image by using a "sliding window". The meaningful features can be extracted by convolution kernels in network training.

3.2. Pooling layer

The pooling layer is usually set after the convolution layer, which the purpose is to reduce the resolution of each feature map in the convolution layer. Thereby, the feature selection of the convolution layer is realized, the complexity of the network calculation is reduced, and the feature is constant. Pooling can be divided into average pooling and max pooling [7]. Max pooling method is mainly used in CNN which central idea is to first divide the image into rectangular regions that do not intersect each other, and then use the largest feature of each region to represent the convolution characteristics after pooling. The general expression of pooling layer is shown in the following equation:

$$x_j^l = f(\beta_j^l \text{down}(x_j^{l-1}) + b_j^l) \quad (2)$$

where $\text{down}(\cdot)$ is the sampling function, β and b are the biases of the output features.

3.3. Fully connected layer

Every neuron on the fully connected layer is interconnected with all neurons in the upper layer of the feature map. Compared to the method of the convolution layer local linking, the full connectivity of the fully connected layer produces more network parameters. The role of the fully connected layer is to re-fit at the end of the CNN, which can reduce the loss of feature information. The previous convolution and pooling layers have reduced the feature dimension, which greatly reduces the computational complexity of the fully connected layer. The output of each neuron can be formulated as:

$$h_{W,b}(x) = f(W^T x + b) \quad (3)$$

where x is the input of the neuron, $h_{W,b}(x)$ is the output of the neuron, W is the connection weight, b is the bias, and $f(\cdot)$ is the activation function.

4. Experimental results and analysis

In order to identify the open and closed state of eyes, the features of the infrared image are extracted in the state recognition network. 40,000 images were selected from the datasets as experimental samples, in which the number of images opened and closed by eyes was 20,000 and 20,000, respectively. The iteration numbers are 100,000, and the basic learning rate is 0.01. During the training process, the experimental data samples are grayed. The extracted image of the eye region is uniformly normalized to 48×40 . Finally, the results of the previous processing are input to ESR-Net for eye state recognition. Parts of the training samples are shown in Figure 7.

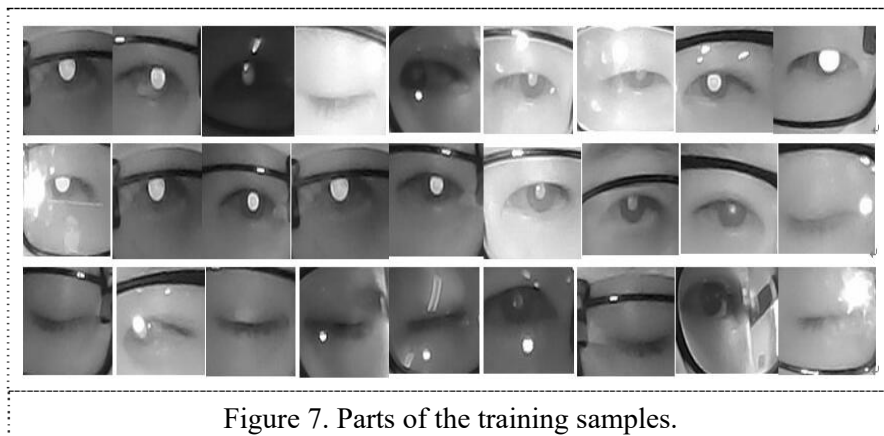


Figure 7. Parts of the training samples.

The network parameters are continuously optimized when the iterative number increases in the process of training experimental samples, and the accuracy of eye state recognition is also improved. When the number of training is sufficient, the recognition rate changes slowly and becomes convergent, and the classification gradually becomes stable. Increasing the pooling layer network can improve the recognition rate and enhance the network generalization ability. The curve of the ESR-Net network recognition accuracy and loss value are shown in Figure 8.

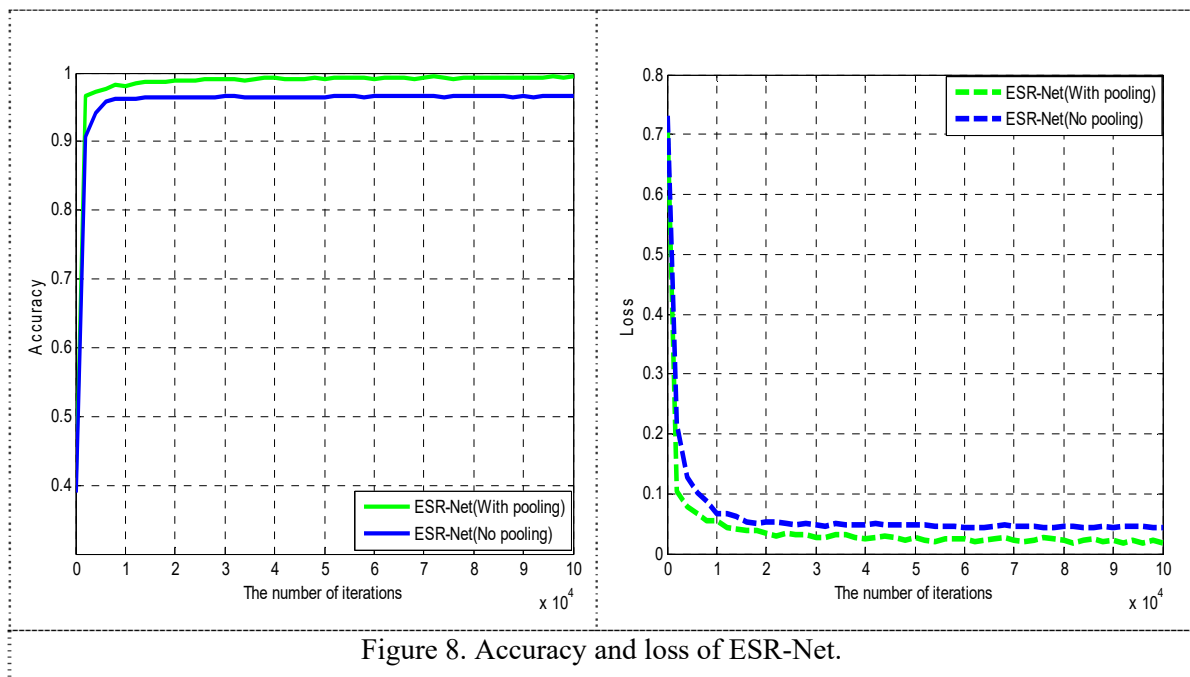


Figure 8. Accuracy and loss of ESR-Net.

To verify the accuracy of the method in this paper, the model generated using the ESR-Net network structure iteration 100,000 times is tested. The test results of eye state recognition are shown in Table 2. It can be seen from Table 2 that the network model has a high recognition rate for eye states in different types of glasses, and the average recognition rate can reach 98.05%. However, the state recognition rate of the eyes is lower than the other two conditions when wearing sunglasses. The reason is that the optical properties of different glasses are different, which is intervened the recognition of some pictures, and finally the state in which the eyes are closed is erroneously recognized as the state in which the eyes are open.

Table 2. The test result of eye state.

Dataset	Model	Label	Number of test	Number of misidentification	Accuracy (%)
TJPU-FDD	Myopia mirror	Open	4265	44	98.97
		Close	3230	37	98.85
TJPU-FDD	Sunglasses	Open	4328	108	97.50
		Close	3106	95	96.94
TJPU-FDD	Polarizer	Open	4182	82	98.04
		Close	3207	65	97.97

The eye state recognition method based on LBP [4] feature, Multi-feature fusion method [8], the method of combined Gabor and SVM [9], and the method of Sclera Gaussian model [3] are compared with the proposed method in the paper. The accuracy rate of ESR-Net is the average of the test accuracy in Table 2. The comparison results are shown in Table 3.

Table 3. Comparison of different eye state recognition methods.

Method	Accuracy (%)
LBP feature	90.60
Multi-feature fusion	91.90
Gabor+SVM	95.59
Sclera Gaussian model	96.77
ESR-Net(Ours)	98.05

5. Conclusions

In this paper, the driver's face images were collected using the infrared acquisition equipment, and then the images of the eye regions were acquired by the MTCNN method. Finally, the eye state was identified by the CNN-based method. The detection method can effectively utilize the multiform infrared features and the eye state recognition rate is significantly improved when wearing different types of glasses. In addition, this method only recognizes the open and closed state of eyes. However, there is also an intermediate state in the eye during the actual detecting process. In the future, we should study the identification of the middle state of the eye, which is of great significance for judging whether the driver is fatigued in subsequent work.

Acknowledgments

This work is supported by National Nature Science Foundation of China, under grant No. 61771340 and supported by the Program for Innovative Research Team in University of Tianjin No.TD13-5034. And this work is supported by Tianjin Science and Technology Major Projects and Engineering, under grant No. 17ZXHLSY00040, No. 17ZXSCSY00060 and No. 17ZXSCSY00090. This work is also funded by the Tianjin Natural Science Foundation. The funding number is 17JCQNJC01400.

References

- [1] Lee, B.G., Jung, S.J., Chung, W.Y. (2011) Real-time physiological and vision monitoring of vehicle driver for non-intrusive drowsiness detection. *J. IET Communications*, 5: 2461-2469.
- [2] Omidyeganeh, M., Javadtalab, A., Shirmohammadi, S. (2011) Intelligent driver drowsiness detection through fusion of yawning and eye closure. In: *IEEE International Conference on Virtual Environments Human-computer Interfaces & Measurement Systems*. Canada. 18-23.
- [3] Kuang, W.T., Mao, K.C., Huang, J.C. (2016) Fatigue driving detection based on sclera Gaussian model. *J. Journal of Image and Graphics*, 21: 1515-1522.
- [4] Yao, S., Li, X.H., Zhang, W.H. (2015) Eyes state detection method based on LBP. *J. Application Research of Computers*, 32: 1897-1901.
- [5] Zhang, K., Zhang, Z., Li, Z. (2016) Joint Face Detection and Alignment Using Multitask Cascaded Convolution Networks. *J. IEEE Signal Processing Letters*, 23: 1499-1503.
- [6] Wang, J.Z., Li, H., Xiong, Y.Q. (2015) Design and preparation of one induced transmission filter with anlistatig and anti-infrared character. *J. Infrared and Laser Engineering*, 44: 3005-3009.
- [7] Song, Z.H., Liu, Y., Song, R. (2018) A sparsity-based stochastic pooling mechanism for deep convolutional neural networks. *J. Neural Networks*, 105: 340-345.
- [8] Qin, H.B., Li, X.M., Tong, X.M. (2014) Eye state recognition in complex environment based on multi-feature decision fusion. *J. Journal of Optoelectronics Laser*, 24: 2392-2398.
- [9] Javier, F.M. (2010) Driver Drowsiness Warning System Using Visual Information for both Diurnal and Nocturnal Illumination Conditions. *J. EURASIP Journal on Advances in Signal Processing*, 2010: 1-19.