Eye and mouth state detection algorithm based on contour feature extraction

Yingyu Ji
Shigang Wang
Yang Lu
Jian Wei
Yan Zhao

Eye and mouth state detection algorithm based on contour feature extraction

Yingyu Ji, Shigang Wang,* Yang Lu, Jian Wei, and Yan Zhao
Jilin University, School of Communication Engineering, Changchun, China

Abstract. Eye and mouth state analysis is an important step in fatigue detection. An algorithm that analyzes the state of the eye and mouth by extracting contour features is proposed. First, the face area is detected in the acquired image database. Then, the eyes are located by an EyeMap algorithm through a clustering method to extract the sclera-fitting eye contour and calculate the contour aspect ratio. In addition, an effective algorithm is proposed to solve the problem of contour fitting when the human eye is affected by strabismus. Meanwhile, the value of chromatism s is defined in the RGB space, and the mouth is accurately located through lip segmentation. Based on the color difference of the lip, skin, and internal mouth, the internal mouth contour can be fitted to analyze the opening state of mouth; at the same time, another unique and effective yawning judgment mechanism is considered to determine whether the driver is tired. This paper is based on the three different databases to evaluate the performance of the proposed algorithm, and it does not need training with high calculation efficiency. *© The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI.* [DOI: 10.1117/1.JEI.27.5.051205]

1 Introduction

In recent years, driver fatigue has become one of the most important factors for traffic accidents, which has come at a great cost to the safety and property of drivers and pedestrians. Researchers have proposed many fatigue detection methods to solve this problem, which can be divided into three types: [1] physiological parameters, vehicle behaviors, and facial feature analysis. The first method measures the driver's physiological parameters[2–6] by using tools such as electroencephalogram and electrocardiogram. However, these methods are invasive and require contact with the driver's body. The second method is used to measure the behaviors of vehicles,[7–11] such as speed, steering wheel rotation angle, and lane departure detection; however, this method is affected by driving conditions, driving experience, and vehicle type. The third method analyzes the driver's face, such as the PERCLOS value, blink frequency, head posture, and yawn detection. PERCLOS is the abbreviation of percentage of eyelid closure over the pupil over time, which is the percentage of the closing time of the eye over a specific period of time. This method is noninvasive and easy to implement and is applied to the fatigue detection in this paper. The following is a brief introduction to some of the algorithms for detecting fatigue by facial feature analysis.

You et al. presented a night monitoring system for real-time fatigue driving detection,[12] but the monitoring system requires additional infrared lighting equipment, which is limited to some specific applications. An eye state detection method based on projection is proposed in Ref. 13, the height and width of the iris are estimated by integral or variance projection, and the eye state is determined according to the aspect ratio. Omidyeganeh et al. calculated the mouth width and height by horizontal and vertical grayscale projection of

the mouth area[14] according to its aspect ratio to determine whether a driver is yawning. However, this method is not effective when the teeth are exposed or the driver has a beard. The authors[15] proposed a fatigue detection system, based on a smartphone, used to detect the eyes according to a progressive locating method (PLM) algorithm. When at least three frames of the adjacent five frames are closed, the driver is considered to be in a state of fatigue. However, the limitation of this PLM algorithm is that it over-relies on the gray distribution of the facial region. Further, there are some neural networks or supporting vector machine methods to identify the state of the eyes and mouth via training classifiers,[16–20] which have a high detection accuracy. However, these methods need to collect a large amount of training data and have a long training time.

The rest of this paper is organized as follows. Section 2 introduces face region detection, and the location of the eyes and mouth on this basis. Section 3 explains the contour feature extraction method, including eye contour extraction and mouth internal contour extraction. Section 4 evaluates the performance of the algorithm based on the experimental results and analyzes its state. Finally, the conclusion and future work are presented in Sec. 5.

2 Proposed Eye and Mouth Region Detection Algorithm

The block diagram of the proposed eye and mouth state detection algorithm is shown in Fig. 1. The image is acquired from three different databases. Subsequently, we introduce each part of the algorithm.

2.1 Face Detection

The images in the database are obtained under different lighting conditions. Because the highlights and shadows caused by the light source have a great influence on the skin color,

---

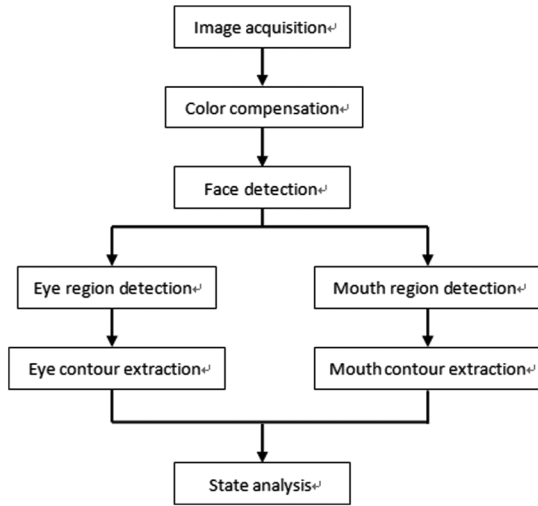*Address all correspondence to: Shigang Wang, E-mail: wangsg@jlu.edu.cn

Fig. 1 Block diagram of the proposed eye and mouth state analysis algorithm.

we use the algorithm of Ref. 21 for color compensation. The facial area is extracted from the database images to obtain the image of the eye and mouth region. Thus, we use the Viola Jones face detection algorithm.[22] The Viola Jones face detection algorithm is a method based on an integral graph, cascade classifier, and Adaboost algorithm, which greatly improves the speed and accuracy of face detection.

2.2 Eye Detection

There is a fixed connection among facial features. For example, the eyes are set in the upper part of the face and the mouth is located in the lower part of the face. In order to improve the accuracy and speed of detection, our algorithm determines the region of interest (ROI) of the eyes and mouth, and then detects the target on the ROI region. After obtaining the facial image, the upper half of the image is extracted and recorded as image I_1, the upper one-eighth of image I_1 is removed, and the lower seven-eighths of image I_1 is reserved and set as the eye ROI, as shown in Fig. 2(a). In this ROI, we use the EyeMap algorithm[23] to locate the eye region. This method builds two EyeMaps

in the YCbCr space,[24] EyeMapC and EyeMapL; then, these two maps are combined into a single map. Experiments find high Cb components and low Cr components around the eyes, and EyeMapC is calculated as follows:

$$\text{EyeMapC} = \left[C_b^2 + (255 - Cr)^2 + \frac{C_b}{C_r} \right] / 3. \qquad (1)$$

The values of C_b^2, $(255 - Cr)^2$, and $\frac{C_b}{C_r}$ are normalized to the range [0, 255]. In addition, eyes contain bright and dark values in the luminance component; therefore, grayscale dilation and erosion with ball structuring elements are used to construct EyeMapL.[25] EyeMapL is calculated as follows:

$$\text{EyeMapL} = \frac{Y(x, y) \bigoplus g(x, y)}{Y(x, y) \Theta g(x, y)}, \qquad (2)$$

where $g(x, y)$ represents the ball structuring element and $\bigoplus$ and Θ denote the grayscale dilation and erosion operations.

Then, EyeMapC is multiplied by EyeMapL to obtain EyeMap

$$\text{EyeMap} = \text{EyeMapC} \times \text{EyeMapL}. \qquad (3)$$

EyeMap of a typical image (from the California Polytechnic University color face database) is constructed as shown in Fig. 2. Among them, the original eyes' ROI is as shown in Fig. 2(a) and EyeMapC, EyeMapL, and EyeMap are as shown in Figs. 2(b)–2(d), respectively.

In order to accurately locate the eye region, the optimal threshold T is obtained by leveraging the OTSU algorithm to convert the EyeMap gray image into a binary image, as shown in Fig. 3(a). We analyze the aspect ratio, position, and other characteristics of every connected component (white part), to exclude the noneye region, and finally consider a pair of connected components as the eye region, as shown in Fig. 3(b). If there is no pair of connected domains, then the threshold is reduced based on the optimal threshold value T and redetected. Experiments demonstrate that eye length is approximately half of the distance between the center of eyes, and eye height is approximately half of the eye length.
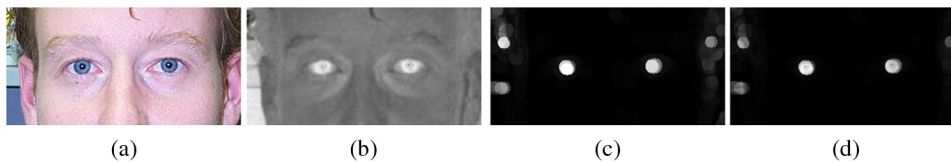

Fig. 2 Illustration of EyeMap construction: (a) the original eye ROI region. (b) EyeMapC. (c) EyeMapL. (d) EyeMap.
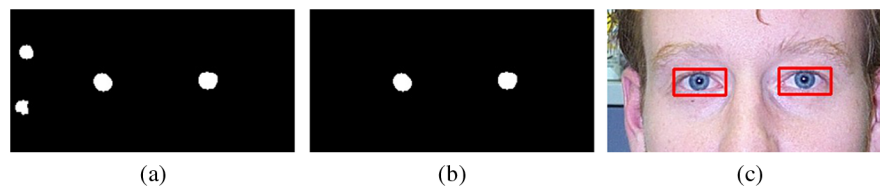

Fig. 3 Different steps of our eye detection method: (a) EyeMap image after binarization. (b) A pair of connected components. (c) Calibrate eye region with a rectangular box.
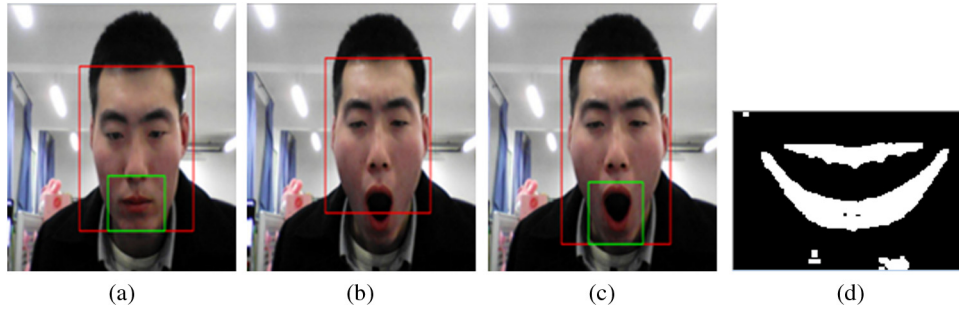
Fig. 4 Different steps of our mouth detection method: (a) Initial mouth ROI. (b) Mouth opens widely. (c) Final mouth ROI. (d) Selected 20% pixels.

Therefore, we locate the region of the left and right eyes, with a rectangular box calibration, as shown in Fig. 3(c).

2.3 Mouth Detection

To improve the speed and accuracy of mouth detection, we set the ROI based on the characteristics of the mouth distribution in the face region. Saeed and Dugelay[26] proposed that the mouth ROI was the lowest one-third of the detected face region. The lower one-third of the face image is extracted and recorded as the image I_2, and the middle half of the image I_2 is extracted and set as the mouth ROI, as shown in the green box in Fig. 4(a). However, when the mouth opens widely (yawning), we cannot obtain a complete mouth region, as shown in Fig. 4(b). When the height of the facial region is expanded one-fifth downward, we obtain the complete mouth region, as shown in Fig. 4(c). However, when the mouth opens narrowly, the ROI is too large, and it will affect the extraction of the mouth internal contour; thus, it is necessary to accurately locate the mouth. Based on the difference between the colors of the lips and the skin, the mouth region is precisely positioned according to lip segmentation, and we split the lips according to the value of chromatism s of the RGB space.[27] The value of chromatism s is defined as follows:

$$s = 2 \times \tan^{-1}\left(\frac{R-G}{R}\right)/\pi. \tag{4}$$

Experiments demonstrate that the value of chromatism s of the lip region is larger than that of the skin. Assuming that the number of pixels belonging to the lip region is N_0, we rank all of the pixel saturation values and select the N_0 pixels with the largest value as the lip region. Talea and Yaghmaie[28] proposed that N_0 is in the range of 10% to 20% in the initial ROI of the mouth. Pan et al. set N_0 to 15% of the ROI of the mouth.[27] In this study, we set N_0 to 20% to extract the complete lip region and the selected pixels as shown in Fig. 4(d) in the white part. Considering that the upper and lower lips are not connected at all times as shown in Fig. 4(d), and the difference between the upper and lower lips size is not large, we select the largest two connected components, according to the size of the connected components and determine whether the upper and lower lips are connected. The external rectangle of the two largest connected components (when the upper and lower lips are not connected) or a connected component (when the upper and lower lips are connected) is the located mouth region. In fact, the final region of the mouth is slightly larger than the rectangular box.

3 Contour Feature Extraction

After locating the eye and mouth region, we judge the state of the eyes as open or closed by extracting the eye contour and analyzing the open state of the mouth by extracting the mouth internal contour.

3.1 Eye Contour Extraction
3.1.1 Sclera extraction

Sclera is the white part of the eye. Based on the difference between the sclera and skin saturation, the large difference between the red and blue components of the skin and the small relative difference between the sclera regions, the sclera region is segmented by a K-means clustering method. First, we exclude the impact of the iris[29,30] and eyelashes; eyebrows are included in certain instances. Given that the gray value of the iris and eyelash region is the smallest and the scleral gray value is larger than the skin region, we obtain the best segmentation threshold T via the OTSU algorithm, the threshold segmentation of the image on the basis of threshold T and then divide the eye into two parts: the iris and eyelash region as shown in the blue region in Fig. 5(a). In the rest of the sclera and the skin region, we use the difference between the red component R and blue component B between the sclera and the skin (R-B) and cluster them into two parts according to a K-means clustering characteristic. The final eye region is divided into three parts, as shown in Fig. 5(b). According to the characteristics of the sclera saturation, the S value is small and adjacent to the iris to obtain the sclera region. In addition, through the saturation S analysis in the HSV space, the pixels with large individual saturation values are removed to accurately locate the sclera region, as shown in the green area of Fig. 5(c).

3.1.2 Contour fitting

According to the extracted sclera region, the boundary point of the sclera is selected to fit the upper and lower eyelids via a quadratic curve, as shown in Fig. 6(a) (the two intersections of the curve are defined as the left and right corners). The
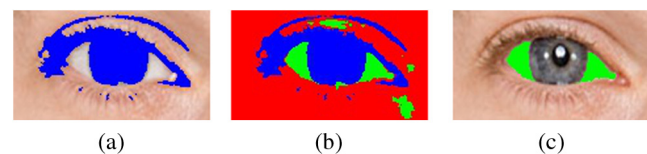

Fig. 5 Illustration of sclera extraction: (a) Iris and eyelash regions. (b) Eye region is divided into three parts. (c) Sclera region.
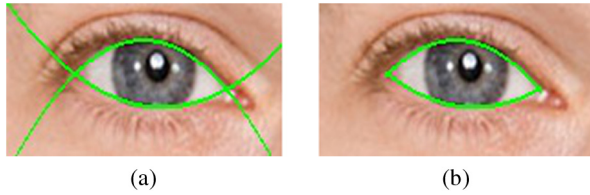
Fig. 6 Illustration of contour fitting: (a) Upper and lower eyelids after contour fitting. (b) Minimum circumscribed rectangle.

minimum circumscribed rectangle of the eye contour is calculated, as shown in Fig. 6(b), and the aspect ratio of the rectangle is used to determine whether the eye is open or closed. The details are described in the next section.

3.1.3 Special circumstances–eye strabismus

There is another special case: when the human eye is strabismus, as shown in Fig. 7(a), the sclera of the iris side and the effect can yield a poor fit to the eye contour with the aforementioned method. By extracting the boundary points of the sclera and iris to fit the contour of the eye, we first estimate the center of the iris. When the sclera is in the side of the iris center, the eye is in a state of strabismus. Given that the iris has the lowest R value in the eye region, we use a rectangle with a side length of r to traverse the entire human eye image; when the sum of the R components of all pixels in the rectangle is the smallest, the rectangular center is regarded as the center of the iris. In Ref. 1, the author proposes that the iris radius R is proportional to one-tenth of the distance between the centers of the eyes; assuming that the distance between the eyes is D, we take $r = 0.2\ D$, as the final location of the center of the iris as shown by the red dot in Fig. 7(b). A partial image of the iris with the iris center as the boundary is considered, and the side length is a quarter of D, as shown in Fig. 7(c). The binary image is obtained by threshold segmentation using the OTSU algorithm, as shown in Fig. 7(d); a morphological operation is used to remove the noise, and then certain iris boundary points are extracted.
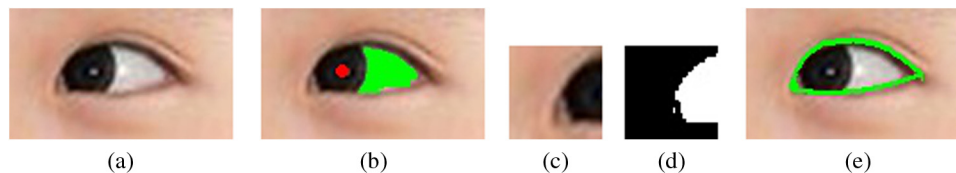
With the previously proposed method, we extract the sclera, as shown in the green area of Fig. 7(b), and part of the sclera boundary points. According to the extracted boundary points of the sclera and iris, the eye contour is fitted via a quadratic curve, as shown in Fig. 7(e).

3.2 Mouth Internal Contour Extraction

In the RGB space, compared with the skin, the difference between the red component R and green component G of the lips is larger.[31] Considering that the mouth is open, particularly when a person yawns, the RGB value of the internal part of the mouth is in balance, even if the teeth are exposed. The relationship between the $R - G$ value of the mouth, lips, and skin pixels is as follows: lips > skin > mouth internal. Owing to the fact that the difference of the lips' R component and G component is the largest, we can effectively separate the internal part of the skin and the mouth. We set the adaptive threshold T according to the following formula via threshold segmentation to obtain the binary image, as shown in Fig. 8(a):

$$h(x, y) = \begin{cases} 255 & R - G \leq T \\ 0 & R - G > T \end{cases}. \tag{5}$$

In order to obtain the internal contour of the mouth more accurately, we calculate the optimal threshold via an iterative method to determine the segmentation threshold T. The algorithm for calculating the threshold T is illustrated by the following steps:

1. Calculate the $R - G$ values of all pixels, and sort them from smallest to largest;
2. Select the top 10% of all pixels and set them to 255. Set the rest of the pixel gray values to 0. Remove image noise via a morphological operation, analyze the position of each connected component (white part), and calculate the number of pixels that do not belong inside the mouth, with skin value;

Fig. 7 Different steps of contour fitting when the eye is strabismus: (a) Original image. (b) Iris center and sclera detection. (c) Image after intercepting part of iris. (d) Binarized image. (e) Contour fitting.
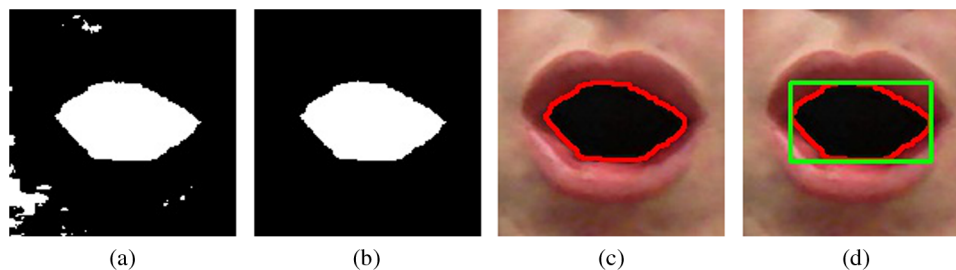

Fig. 8 Illustration of mouth internal contour extraction steps: (a) Binary image after threshold segmentation. (b) Mouth inner area. (c) Mouth contour fitting. (d) Minimum circumscribed rectangle of the mouth contour.

3. Determine whether the skin value is greater than 1/20th of the total number of pixels of the image, and if so, T is the maximum value of $R - G$ in these selected pixels. Otherwise, proceed to the next step;

4. Select n pixels with a small $R - G$ value from the remaining pixels where n is 1/15th of the total number of pixels of the image, merge into the previously selected pixels, and calculate the skin value according to the method in 2. Then, return to step 3.

The mouth image we obtain is located in the center of the region; therefore, if the center of mass of the connected component (white part) is near the center of the image, the connected component is regarded as the inner area of the mouth. The internal area of the mouth is obtained by position analysis of each connected component, as shown in Fig. 8(b). The extracted external contour of the connected component is the internal contour of the mouth, as shown in Fig. 8(c). In addition, calculating the minimum circumscribed rectangle of the contour of the mouth, as shown in Fig. 8(d), determines whether the person is yawning by the aspect ratio of the external rectangle.

4 State Analysis and Test Results

In order to verify the efficiency of our algorithm, we conduct the experiment against some image databases, including the color face database of the California Institute of Technology (CIT), part of the FERET face database, and self-built face database. These pictures are taken under different conditions, such as different lighting conditions, different indoor and outdoor environments, and different head postures. Our experiments are conducted on the basis of the acquisition of the face region, and the face region is obtained by applying the Viola Jones face detector on the OpenCV 2.4 platform.

4.1 State Analysis

4.1.1 Eye state analysis

We determine the eye state according to the aspect ratio of the smallest external rectangle of the eye contour. Assuming that the length of the rectangle is L, the width (or height) is H, the eyelid closure value M is defined as follows:

$$M = \frac{H}{L}. \tag{6}$$

We define the threshold value T according to the following criteria to determine the eye state:

$$\text{Eye state} = \begin{cases} \text{open } M > T \\ \text{closed } M \leq T \end{cases}. \tag{7}$$

For the determination of the threshold T, we use the P80 standard in the PERCLOS parameter; when the degree of eye closure is more than 80%, the eye is in a closed state. In order to evaluate the eyelid aspect ratio when the eyes open normally, we performed related experiments. First, we collected eye images from different people whose eyes are open normally in the images, and then calculated the eyelid aspect ratio for each image. Experimental results show that the eyelid aspect ratio is ~16:9 when the eye is open normally. According to the P80 standard, the eyelid aspect ratio is less

than 0.1125 when the eyes are closed, i.e., the T value should be 0.1125. However, the experimental results show that when $M < 0.15$, i.e., $T = 0.15$, the eye is in a closed state. In this case, the judgment of the eye state is more accurate, with an accuracy rate of 98.67%. In addition, when the eye is completely closed, the eye region through the OTSU algorithm is divided into two parts, and this time, we cannot detect the sclera region. Therefore, when the sclera region is not detected or the eyelid closure value $M < T$, the eye is considered to be in a closed state.

4.1.2 Mouth state analysis

According to the internal contour of the mouth, the mouth opening degree N is defined by the aspect ratio of the external rectangle

$$N = \frac{H_1}{L_1}, \tag{8}$$

where L_1 is the length and H_1 is the width (or height) of the rectangle.

In general, when people yawn, the mouth is open widely and this lasts for a few seconds. There is no universal international standard to determine the degree of mouth opening; therefore, we perform experiments to determine the degree of mouth opening when people yawn. First, we collect more than 100 pictures of people yawning from different groups, and the degree of mouth opening in these pictures is different. We calculate the mouth opening degree N of these pictures according to the proposed algorithm. The experimental results show that when $N > 0.75$, the degree of mouth opening reaches the degree of mouth opening when yawning. In this case, the detection of the mouth state is more accurate, with an accuracy rate of 97.5%. Yawning is a process that lasts for a few seconds, with the mouth widely open in continuous multiframe pictures. Although the N value has a little error, it has little effect on the result. In addition, the degree of mouth opening is very small at times, but the N value is greater than 0.75. In order to eliminate this situation and improve accuracy, the state detection method needs to satisfy the condition that the internal area of the mouth is greater than one-eighth of the image size. When the mouth opening degree is detected to be $N > 0.75$, the internal area of the mouth is large enough; if the mouth is open widely and it lasts for a few seconds, then it can be determined that the driver is yawning.

4.2 Test and Experimental Results

The performance of the proposed algorithm is evaluated on three databases. During the eye state detection, we observe the eyes in the image to decide their state (open or closed) and compare them with the state obtained from the algorithm. The experimental results demonstrate that the algorithm can fit the contour of the eyes with different opening states, similar to the experimental results shown in Table 1, and six fitting effect diagrams shown in Fig. 9, including the situation of different lighting conditions, different groups, different open eye sizes, and wearing glasses.

In the CIT and FERET databases, there are many pictures where the teeth are exposed or with a beard. Under this situation, the gray-projection algorithm to detect the mouth state may lose its accuracy;[14] however, the algorithm discussed

in this paper can obtain a better result of the internal contour of the mouth, as the experimental results shown in Table 2. Then, we provide six fitting effect diagrams of the mouth contour as shown in Fig. 10, including the situation of different lighting conditions, beards, different opening sizes, and exposed teeth.

In fact, the pictures in the database were obtained under different lighting conditions. Reference 14 uses the projection method to determine the eye and mouth state, that is, the image is projected horizontally and vertically to calculate the sum of the gray values of the pixels in the horizontal and vertical directions. This method is greatly affected by the light intensity, and the detection effect is decreased significantly under uneven illumination. The algorithm proposed in this paper performs illumination compensation on the image before face detection, which reduces the impact of highlights and shadows on the experiment. Thus, the proposed algorithm exhibits robustness against illumination changes. The performance of the algorithm will be reduced for images obtained under dramatic lighting changes. In addition, there are several mouth pictures where the teeth are exposed or with a beard. Owing to the large differences between the gray value of teeth and beards and the skin color, the projection method does not work well in this case. Our algorithm takes into account the color difference between teeth, beard, and skin and determines the degree of mouth opening by obtaining the internal contour of the mouth. The experimental results show that the detection effect is significantly improved, and the performance comparison is shown in Tables 3 and 4.

Table 1 Performance of the proposed eye state detection algorithm.

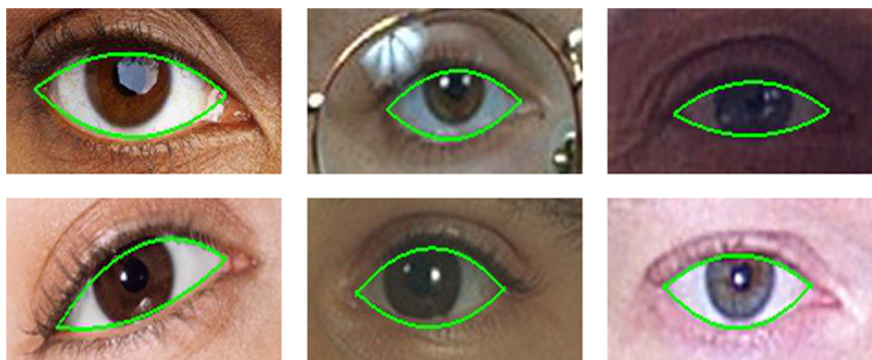| Database	Number of images	Number of open	Number of closed	Open state determination accuracy	Closed state determination accuracy
CIT	497	442	55	96.15%	96.36%
FERET	418	374	34	97.06%	94.18%
Self-built database	146	84	62	97.95%	96.77%

Fig. 9 Eye contour fitting results obtained by the proposed algorithm.

Table 2 Performance of the proposed mouth state detection algorithm.

Database	Number of images	Number of open	Number of closed	Open state determination accuracy	Closed state determination accuracy
CIT	438	386	52	95.34%	96.15%
FERET	418	374	34	96.52%	94.18%
Self-built database	146	84	62	98.63%	98.39%

Fig. 10 Mouth contour fitting results obtained by the proposed algorithm.

Table 3 Performance comparison.

	Eye state recognition accuracy	
Number of images	Proposed method	Projection method in Ref. 14
254	97.64%	92.52%

Table 4 Performance comparison.

	Mouth state recognition accuracy	
Number of images	Proposed method	Projection method in Ref. 14
378	96.56%	91.27%

5 Conclusion

In this paper, we proposed a method for detecting the eye and mouth state by extracting contour features. In each step, we presented new algorithms and modifications to achieve better results. The eye contour is fitted by extracting the sclera border points, and the eyelid closure value M is defined according to the smallest circumscribed rectangle of the eye contour to determine whether the eye is open or closed. When the mouth is open, the internal contour of the mouth is extracted according to its external rectangle to define the mouth opening degree N for analyzing the mouth open state. The experimental results demonstrate that the proposed algorithm has a high accuracy rate in different environments, and it does not require training data to improve the computational efficiency. The proposed algorithm can be used in future energy vehicles,[32] intelligent vehicles, and advanced driver assistant systems and can be further analyzed to determine whether a driver is fatigued by detecting the state of the eyes and mouth. When driver fatigue is detected, it can issue a warning to remind the driver to pay attention to driving in order to reduce traffic accidents. Our future work is to build a fatigue model, which can provide a warning signal to remind the driver when he/she appears to be demonstrating fatigue symptoms.

References

1. P. R. Tabrizi and R. A. Zoroofi, "Open/closed eye analysis for drowsiness detection," in *1st Workshops on Image Processing Theory, Tools and Applications*, pp. 1–7, IEEE, Sousse, Tunisia (2008).
2. A. Jain et al., "Fatigue detection and estimation using auto-regression analysis in EEG," in *Int. Conf. on Advances in Computing, Communications and Informatics*, pp. 1092–1095, IEEE, Jaipur (2016).
3. D. G. K. Madusanka et al., "Hybrid vision based reach-to-grasp task planning method for trans-humeral prostheses," *IEEE Access* **5**(99), 16149–16161 (2017).
4. X. Q. Huo, W. L. Zheng, and B. L. Lu, "Driving fatigue detection with fusion of EEG and forehead EOG," in *Int. Joint Conf. on Neural Networks (IJCNN)*, pp. 897–904, IEEE, Canada (2016).
5. X. Song et al., "The anti-fatigue driving system design based on the eye blink detect," *Proc. SPIE* **10322**, 103221R (2017).
6. M. Ramasamy and V. K. Varadan, "Real-time monitoring of drowsiness through wireless nanosensor systems," *Proc. SPIE* **9802**, 98021G (2016).
7. Z. Wang et al., "The effect of a haptic guidance steering system on fatigue-related driver behavior," *IEEE Trans. Hum. Mach. Syst.* **47**(5), 741–748 (2017).
8. P. N. Bhujbal and S. P. Narote, "Lane departure warning system based on Hough transform and Euclidean distance," in *Third Int. Conf. on Image Information Processing (ICIIP)*, pp. 370–373, IEEE, Waknaghat (2015).
9. R. H. Zhang et al., "Vehicle detection method for intelligent vehicle at night time based on video and laser information," *Int. J. Pattern Recognit. Artif. Intell.* **32**(4), (2017).
10. R. H. Zhang et al., "Study on self-tuning tyre friction control for developing main-servo loop integrated chassis control system," *IEEE Access* **5**(99), 6649–6660 (2017).
11. E. J. Delp, "Video-based real-time surveillance of vehicles," *J. Electron. Imaging* **22**(4), 041103 (2016).
12. F. You et al., "Monitoring drivers' sleepy status at night based on machine vision," *Multimedia Tools Appl.* **76**(13), 14869–14886 (2017).
13. Y. F. Lu and C. L. Li, "Recognition of driver eyes' states based on variance projections function," in *3rd Int. Congress on Image and Signal Processing*, pp. 1919–1922, IEEE, China (2010).
14. M. Omidyeganeh, A. Javadtalab, and S. Shirmohammadi, "Intelligent driver drowsiness detection through fusion of yawning and eye closure," in *IEEE Int. Conf. Virtual Environments, Human-Computer Interfaces and Measurement Systems Proc.*, pp. 1–6, IEEE, Ottawa (2011).
15. L. N. Jia et al., "Smartphone-based fatigue detection system using progressive locating method," *Inst. Eng. Technol.* **10**(3), 148–156 (2016).
16. B. Mandal et al., "Towards detection of bus driver fatigue based on robust visual analysis of eye state," *IEEE Trans. Intell. Transp. Syst.* **18**(3), 545–557 (2017).
17. A. Punitha, M. K. Geetha, and A. Sivaprakash, "Driver fatigue monitoring system based on eye state analysis," in *Int. Conf. Circuits, Power and Computing Technologies*, pp. 1405–1408, IEEE, India (2014).
18. C. Du and S. Gao, "Image segmentation-based multi-focus image fusion through multi-scale convolutional neural network," *IEEE Access* **5**(99), 15750–15761 (2017).
19. L. Zhao et al., "Human fatigue expression recognition through image-based dynamic multi-information and bimodal deep learning," *J. Electron. Imaging* **25**(5), 053024 (2016).
20. R. Kavi et al., "Multiview fusion for activity recognition using deep neural networks," *J. Electron. Imaging* **25**(4), 043010 (2016).
21. Y. Y. Liu et al., "Fast robust face detection under a skin color model with geometry constraints, in *Int. Conf. on Computational Intelligence and Security*, pp. 515–519, IEEE, Beijing (2009).
22. P. Viola and M. J. Jones, "Robust real-time face detection," in *Proc. of Eighth IEEE Int. Conf. on Computer Vision*, pp. 747–747, IEEE, Canada (2001).
23. R. L. Hsu, M. Abdel-Mottaleb, and A. K. Jain, "Face detection in color images," *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(5), 696–706 (2002).
24. V. H. Jimenez-Arredondo, J. Cepeda-Negrete, and R. E. Sanchez-Yanez, "Multilevel color transfer on images for providing an artistic sight of the world," *IEEE Access* **5**, 15390–15399 (2017).
25. H. Kalbkhani, M. G. Shayesteh, and S. M. Mousavi, "Efficient algorithms for detection of face, eye and eye state," *IET Comput. Vision* **7**(3), 184–200 (2013).
26. U. Saeed and J. L. Dugelay, "Combining edge detection and region segmentation for lip contour extraction," *Lect. Notes Comput. Sci.* **6169**, 11–20 (2010).
27. J. Pan, Y. Guan, and S. Wang, "A new color transformation based fast outer lip contour extraction," *J. Inf. Comput. Sci.* **9**(9), 2505–2514 (2012).
28. H. Talea and K. Yaghmaie, "Automatic combined lip segmentation in color images," in *IEEE 3rd Int. Conf. on Communication Software and Networks*, pp. 109–112, IEEE, China (2011).
29. T. N. Tan, Z. F. He, and Z. N. Sun, "Efficient and robust segmentation of noisy iris images for non-cooperative iris recognition," *Image Vision Comput.* **28**(2), 223–230 (2010).
30. A. Mollahosseini, A. D. Anjos, and H. R. Shahbazkia, "Accurate extraction of iris and eye corners for facial biometrics," in *Int. Conf. on Environment and BioScience*, pp. 1–5, IACSIT Press, Singapore (2011).
31. V. E. C. Ghaleh and A. Behrad, "Lip contour extraction using RGB color space and fuzzy c-means clustering," in *IEEE 9th Int. Conf. on Cybernetic Intelligent Systems*, pp. 1–4, IEEE, United Kingdom (2010).
32. R. H. Zhang et al., "Exploring to direct the reaction pathway for hydrogenation of levulinic acid into gamma-valerolactone for future clean-energy vehicles over a magnetic Cu-Ni catalyst," *Int. J. Hydrogen Energy* **42**(40), 25185–25194 (2017).

Yingyu Ji received his BS degree from the School of Information Technology, Hebei University of Economics and Business, Hebei, China, in 2016. He is currently pursuing his MS degree with the College of Communication Engineering of Jilin University. His major research interests include pattern recognition and image processing.

Shigang Wang received his BS degree from Northeastern University in 1983, his MS degree in communication and electronics from Jilin

University of Technology in 1998, and his PhD in communication and information system from Jilin University in 2001. Currently, he is a professor of communication engineering. His research interests include multidimensional signal processing and stereoscopic, multiview video coding, and so on.

Yang Lu received her BS degree from the College of Communication Engineering, Jilin University, Jilin, China, in 2014. Currently, she is pursuing her PhD in the College of Communication Engineering of Jilin University. Her major research interests include pattern recognition and image processing.

Jian Wei received his BS degree from Jilin University in 2008, his MS degree in communication and information systems from Jilin

University in 2011, and his PhD in informatics from Tuebingen University in 2016. Currently, he is working in the Department of Communication Engineering, Jilin University. His research interests include multiview stereo and 3D display technology.

Yan Zhao received her BS degree in communication engineering from Changchun Institute of Posts and Telecommunications in 1993, her MS degree in communication and electronic from Jilin University of Technology in 1999, and her PhD in communication and information system from Jilin University in 2003. Currently, she is a professor of communication engineering. Her research interests include image and video processing, image and video coding, and so on.