Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

August 2016, 23(4): 91–100
www.sciencedirect.com/science/journal/10058885

The Journal of China Universities of Posts and Telecommunications

http://jcupt.bupt.edu.cn

Fatigue driving detection based on Haar feature and extreme learning machine

Chang Zheng, Ban Xiaojuan(✉), Wang Yu

School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100086, China

Abstract

As the significant branch of intelligent vehicle networking technology, the intelligent fatigue driving detection technology has been introduced into the paper in order to recognize the fatigue state of the vehicle driver and avoid the traffic accident. The disadvantages of the traditional fatigue driving detection method have been pointed out when we study on the traditional eye tracking technology and traditional artificial neural networks. On the basis of the image topological analysis technology, Haar like features and extreme learning machine algorithm, a new detection method of the intelligent fatigue driving has been proposed in the paper. Besides, the detailed algorithm and realization scheme of the intelligent fatigue driving detection have been put forward as well. Finally, by comparing the results of the simulation experiments, the new method has been verified to have a better robustness, efficiency and accuracy in monitoring and tracking the drivers' fatigue driving by using the human eye tracking technology.

Keywords Haar feature, extreme learning machine, fatigue driving detection

1 Introduction

At present, tens of thousands of people have lost their lives, and numerous properties have been damaged each year due to the traffic accident caused by drivers' fatigue driving. According to incomplete statistics, at least one hundred thousand traffic accidents have been caused by the drivers' fatigue driving each year around the world, which has directly led to economic loss up to 12.5 billion dollar. Therefore, the fatigue driving phenomenon is a critical problem in the field of traffic safety. Statistics show that 90% information gathered by the vehicle drivers is obtained by their eyes, thus the measurement of eye opening and closing, movement and position is a very natural method to detect the fatigue state of the vehicle driver. In conclusion, the basic principle is that eye movement trajectory can work as an important indicator to reflect the fatigue state of the vehicle driver. Therefore, it is very effective and natural to recognize the fatigue state of the vehicle driver and avoid the traffic accident caused by the fatigue driving.

Traditionally, aimed at tracking the human eye movement trajectory, the human-computer interaction research mainly focuses on the gray distribution model of the human eyeball, while the tracking of the continuous dynamic eye movement is based on the image gray attribution model or color attribution model. Lu proposed a new face detection method which was based on the human skin color. He segmented human face in the hue- saturation-value (HSV) color space clustering [1]. With the rapid development of technology on the structure light sensor, some researchers applied the structure light sensor to human body. For example, Yang et al. proposed the structure light sensor to detect the position of pupil [2]. And Zhu et al. applied Kinect sensor to detect the real-time state of the human eye [3].

However, due to the diversity, ambiguity and discrepancy in time and space of the human eye movement, the traditional eye movement trajectory tracking technology has faced many limitations. This paper attempts to introduce the image topological analysis, Haar like features and extreme learning machine algorithm into

Received date: 25-11-2015
Corresponding author: Ban Xiaojuan, E-mail: banxj@ustb.edu.cn
DOI: 10.1016/S1005-8885(16)60050-X

the human eye movement trajectory field with the help of the smart glasses, which will be more robust, faster and more accurate in monitoring and tracking the drivers' fatigue driving [4]. In the machine vision area, many researches have proved that the performance of Haar like features model and extreme learning machine model is much better than any other tradition method in real-time application. Dai et al. proposed to use Haar like features model to extract visual saliency in images [5]. Gao proposed Haar like features model and AdaBoost algorithm to detect human movement in his paper, and his new method is verified to have a higher accuracy in identifying the images and video streaming [6]. Moreover, Huang et al. used calculation and experiments to prove that the performance of the extreme learning machine model is much better than other machine learning algorithms in many real-time applications [7].

The paper is organized as follows. In Sect. 1, we describe the background of the fatigue driving detection research and the importance of the fatigue driving detection. In Sect. 2, we describe the main problems about the human eye movement recognition in the fatigue driving detection, followed by a summary of the problems of traditional human eye movement tracking technology. In Sect. 3, we mainly introduce the new method and the main steps including the image topology processing, Haar like features eye tracking, and the fatigue detection based on extreme learning machine. In Sect. 4, a comparison of the fatigue detection performance between the new method and the other algorithms are made. The conclusions of the study are made in Sect. 5.

2 Problem description

Human eye movement trajectory recognition problem is to find the exact position of the human eyeball by analyzing and processing human eyes' real motion images that have been captured by the sensor installed on the smart optical devices. Traditionally, the eye tracking research mainly based on the image pixels, though simple and easy, it is unable to reflect the light and dark part of each image. To this point, the human vision has taken full advantages of the light and dark part of each image in the process of millions years of natural evolution. The human vision is much more sensitive to the great contrast between the light and dark in the adjacent regions than the absolute color value in a certain region. Even under the condition of poor lighting or complex background, people can still identify clearly the position of human body's each part by perceiving the contrast between the light and dark in different area.

Here are four limiting factors of the traditional human eye movement tracking technology:

1) Light

When the light condition changes, the information of bright and grey distribution of human eye area alters with it, thus the images captured by the sensor are susceptible to the light condition.

2) Obstruction

In the recognition process, since the eye movement trajectory may be obscured by the surrounding eyelids, eyelashes and eye corner skin, which may lose some important identification information and weaken the reliability of the tracking.

3) Threshold

In the actual process of recognition, the recognition threshold in the algorithm should be adjusted repeatedly under different conditions instead of keeping the same, which will make the recognition more difficult.

4) Computing time

Based on the image pixels, the computation work will increase with the image resolution by using the traditional eye movement tracking algorithm, thus processing the high resolution image will need a lot of computing time.

The paper combines the image topological analysis technology, Haar like features and extreme learning machine algorithm to track and recognize the position of the eyeball. The new method adopted the image topological analysis technology and Haar like features, has partly overcome the limitations on light, obstruction, threshold and computing time of the traditional methods, and presented a better result.

3 Method

In the paper, the whole process of eye tracking is divided into three parts, as is shown in Fig. 1. First of all, the image topological algorithm is adopted to remove the noise, realize the binarization and keep the important part around the eyeball in the eye video streaming that has been captured by the infrared camera installed on the smart optical devices. Then the Haar like features is adopted to recognize the binarized eye region image preprocessed in the first step. Next the Haar wavelet transformation and the cascaded enhanced classifier by AdaBoost algorithm are used to obtain the position and closing and opening of the eyeball [8–9]. The recognition speed is very fast. At last, the extreme learning machine algorithm is applied to analyze the fatigue degree of the vehicle driver.

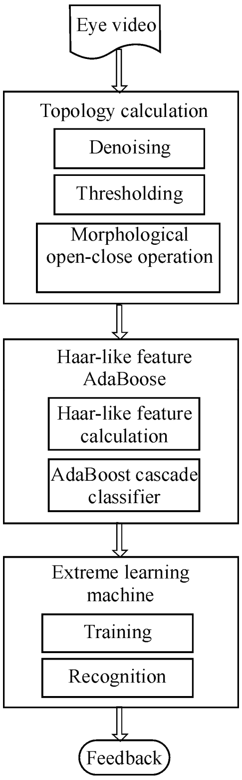


Fig. 1 System structure

3.1 Image topology processing

To locate the pupil position in the following section and improve the recognition performance of fatigue driving, the paper attempts to use image topology method in three steps to preprocess every frame of raw eye video streaming captured by the smart glasses, as is shown in Fig. 2. First of all, the system gray process and histogram equalize raw eye images which are captured by infrared camera of smart optical devices. And then the system needs to threshold every frame of eye video streaming. At last, the paper adopts the open-close topology method to obtain the main part of the pupil image.

The brightness of the eye images is likely to be captured by infrared camera of smart optical device, and the gradient reflections of these images are very strong. The pixel distributions of these eye images are not homogeneous. Therefore, it is difficult to obtain the perfect threshold process performance of these images within one or two fixed threshold values. Thus our research needs to adopt an adaptive threshold method which can adjust the threshold value according to the gradient of the images, that is why we introduce the Gaussian adaptive threshold value method into the paper.

For every pixel $I(x, y)$ of the eye image, the system calculates the weighted sum of other pixels in n region adjacency of $I(x, y)$, and then the system uses weighted sum of n region adjacency minus a constant P to obtain the Gaussian adaptive threshold value of the pixel $I(x, y)$. As is shown in the following formula:

$$I'(x, y) = \begin{cases} 0; & I(x, y) < T(x, y) \\ 255; & I(x, y) \geqslant T(x, y) \end{cases} \tag{1}$$

$$T(x, y) = \sum_{i=1}^{n} (I_i W_i) - P \tag{2}$$

$$W_i = W(x, y) = \mathrm{e}^{-x^2 - y^2} \tag{3}$$

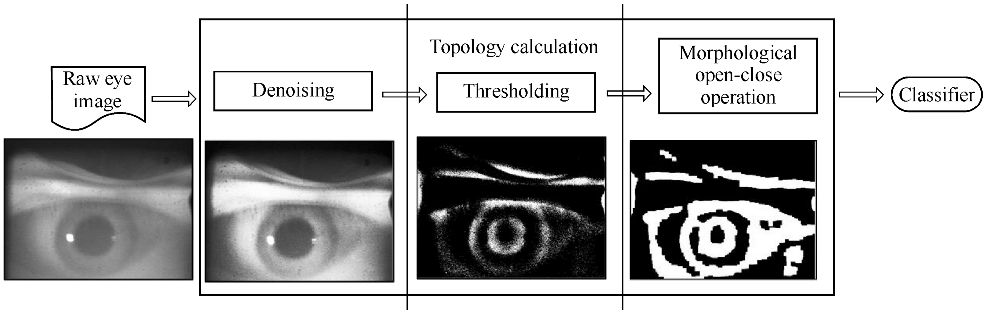


Fig. 2 Eye image preprocessing

In Eqs. (1)–(3) $I'(x, y)$ is the pixel gray value of the pixel $I(x, y)$ after the Gaussian adaptive threshold process, and $T(x, y)$ is the Gaussian adaptive threshold value in the pixel (x, y). I_i is the other pixels in n region adjacency of pixel $I(x, y)$. W_i is the weight of every pixel by the standard Gaussian formula.

The system applies Gaussian adaptive threshold method to process eye images which have been processed by gray-level histogram equalization method, and then the system can obtain a source image which includes a clear principal component. For those images which have a higher brightness and gradient reflection, the threshold performance of this adaptive threshold method is much better than global fixed threshold value method, as is shown in Figs. 3 and 4.

Fig. 3 Global fixed threshold method

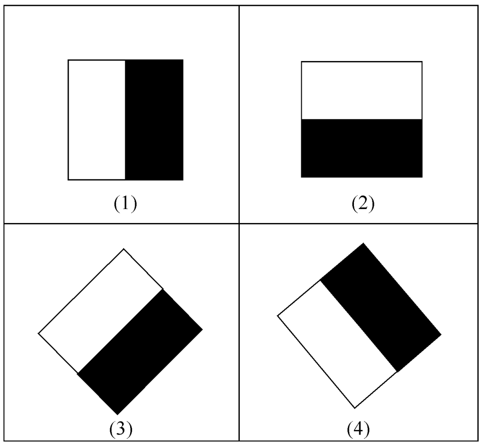
Fig. 4 Gaussian adaptive threshold method

3.2 Eye tracking

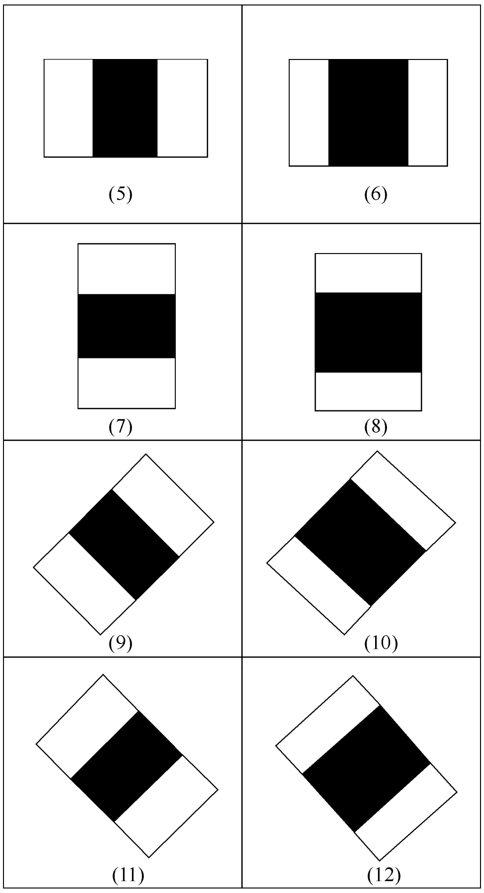
In 1988, Kearns and Valiant proposed prototype of the Boosting algorithm, and this new algorithm is based on the probably approximately correct (PAC) model. In 1995, Freund and Schapire proposed an improved Boosting algorithm, the adaptive boosting (AdaBoost) algorithm, which was based on the traditional boosting algorithm [10]. In 2001, Viola and Jones designed a kind of new human face detection algorithm which was based on AdaBoost algorithm and cascade weak classifiers method, and later Rainer and Jochen extended the face detection model [11–12].

This paper adopts the method of Haar wavelet transform and AdaBoost cascade model to detect and track human pupil. The method to extract the characteristics of the eye by using the Haar wavelet transform is similar to the biological visual channel that is more sensitive to the great contrast between the light and dark in the adjacent regions than the absolute color value in a certain region.

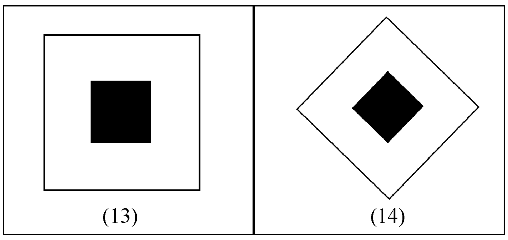
There are 14 Haar features in Fig. 5 [4,8]. In Fig. 5, features (1)~(4) are used to detect edge feature. Features (5)~(12) are used to detect line feature. Features (13)~(14) are used to detect center-surround feature.



(a) Features are used to detect edge feature



(b) Features are used to detect line feature



(c) Features are used to detect center-surround feature

Fig. 5 Haar features

In Haar features extraction section, the paper assumes that the size of filter window is $W \times H$. System calculate weighted sum of rectangle area which is covered by some Haar features. A rectangle area in filter window is specified by the tuple $r = (x, y, w, h, \alpha)$. In the tuple, x, y represents top left corner coordinate of this rectangle area. $w, h, (0 \leqslant w \leqslant W, 0 \leqslant h \leqslant H)$ is the width and length of the rectangle area. α $(\alpha = \{0^\circ, 45^\circ\})$ represents the rotation angle of the rectangle area.

Therefore, the pixel sum of the rectangle area is denoted by $S_{\text{rectangle}}(r)$.

The calculation formula of Haar feature is as follow:

$$F_I = \sum_{i=1}^{I} (\omega_i S_{\text{rectangle}}(r_i)) \tag{4}$$

Among them, ω_i means the weight of rectangle area. In this paper, the weight of the white rectangle area of Haar feature is +1, and the black one is -1.

This paper proposes to apply integral image method to compute the pixel sum of rectangle area. The integral image method is a kind of matrix computing method which can describe the whole information of image. The construction of integral image is that the value of pixel (x, y) is defined as the sum of the pixels of the upright rectangle R ranging from the top left corner at original point (0, 0) to the bottom right corner at (x, y) :

$$S_{\text{pixel}}(x, y) = \sum_{x, y \subseteq R} (I(x, y)) \tag{5}$$

For those rectangle area which is rotated in $45°$, system can rotate the integral image in $45°$. Therefore, the pixel sum of rectangle R is defined as follow:

$$S_{\text{rectangle}}(D) = S_{\text{pixel}}(d) + S_{\text{pixel}}(a) - S_{\text{pixel}}(b) - S_{\text{pixel}}(c) \tag{6}$$

Among them, the four corners of the rectangle area D are a, b, c, d , and the values of the four corners in the integral image are $S_{\text{pixel}}(a), S_{\text{pixel}}(b), S_{\text{pixel}}(c), S_{\text{pixel}}(d)$, as is shown in Fig. 6.

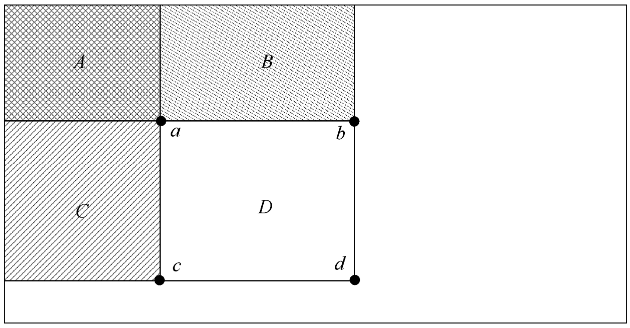


Fig. 6 Pixels sum calculation

By searching each frame of the eye video streaming in the filter windows, system can obtain a lot of features. In the paper, system cascades many weak classifiers to strong classifiers, and this cascade model can help us to find out the key features efficiently among lots of Haar features.

As shown in Fig. 7, system can cascade many weak classifiers. Only if the recognition result of the first level weak classifier is true, the second level weak classifier should work. The third level weak classifier should work, only if the second level classifier has a positive result. However, if the result of some level classifier is negative, the image will not pass the detection process. In this way, many cascade classifiers build up a decision tree. Some primary weak classifiers can negative a lot of unqualified images efficiently, thus the amount of images that need to be tested by cascade classifiers will decrease greatly.

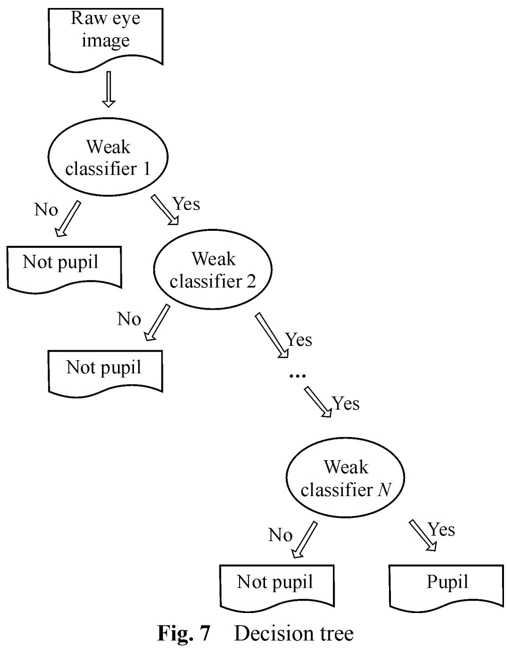


Fig. 7 Decision tree

By above process of many weak classifiers with cascade Haar features and AdaBoost algorithm, the system can obtain the information of human eye, such as the open-close states, movement direction, and real-time position, as is shown in Fig. 8.

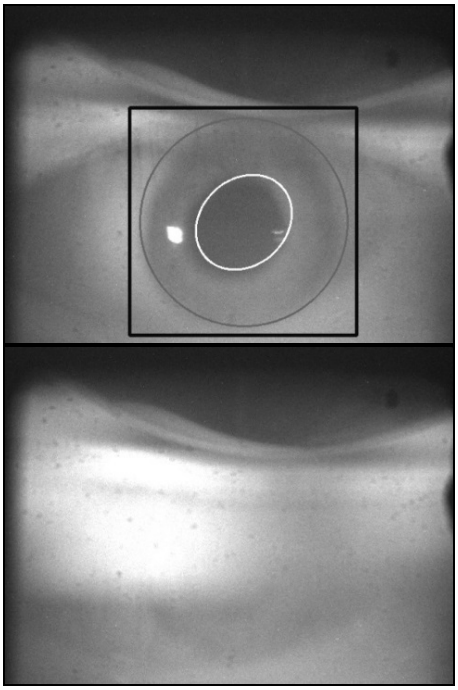


Fig. 8 Pupil detection

3.3 Fatigue driving detection

This paper also proposes the extreme learning machine model to detect drivers' fatigue driving. The extreme learning machine includes two steps: network training and network recognition. The training step is that system put positive and negative features set into the computing model, and the parameters of extreme learning machine artificial network would be determined. The recognition step is that system can analyze the real-time eye video streaming captured by smart glasses devices after system determines the parameters of extreme learning machine artificial network. In the last section, system can obtain some information of human eye, such as the open-close states, movement direction, and real-time position. Therefore, the real-time information of pupil in a frame is specified by the tuple $F_{\mathrm{EyeState}} = (x, y, m_{\mathrm{Dir}}, m_{\mathrm{Dis}}, o, t)$. In the tuple, x, y represents the center coordinate of the pupil. m_{Dir} represents the relative direction of pupil movement with respect to the last frame. m_{Dis} represents the displacement of pupil with respect to the last frame. o means the open-close states of pupil. t represents the time interval of close eye.

Therefore, this paper proposes a machine learning artificial neural network to process the tuple $F_{\mathrm{EyeState}} = (x, y, m_{\mathrm{Dir}}, m_{\mathrm{Dis}}, o, t)$ and detect the fatigue driving actions. According to the tuple $F_{\mathrm{EyeState}} = (x, y, m_{\mathrm{Dir}}, m_{\mathrm{Dis}}, o, t)$, the artificial neural network can detect the fatigue driving conditions. However, our real-time fatigue driving detection application needs a higher requirement of robustness, efficiency and accuracy. The traditional machine learning artificial neural network cannot achieve the high requirement of robustness, efficiency and accuracy. These traditional methods still have some limitations:

1) Average eigenvalue method

In this method, system needs to compute the average eigenvalue of the pre-defined sample human eye movement trajectory. So the average eigenvalue method is influenced greatly by some special pre-defined sample human eye movement trajectory such as the wide margin action or the narrow range action.

2) *K* nearest neighbor method

The *K* nearest neighbor method (KNN) needs to compute the distance (such as Euclidean distance, Mahalanobis distance or pearson correlation) between the real human eye movement trajectory captured by the sensor and the pre-defined sample human eye movement trajectory. Therefore, due to the heavy computation, the method has been limited to very few application fields.

3) Gradient descent-based feed-forward network learning methods

The back propagation (BP) artificial network is one of the typical gradient descent-based feed-forward network learning methods. All the parameters of the feed forward networks need to be turned and thus there exists the dependency between different layers of parameters (weight and biases) in the BP artificial network. Therefore, the training process is very slow. Moreover, the gradient descent-based learning method may easily cause the local minima and over-fitting problem.

Therefore, in the paper, our team decides to use the extreme learning machine network model to detect the fatigue driving condition.

The paper attempts to build up an extreme learning machine (ELM) artificial neural network (Fig. 9) as follow [13]:

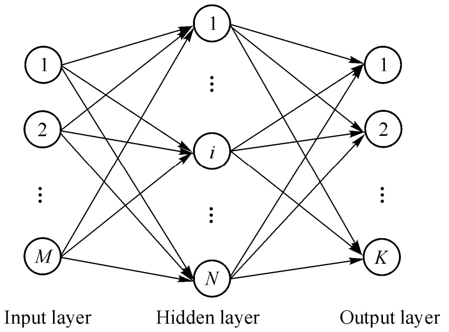


Fig. 9 ELM artificial neural network

There are N hidden nodes and M input lay nodes in ELM artificial neural network. The training sample set is $(\boldsymbol{X}_j, \boldsymbol{T}_j)$, where $\boldsymbol{X}_j = \left[\boldsymbol{X}_{j1}, \boldsymbol{X}_{j2}, ..., \boldsymbol{X}_{jM}\right]$ is the feature of eye state samples, $\boldsymbol{T}_j = \left[\boldsymbol{T}_{j1}, \boldsymbol{T}_{j2}, ..., \boldsymbol{T}_{jM}\right]^{\mathrm{T}}$ is the results of recognition, including sleeping, wake, and tired. Therefore, there are 3 output layer nodes, which means that there are 3 types of drivers' states. Each node in the output layer presents a type of drivers' state. And activation function $f(x)$ is mathematically modeled as:

$$\sum_{i=1}^{N} \beta_i f(\boldsymbol{W}_i \boldsymbol{X}_i + b_i) = \boldsymbol{R}_j \tag{7}$$

where $\boldsymbol{R}_j = \left[\boldsymbol{R}_{j1}, \boldsymbol{R}_{j2}, ..., \boldsymbol{R}_{jM}\right]^{\mathrm{T}}$ is the output vector from ELM artificial neural network, $\boldsymbol{W}_i = [W_1, W_2, ..., W_N]^{\mathrm{T}}$ is the weight vector which connects the ith hidden layer node, and the input nodes, $\boldsymbol{\beta}_i = [\beta_1, \beta_2, ..., \beta_N]^{\mathrm{T}}$ is the weight vector which connect the ith hidden layer node and the output nodes, b_i is the threshold of the ith hidden layer node.

Each element in the two vectors $\boldsymbol{T}_j = \left[\boldsymbol{T}_{j1}, \boldsymbol{T}_{j2}, ..., \boldsymbol{T}_{jM}\right]^{\mathrm{T}}$ and $\boldsymbol{R}_j = \left[\boldsymbol{R}_{j1}, \boldsymbol{R}_{j2}, ..., \boldsymbol{R}_{jM}\right]^{\mathrm{T}}$ is a K dimensional vector. In other words, $\boldsymbol{T}_{ji}$ and $\boldsymbol{R}_{ji}$ serves as the real type and computational type of result of recognition of training sample $\boldsymbol{X}_{ji}$, respectively.

Our goal is to make $\boldsymbol{T}_j$ close to $\boldsymbol{R}_j$, which is equivalent to minimizing the cost function

$$C = \sum_{i=1}^{N} \beta_i f(\boldsymbol{W}_i \boldsymbol{X}_i + b_i) - \boldsymbol{T}_j = \boldsymbol{R}_j - \boldsymbol{T}_j \tag{8}$$

Previous excellent researches have proved that for any infinitely differentiable activation function ELM artificial neural network with N hidden nodes can distinct N samples exactly, and ELM artificial neural network may require less than N hidden nodes if learning error is allowed [7,14]. Of course, the numbers of the hidden layer nodes in the artificial neural network could have some impacts on the drivers' state recognition performance. In the experiment section, this paper will show these impacts under different numbers of the hidden layer nodes.

The equation is presented as follow:

$$\sum_{i=1}^{N} \beta_i f(\boldsymbol{W}_i \boldsymbol{X}_j + b_i) = \boldsymbol{T}_j \tag{9}$$

The above equations could be abbreviated as

$$\boldsymbol{BF} = \boldsymbol{T} \tag{10}$$

where

$$F = \begin{bmatrix} f(\boldsymbol{W}_1 \boldsymbol{X}_1 + b_1) & \cdots & f(\boldsymbol{W}_1 \boldsymbol{X}_M + b_1) \\ \vdots & & \vdots \\ f(\boldsymbol{W}_N \boldsymbol{X}_1 + b_N) & \cdots & f(\boldsymbol{W}_N \boldsymbol{X}_M + b_N) \end{bmatrix}_{N \times M}$$

$$\boldsymbol{B} = [\beta_1, \beta_2, ..., \beta_N]$$

$$\boldsymbol{T} = [T_1, T_2, ..., T_M]$$

On both sides of the equation, the transpose operation could be applied in order to obtain the standard ELM model:

$$\boldsymbol{H\beta} = \boldsymbol{S} \tag{11}$$

where

$$\boldsymbol{H} = \boldsymbol{F}^{\mathrm{T}}$$

$$\boldsymbol{\beta} = \boldsymbol{B}^{\mathrm{T}}$$

$$\boldsymbol{S} = \boldsymbol{T}^{\mathrm{T}}$$

So we could obtain the following result:

$$\boldsymbol{\beta} = \boldsymbol{H}^{\dagger} \boldsymbol{S} \tag{12}$$

where $\boldsymbol{H}^{\dagger}$ is the Moore-Penrose generalized inverse of matrix $\boldsymbol{H}$.

If $N = M$ and $\boldsymbol{H}$ has an inverse, the solution is existing and unique. And the answer is, evidently, $\boldsymbol{H}^{\dagger} = \boldsymbol{H}^{-1}$. The result is shown as follow:

$$\boldsymbol{\beta} = \boldsymbol{H}^{-1} \boldsymbol{S} \tag{13}$$

There are several methods which could be used to calculate the Moore-Penrose generalized inverse of matrix $\boldsymbol{H}$. This paper attempts to apply the spectral theorem and Tikhonov's regularization method to calculate $\boldsymbol{H}^{\dagger}$ Moore-Penrose generalized inverse of matrix $\boldsymbol{H}$ [15].

$$\boldsymbol{H}^{\dagger} = \sum_{\substack{b=1 \\ \alpha_b \neq 0}}^{S} \frac{1}{\alpha_b} \left[\prod_{\substack{l=1 \\ l \neq b}}^{S} (\alpha_b - \alpha_l)^{-1} \right] \left[\prod_{\substack{l=1 \\ l \neq b}}^{S} \left(\boldsymbol{H}^* \boldsymbol{H} - \alpha_l \boldsymbol{E}_n \right) \right] \boldsymbol{H}^* \tag{14}$$

where $\boldsymbol{H}^*$ represents the adjoint matrix of $\boldsymbol{H}$ and $\alpha_b, b = 1, 2, ..., s$ are the eigenvalues of $\boldsymbol{H}^* \boldsymbol{H}$ (or the singular values of $\boldsymbol{H}$).

According to the training sample set, could obtain the weight vector $\boldsymbol{\beta}$ and then complete the training of ELM. As presented in the previous outstanding paper, ELM has many important properties: the smallest norm of weights, the minimum approximation error, and the minimum norm least-squares solution of $\boldsymbol{H\beta}=\boldsymbol{S}$ is unique.

After the system gets the parameters of artificial network, the training step is completed. And then, system can detect fatigue driving state by the real-time eye images which are captured by smart optical devices. If the system has detected the driver's fatigue driving, it will give drivers some alert.

4 Results of the experiment

4.1 Experiment data pre-processing

The human eye movement trajectory recognition experiments are done in normal laboratory environment. In the experiment, people should keep their head forward, perpendicular to the horizontal plane, wear a smart glasses device, and keep about 1 cm to 1.5 cm to the infrared camera installed on the smart glasses devices.

In the paper, the eye movement cannot be too fast or too slow. The human eye movements monitored are debounced and the center position data of the prior frame are recorded to compare with the center position data of the current frame. If the deviation is within the threshold range, the position data of the prior frame will be chosen regardless of the jitter of the current frame. During the comparative experiments, the experiments are repeated for many times to determine the value of the threshold.

When using the real-time human eye movement trajectory, invalid frames will appear at the beginning and the end of the movement. By eliminating the jitter of the physical movements, the system could remove the useless part of the physical movements, and all the frames left are presented clearly.

4.2 The result of the experiment

In order to verify the robustness of the method, here some comparative experiments are carried out under different conditions, such as different light environment and different obstruction. We could compare with the recognition rate of different traditional recognition algorithms (KNN, BP and SVM) in the same raw data and experimental environments.

In these comparing experiments, our team puts the same Haar like features process result into different machine learning algorithm. In other words, we use different machine learning algorithm to study the same eye features in order to compare their different performance in the emulational fatigue driving detection application.

We repeat the comparative experiments for 50 times in different light environments. Table 1 and Figs. 10 and 11 are the recognition accuracy in normal light and the weak light environments for human eye movements. The recognition accuracy rate declines sharply by the traditional method in weak light conditions. The Table 1 shows that the method in this paper could capture the human pupil movement very well under the normal and the weak light environments. The experiment shows that the new method has a better robustness under the weak light conditions than the traditional method, as is shown in Figs. 10 and 11.

Table 1 Recognition rate under different light

Light environment	Recognition rate/(%)			
	KNN	BP	SVM	Haar+ELM
Ordinary light	85	91	95	96
Weak light	67	75	82	91

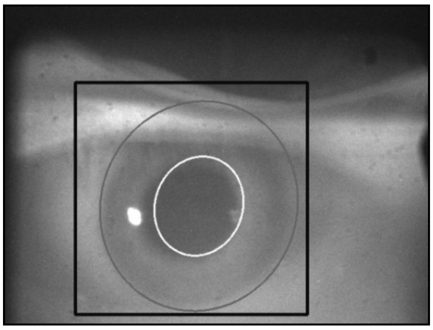


Fig. 10 Pupil image under normal light environment

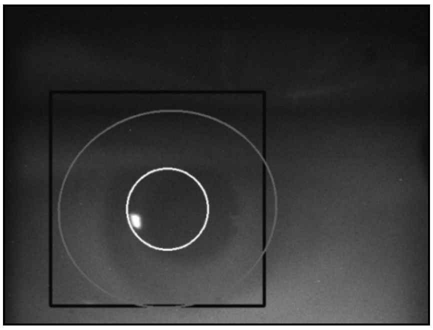


Fig. 11 Pupil image under weak light environment

We repeat the comparative experiments for 50 times in different obstruction environment. Table 2 and Fig. 12 are the recognition accuracy in normal obstruction and heavy obstruction of the human eye movements. The recognition accuracy rate declines sharply by the traditional method in heavy obstruction conditions. The Table 2 shows that the new method could capture the human pupil movements very well under both obstruction environments. The experiment shows that the new method has a better accuracy under the heavy obstruction conditions than the traditional method, as is shown in Figs. 12 and 13.

Table 2 Recognition rate under different obstruction

Obstruction Environment	Recognition rate/(%)			
	KNN	BP	SVM	Haar+ELM
Normal obstruction	86	92	95	97
Heavy obstruction	63	69	86	92

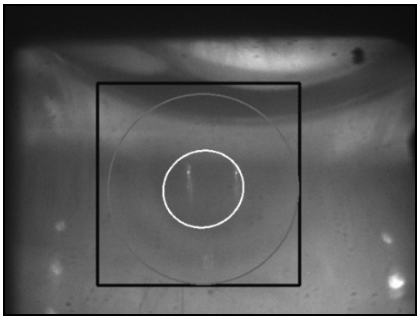


Fig. 12 Pupil image under heavy obstruction

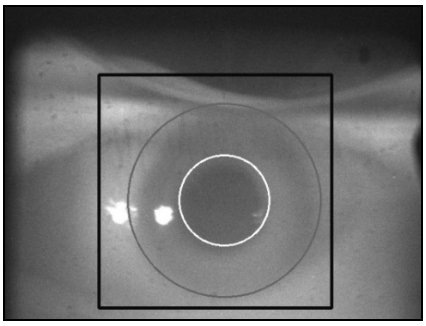


Fig. 13 Pupil image under normal obstruction

We repeat the comparative experiments for 20 times in different algorithms. Table 3 records the recognition time in different algorithms.

Table 3 Recognition time under different algorithm

Experiment number	Recogniton time/ (frame·s^{-1})			
	KNN	BP	SVM	Haar+ELM
1	19	28	30	33
2	18	27	31	35
3	19	26	29	32
4	21	29	26	31
5	17	25	28	33
6	18	28	25	34
7	15	30	29	33
8	19	25	26	32
9	18	29	29	33
10	19	27	28	34
11	20	28	29	31
12	19	30	29	35
13	17	32	30	34
14	18	26	31	33
15	19	29	26	30
16	20	28	28	31
17	21	29	26	32
18	17	27	27	33
19	19	29	29	30
20	18	25	30	31

The experiment shows that the new method has a better efficiency than the traditional method. In order to verify the performance of the method in the real application, here some comparative experiments are carried out under emulational fatigue driving detection environment. We could compare the recognition rate of different traditional recognition algorithms (KNN, BP and SVM) with the same raw data and experimental conditions.

We repeat the comparative experiments for 50 times under different algorithms condition in emulational fatigue driving detection environment and Table 4 records the recognition accuracy in different machine learning algorithms in the emulational fatigue driving detection environment. The experiment shows that the new method has a better efficiency than the traditional method.

Table 4 Recognition performance under different algorithms

Machine learning algorithm	Recognition performance		
	Training time/s	Recognition time/s	Recognition rate/(%)
KNN	8.390	0.148	81.9
BP	4.250	0.093	86.9
SVM	2.630	0.056	93.5
Haar+ELM	0.094	0.049	93.9

5 Conclusions

In the paper, by combining the image topological analysis technology, Haar features and extreme learning machine algorithm, the new method proposed has partly overcome the limitations on light, obstruction, threshold and computing time of the traditional methods, and presents a better result. By comparing the results of the simulation experiments, the new method has been verified to have a better robustness, efficiency and accuracy in monitoring and tracking the drivers' fatigue driving by using the human eye tracking technology.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (61272357, 61300074, 61572075).

References

1. Lu K. A study on OpenCV-based algorithms for face detection and tracking. Master Thesis. Shanghai, China: East China University of Science and Technology, 2015 (in Chinese)
2. Yang W L, Zheng J Y, Zhan C, et al. Pupil localization algorithm based on Kinect sensor. Transducer and Microsystem Technologies, 2015, 34(2): 125−128 (in Chinese)
3. Zhu Z Z, Wang W, Duan X D, et al. Real-time detection of facial eye status

based on Kinect sensor. Journal of Dalian Nationalities University, 2015, 17(1): 81−84 (in Chinese)
4. Viola P, Jones M J. Robust real-time object detection. CRL 2001/01. Cambridge, MA, USA: Cambridge Research Laboratory, 2001: 25p
5. Dai B, Wu X F, Wang B. A visual saliency model based on Haar like featuress and structural complexity. Journal of Fudan University (Natural Science), 2014, 53(5): 651−658 (in Chinese)
6. Gao Y. Human movement detection based on Haar features. Master Thesis. Changchun, China: Jilin University, 2013 (in Chinese)
7. Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: a new learning scheme of feedforward neural networks. Proceedings of the 2004 IEEE International Joint Conference on Neural Networks (IJCNN'04): Vol 2, Jul 25−29, 2004, Budapest, Hungary. Piscataway, NJ, USA: IEEE, 2004: 985−990
8. Lienhart R, Maydt J. An extended set of Haar-like feature for rapid object detection. Proceedings of the 2002 International Conference on Image Processing (ICIP'02): Vol 1, Sept 10−13, 2002, Vancouver, Canada. Piscataway, NJ, USA: IEEE, 2002: 900−903
9. Lienhart R，Kuranov A，Pisarevsky V. Empirical analysis of detection cascades of boosted classifiers for rapid object detection. Pattern Recognition: Proceedings of the 25th DAGM Symposium. Sept 10−12, 2003, Magdeburg, Germany. LNCS 2781. Berlin, Germany: Springer, 2003: 297−304
10. Huang G F, Cheng X P, Ren F. An automatic recognition approach of human gestures. Journal of Southwest China Normal University, 2010, 35(4): 136−140 (in Chinese)
11. Guo K, Ishwar P, Konrad J. Action recognition using sparse representation of covariance manifold of optical flow. Proceedings of the 7th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS'10), Aug 29-Sept 1, 2010, Boston, MA, USA. Piscataway, NJ, USA: IEEE, 2010: 188−195
12. Gu Y, Liu J F, Chen Y Q, et al. Constraint online sequential extreme learning machine for lifelong indoor localization system. Proceedings of the 2014 International Joint Conference on Neural Networks (IJCNN'14), Jul 6-11, 2014, Beijing, China. Piscataway, NJ, USA: IEEE, 2014: 732−738
13. Chang Z, Ban X J, Shen Q, et al. Research on three-dimensional motion history image model and extreme learning machine for human body movement trajectory recognition. Mathematical Problems in Engineering, 2015: ID 528190/ 1−15
14. Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: theory and applications. Neurocomputing, 2006, 70(1/2/3): 489−501
15. Chang Z, Ban X J, Liu X. Recognition of human head movement trajectory based on the three-dimensional somatosensory technology. Knowledge Engineering and Management: Proceedings of the 7th International Conference on Intelligent Systems and Knowledge Engineering (ISK'12), Dec 15−17, 2012, Beijing, China. Advances in Intelligent Systems and Computing 214. Berlin, Germany: Springer, 2014: 611−622

(Editor: Wang Xuying)

From p. 82

17. Mittal A, Soundararajan R, Bovik A C. Making a “completely blind” image quality analyzer. IEEE Signal Processing Letters, 2013, 20(3): 209−212
18. Reichl P, Egger S, Schatz R, et al. The logarithmic nature of QoE and the role of the Weber-Fechner law in QoE assessment. Proceedings of the 2010 IEEE International Conference on Communications (ICC'10), May 23−27, Cape Town, South Africa. Piscataway, NJ, USA: IEEE, 2010: 5p
19. Vu P V, Vu C T, Chandler D M. A spatiotemporal most-apparent-distortion model for video quality assessment. Proceedings of the 18th IEEE International Conference on Image Processing, Sept 11−14, 2011, Brussels, Belgium. Piscataway, NJ, USA: IEEE, 2011: 2505−2508
20. Pinson M H, Wolf S. A new standardized method for objectively measuring video quality. IEEE Transactions on Broadcasting, 2004, 50(3): 312−322
21. Soundararajan R, Bovik A C. Video quality assessment by reduced reference spatio-temporal entropic differencing. IEEE Transactions on Circuits and System for Video Technology, 2013, 23(4): 684−694
22. Sheikh H R, Sabir M F, Bovik A C. A statistical evaluation of recent full reference image quality assessment algorithms. IEEE Transactions on Image Processing, 2006, 15(11): 3440−3451

(Editor: Wang Xuying)