Vision-Based Fatigue Driving Recognition Method Integrating Heart Rate and Facial Features

Guanglong Du, Tao Li, Chunquan Li, Peter X. Liu, *Fellow, IEEE*, and Di Li

***Abstract*— Driving fatigue can be detected by measuring drivers' heart rate with a wearable device or extracting their facial features with an RGB camera. However, a wearable device causes inconvenience and discomfort to the driver, and an RGB camera's detection accuracy may be affected by light, glasses, and head orientation. Furthermore, most existing methods ignored the temporal information of fatigue features and the relationship between the features, lowering recognition accuracy. Additionally, some existing fatigue detection methods focused on dealing with fatigue features with a temporal slice, ignoring temporal variations in the features. To address these problems, a single RGB-D camera is first used to extract three fatigue features: heart rate, eye openness level, and mouth openness level. More importantly, this paper proposes a novel multimodal fusion recurrent neural network (MFRNN), integrating the three features to improve the accuracy of driver fatigue detection. Specifically, a recurrent neural network (RNN) layer is applied in the MFRNN to obtain the temporal information of the features. Since the heart rate feature is a physiological signal extracted indirectly, it contains more noise and is fuzzier than the other features. To deal with the fuzziness and noise, we combine fuzzy reasoning with RNN to extract the temporal information of the heart rate. To identify the relationship between the features, we develop a new relationship layer containing a two-level RNN, for which the input is the temporal information of the features. Both the simulation and field experiment results show that the proposed method provides better performance than similar methods.**

***Index Terms*— Fatigue driving, recurrent neural network (RNN), fuzzy reasoning, RGB camera, heart rate.**

Manuscript received October 13, 2018; revised March 14, 2019, August 15, 2019, and December 31, 2019; accepted February 10, 2020. Date of publication March 16, 2020; date of current version May 3, 2021. This work was supported in part by the Natural Science Foundation of China under Grant 61973126, Grant 61863028, Grant 81660299, and Grant 61503177, in part by the Guangdong Natural Science Funds for Distinguished Young Scholar under Grant 2017A030306015, in part by the Pearl River S&T Nova Program of Guangzhou under Grant 201710010059, in part by the Guangdong Special Projects under Grant 2016TQ03 × 824, in part by the Fundamental Research Funds for the Central Universities under Grant 2019ZD27, in part by the Science and Technology Department of Jiangxi Province of China under Grant 20161ACB21007, Grant 20171BBE50071, and Grant 20171BAB202033, in part by the Science and Technology Planning Project of Guangdong Province under Grant 2017B090914002, and in part by the Innovation Team of the Modern Agriculture Industry Technology System in Guangdong Province under Grant 2019KJ139. The Associate Editor for this article was L. M. Bergasa. *(Corresponding author: Chunquan Li.)*

Guanglong Du and Tao Li are with the School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China (e-mail: csgldu@scut.edu.cn; cstaoli@mail.scut.edu.cn).

Chunquan Li is with the School of Information Engineering, Nanchang University, Nanchang 330029, China (e-mail: lichunquan@ncu.edu.cn).

Peter X. Liu is with the Department of Systems and Computer Engineering, Carleton University, Ottawa, ON K1S 5B6, Canada (e-mail: xpliu@sce.carleton.ca).

Di Li is with the School of Mechanical and Automotive Engineering, South China University of Technology, Guangzhou 510006, China.

Digital Object Identifier 10.1109/TITS.2020.2979527

I. Introduction

WHEN drivers are in a fatigue-driving state, their ability to recognize road conditions and driving skills are significantly degraded. Research results show that 25%-30% of traffic accidents are caused by fatigue-driving [1]. To overcome this problem, it is important to develop a system that can effectively detect drivers' fatigue-driving and warn them in a timely manner. In the literature, vehicle-based behavior, drivers' physiological state, and facial expression have been used to recognize fatigue driving [2].

The vehicle behavior-based methods mainly measure vehicle data such as steering angle, speed, acceleration, and turning angle [3]–[5] without considering the physiological signals for detecting driver fatigue and make early warnings. The physiological signal-based methods mainly apply electrooculogram (EOG) and electrocardiogram (ECG) [6], heart rate (HR) [7], heart rate variability (HRV) [8], electroencephalogram (EEG) [9] and other physiological signals. However, drivers must wear relevant devices that are invasive, hinder driving, and lead to unfavorable user experiences.

The behavior-based methods detect fatigue driving by visually analyzing facial features such as eyelid closure duration, blinking, yawning, head posture, eyelid movements, and facial expressions [10]. Because methods based on visual behavior do not interfere with driving, they are more acceptable to the drivers. In terms of behavior-based methods, some related investigations [11]–[15] have been presented to determine the fatigue as a percentage of eyelid closure (PERCLOS) by detecting the eyelid closure frequency. As an unconscious behavior caused by fatigue, yawning is also used for visual fatigue detection. References [16]–[18] achieve good results in fatigue testing using yawning facial video, which verifies the feasibility of fatigue detection through facial expressions. However, a driver suddenly opening his or her mouth or an eye blinking motion caused by high beams can decrease recognition accuracy.

Note that the above methods are all single source-based detection methods that use only one source of information, such as vehicle-based behavior, physiological signals, or facial expression, instead of combinations of these. Therefore, methods based on a single source of information have some limitations in robustness and reliability.

To overcome the limitations associated with using a single source of information, some methods have combined multiple information sources for fatigue detection. These methods, which involve the extraction of fatigue features and the construction of fusion models, play an important role in improving fatigue-driving recognition. Sun *et al.* [19] proposed a contextual feature two-level fusion method based on a multiclass support vector machine (MCSVM). References [20] and [21] both use dynamic Bayesian networks to fuse multiple fatigue features. In particular, these fusion methods combine transient fatigue features into the input vector of the classifier to estimate the driver fatigue state. However, they ignore the fact that the fatigue state of an individual is a continuous-time process. Therefore, we need to extract feature changes over a period of time to better recognize fatigue driving. Although the above methods are promising, further improvements are required for the following reasons:

1) Driver fatigue is a continuous-time process, so temporal variations in fatigue features are important for fatigue-driving recognition. However, existing algorithms [19]–[21] focus on dealing with features with a temporal slice, ignoring temporal variations in fatigue features.
2) Some methods can detect fatigue by measuring heart rate with a wearable device. However, wearing a device is inconvenient for drivers and may make them uncomfortable.
3) Existing fatigue-driving detection methods extract eye-opening by using RGB images, which can be affected by light, glasses and head orientation.
4) Few existing models capture both temporal information of features and temporal relationship information between features for driver fatigue detection. For example, a greater degree of mouth opening does not necessarily indicate fatigue, which can be reassessed by changes in heart rate and eye openness level over time.

To address the aforementioned limitations, we develop a new multimodal fusion neural network for fatigue identification and a new vision-based framework for calculating heart rate statistics, eye openness level, and mouth openness level. Since the recurrent neural network (RNN) can record and analyze the changes in data over a period of time, it can better identify driver fatigue by extracting temporal information related to each fatigue feature. Because of the fuzziness of fatigue and the noise of heart rate, we combine fuzzy reasoning with a recurrent neural network to address the fuzziness and noise and to extract temporal information related to heart rate. Then, the information extracted from the previous layers is input to the proposed relationship layer. The relationship layer is a two-level RNN that determines the relationships between the features. In addition, considering that wearing a monitor to detect heart rate causes driver discomfort, we designed a non-contact method to extract heart rate from RGB and infrared video. The main contributions of this paper can be summarized as follows.

1) We propose an MFRNN to extract temporal information related to fatigue-driving features and the relationship information among the features to improve the performance of driver fatigue detection.
2) We combine fuzzy reasoning with RNN to address the fuzziness and noise and to extract temporal information related to heart rate.

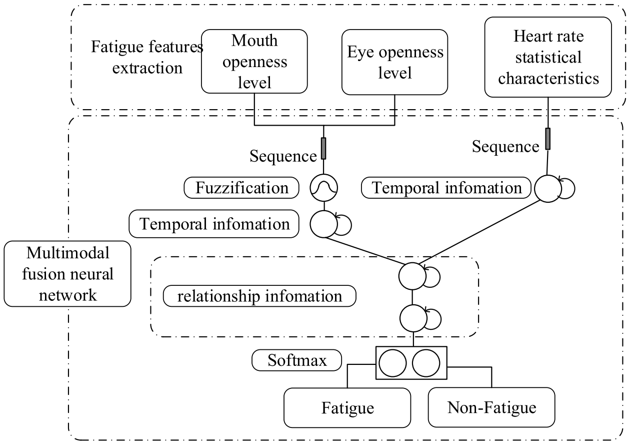


Fig. 1. Framework of the proposed fatigue detection method.

3) A noncontact method to obtain heart rate is introduced to improve the accuracy of driver fatigue detection and avoid the driver discomfort caused by wearable devices. Moreover, because the depth image is acquired by emitting infrared light, it is less affected by illumination, wearing glasses and changes in head orientation. Using depth images is more accurate for determining eye and mouth opening degrees.

This paper is organized as follows. Section II introduces an overview of the vision-based fatigue detection framework. Section III illustrates the methods for estimating heart rate, eye openness level, and mouth openness level. Section IV demonstrates the structure of the MFRNN. The experimental environment and results are presented in Section V. The last section provides a conclusion and discusses further work.

II. Overview

Fig. 1 shows the whole framework of the proposed fatigue detection method, involving two parts: fatigue feature extraction and fatigue feature fusion. The eye openness level, mouth openness level, and heart rate are used as the fatigue-driving features. See below for a more detailed description of the fatigue detection framework.

(1) Fatigue feature extraction. We use an RGB-D camera to obtain face RGB and infrared video. The face video is analyzed by independent quantity using the joint approximate diagonalization of the eigen matrices (JADE) algorithm [26], which obtains the independent quantity matrix of the four channels of RGB and infrared. A Fourier transform is then applied to extract the image change frequency, which obtains heart rate (HR). The triangular surface patch (TSP) descriptor [30] algorithm is further employed to track facial orientation and to extract facial feature points based on depth images captured by the RGB-D camera. The eye and mouth openness degrees can be calculated according to the feature points of the mouth and eyes.

(2) Fatigue feature fusion. We devise a novel MFRNN for fatigue identification. MFRNN can classify the above feature sequences into two states: the fatigue state and the non-fatigue state.

Note that the fatigue features include two data sources: one is the physiological signals, namely heart rate obtained through frequency analysis, and the other is the facial features including eye openness level and mouth openness level. Therefore, we develop two different recurrent neural network layers to extract the data from the two different data sources. For the statistical characteristics of heart rate, there is a fuzzy relationship between heart rate and fatigue, and heart rate contains substantial noise. We combine fuzzy reasoning with an RNN to improve its anti-noise ability. Fuzzy reasoning allows the neural network to better handle the relationship between heart rate changes and fatigue.

There may be a relationship associated with fatigue among the above three features. To extract the relationship, we use a relationship layer to extract their temporal and logical relationships. Finally, to output the classification results, Softmax is employed to amplify the discrimination of the output. The proposed MFRNN model classifies the input feature sequence into two categories, namely, fatigue and non-fatigue.

III. Fatigue Features Extraction

This section describes how to estimate the statistical characteristics of heart rate, eye openness level, and mouth openness level. The data used are the face video captured by the RGB-D camera. The depth data for each point in the depth image reflects the distance between the object being photographed and the depth camera. For example, the opening and closing of either the eyelids or the mouth can change the distance between the eyes and the mouth from the RGB-D camera. All of these changes are well reflected in the depth data from the depth image captured by the RGB-D camera. The triangular surface patch (TSP) descriptor [30] algorithm is utilized to track facial orientation and extract facial feature points based on depth images captured by the RGB-D camera.

A. Estimation of Statistical Characteristics of Heart Rate

The blood color of the face changes slightly with the relaxation and contraction of the heart. Therefore, the heart rate can be obtained by recording the face color change frequency with the camera. Here, we use an RGB-D camera to record RGB and IR images of the driver's face and analyze the subtle changes in brightness for identifying the driver's heart rate. In fact, the facial video captured by the RGB-D camera can be regarded as the superposition of heart rate signals and facial image signals. Since heart rate and facial image are independent data points, the channels of RGB and infrared images are equivalent to the observations of heart rate and facial image by multiple sensors. Therefore, this is a blind source signal separation problem [25].

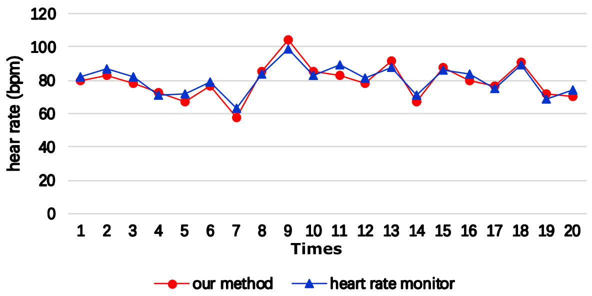


Fig. 2. Comparison of heart rate measurement between our method and the heart rate monitor.; "our method" represents the non-contact measurement of heart rate using the RGB-D camera; "heart rate monitor" means the direct contact measurement of heart rate using Polar H10.

From the above analysis, the heart rate measurement can be implemented via a non-contact method with an RGB-D camera, following three steps. First, the interesting facial image region can be extracted and tracked from the facial video. Then, the four-channel image signals of R, G, B, and IR from the video collected by RGB-D camera are normalized. Furthermore, the JADE algorithm [26] based on the feature matrix and approximate diagonalization is employed to perform independent component analysis on the mixed signals, including heart rate and facial data, generating a fourth-order separation matrix. Finally, a fast Fourier transform is performed on the separation matrix to match the heart rate range so that the real-time heart rate can be obtained.

Heart rate measurements were performed under ambient light conditions using both an RGB-D camera and a Polar H10 heart rate monitor (heart rate measurement wristbands) to measure the real-time heart rate of the same volunteer. The heart rate under a random emotional state was measured 20 times; each measurement time was 10 seconds; 20 measurements were recorded, as shown in Fig. 2. The measurement error between the RGB-D camera and the Polar H10 heart rate monitor is within ±3 bpm. This indicates that the heart rate calculated by our method that uses the non-contact measurement of heart rate with RGB-D is comparable to that measured by the Polar H10 heart rate monitor. Therefore, Fig. 2 verifies the effectiveness of our non-contact method on the heart rate measurement.

Some investigations [8], [27] have demonstrated that heart rate variability (HRV) is related to fatigue. HRV is used to measure the change features of heart rate [28], [29]. Specifically, we can extract the statistical features of a heart rate signal for fatigue-driving detection by measuring the signal change features. Here, we extracted four statistical features of the heart rate signals (mean, root mean square, and maximum and minimum amplitude) as the physiological signal characteristics for detecting fatigue. In the feature fusion neural network, the statistical features of heart rate at the i^{th} moment can be expressed as $E_i = [MEAN, RMSE, MAX, MIN, RANGE]$.

B. Estimation of Eye Openness Level and Mouth Openness Level

We use TSP descriptors [30] to track facial orientation and extract facial feature points based on the depth images captured by the RGB-D camera. According to the feature points of the mouth and eyes, the eye and mouth openness level can be calculated. The depth images can reflect the distance from the object surface to the depth camera so that they are of great advantage in recognizing eye and mouth opening. Because the RGB-D camera acquires depth data by

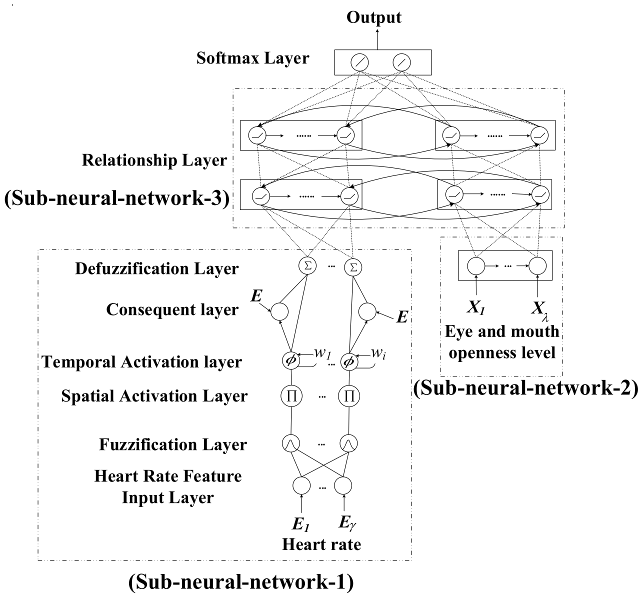


Fig. 3. The feature fusion network structure.

actively emitting infrared light, the depth data are less affected by light.

IV. Features Fusion and Fatigue Detection

Fig. 3 shows that the architecture of the proposed MFRNN contains three subnetworks, which can extract the time information related to the statistical features of heart rate, the time information related to eye and mouth openness level, and the temporal and logical relationship among the three features, respectively. The heart rate statistical features sequence are defined as E; the eye and mouth openness level feature sequence is done as X. Fig. 3 also shows that both E and X are input into the MFRNN, and the output result shows fatigue or non-fatigue driving. For the entire MFRNN model, all the weights and other parameters are optimized by the adaptive moment estimation (ADAM) algorithm [32].

A. Structure of Subneural-Network-1 for Heart Rate

Subneural-network-1 including fuzzy inference and RNN is used to extract the temporal information related to heart rate and to reduce the noise interference in the heart rate feature. $u^{(l)}$ represents the output of the layer l in the Subneural-network-1; $u_i^{(l)}$ does the i^{th} node (neuron) of the layer l.

1) The First Layer (Heart Rate Feature Input Layer): The statistical characteristics sequence of heart rate is the input of the first layer. The inputs are denoted by $E = \left(E_1 \ldots E_\gamma\right)$ where γ denotes the number of neurons in the input layer. This layer does not participate in any calculation. Each neuron is directly passed to the next layer as follows:

$$u_i^{(1)} = E_i \in \left[1, \gamma\right] \tag{1}$$

2) The Second Layer (Fuzzification Layer): Fuzzification can attenuate the interference noise in the heart rate feature. Note that the fuzzification layer is also called the membership function layer. In this layer, the Gaussian membership function is used to calculate the membership value of the data from the first layer. The Gaussian membership function calculation formula is given as follows:

$$u_{ij}^{(2)} = \exp\left(-\frac{\left[u_{ij}^{(1)} - m_{ij}\right]}{\sigma_{ij}^2}\right) i \in \left[1, \gamma\right], \quad j \in [1, n] \tag{2}$$

where m_{ij}, and σ_{ij}^2 are the mean and the variance of the Gaussian membership function, respectively; γ is the number of neurons in the input layer and also is the length of the heart rate feature sequence; n is the number of fuzzy subsets; $u_{ij}^{(1)}$ and $u_{ij}^{(2)}$ are the output of the first layer and second layer, respectively. Before training, the initial values of m_{ij} and σ_{ij} .in equation (2) can be given in the following equation (3):

$$\begin{aligned} m_i &= E_i^{(0)} \\ \sigma_{ij} &= 0.5 \end{aligned} \tag{3}$$

Here, $E_i^{(0)}$ refers to the fatigue features of the first input sample at time i.

3) The Third Layer (Spatial Activation Layer): The purpose of this layer is to calculate the membership degree of the whole feature. Each node of the third layer corresponds to a fuzzy rule as a spatial rule node function. The node of the third layer obtains the spatial activation strength $u_i^{(3)}$ using the membership degree operation received from the second layer. Then, the membership degree was calculated by the fuzzy method. Here, we used continuous cumulative multiplication as the fuzzy operator. The spatial activation strength $u_i^{(3)}$ can be calculated as follows:

$$u_i^{(3)} = \prod_{j=1}^{n} u_{ij}^{(2)}, \quad i \in \left[1, \gamma\right] \tag{4}$$

where $u_{ij}^{(2)}$ and $u_i^{(3)}$ are the output of the second layer and third layer, respectively.

4) The Fourth Layer (Temporal Activation Layer): RNN can effectively extract the temporal information of signals. Therefore, an RNN layer is used in this layer for extracting the temporal information of the heart rate. Each neuron can be computed as follows:

$$u_i^{(4)} = \phi_q^i(t) = \frac{1}{\frac{1}{u_i^{(3)}} - \log\left(sigmoid\left(w_i \phi_q^i(t-1)\right)\right)} \tag{5}$$

where t is the time step, and w_i is the weight of the previous state value in the current state value. $i, q \in \left[1, \gamma\right]$, $j \in [1, n]$. $\phi_q^i(t)$ is the output of the recurrent structure, which is a temporal activation strength. This output $\phi_q^i(t)$ combines the spatial activation strength $u_i^{(3)}$ transmitted by the previous layer with the temporal activation strength $\phi_q^i(t-1)$ at the last time $t-1$.

5) Fifth Layer (Consequent Layer): This layer uses the input of the first layer and the output of the fourth layer for weighted linear sum calculation. Each node in the fourth layer has an output to the next layer.

$$u_i^{(5)} = \sum_{j=1}^{\gamma} \rho_{ij} E_j + b_i + \omega_i u_i^{(4)} \tag{6}$$

where $i, j \in [1, \gamma], \rho_{ij}$ and ω_i is the weight parameter. $\sum_{j=1}^{\gamma} \rho_{ij} E_j$ is the weighted sum of the elements of the input data E.

6) The Sixth Layer (Defuzzification Layer): The task of the sixth layer is to perform fuzzy defuzzification. This layer adopts the weighted average defuzzification method as shown below:

$$u_q^{(6)} = \frac{\sum_i \phi_q^i(t)\, u_i^{(5)}}{\sum_i \phi_q^i(t)} \tag{7}$$

where $i, q \in [1, \gamma]$ and $u_q^{(6)}$ are the output of the defuzzification layer. $\phi_q^i(t)$ is the output of the fourth layer. The output of all neurons in this layer is passed to the relationship layer to extract the relationship between the features.

B. Structure of the Subneural-Network-2 for Eye and Mouth Openness Level

Subneural-network-2 only needs a recurrent neural network to extract the temporal information from the eyes and mouth features. Because the number of the eyes and mouth feature sequence X over a period of time is equal to λ, the number of RNN layers in the sub-network-2 is also set to λ for effectively extracting temporal information of the eyes and mouth feature at λ time series.

Due to the characteristics of RNN, RNN layer t, $t \in [1, \lambda]$ has two input vector. One is the $(t-1)^{\text{th}}$ layer hidden state h_{t-1}. The other is the t^{th} element of the input sequence X, expressed as X_t. The t^{th} layer hidden state h_t can be expressed as:

$$h_t = sigmoid\left(U_t X_t + V_t h_{t-1} + d_t\right) \tag{8}$$

where the 0th hidden state h_0 is 0, $t \in [1, \lambda]$, U_t is the weight of the links from layer $t-1$ to layer t, V_t is the weight of the input links to layer t, and d_t is the bias of layer t.

The output of layer t is expressed by the following equation:

$$y_t = sigmoid\left(Z_t h_t + s_t\right) \tag{9}$$

where $t \in t, \lambda]$ and Z_t are the weight of the output link of layer t, s_t is the corresponding bias, and y_t is the output of layer t.

C. Structure of Subneural-Network-3 for Feature Fusion and Classification

Subneural-network-3 is also called the feature relationship extraction layer (the relationship layer), and its task is to extract the relationship of the three features and fuse them.

As shown in Fig. 3, subneural-network-3 consists of two symmetric neural networks with the same structure. The input to subneural-network-3 comes from the output of the first two subnetworks. We use the features of the same time with the linear weighted sum method to obtain their quantitative relationship. The output of the previous layer is passed to the next layer in a fully connected manner. To better extract the nonlinear relationship, the rectified linear unit (ReLU) activation function is used for the output of each neuron. $R^{(l)}$ represents the output of the layer l in the relationship layer (subneural-network-3). The sequence length of the heart rate, eyes openness level, and mouth openness level varies. Therefore, we use zero padding to achieve the same time sequence length. The output sequence lengths of the two processed subneural-networks are both $\eta = max(\lambda, \gamma)$.

The first layer in the relationship layer can be expressed as follows:

$$\begin{cases} R_i^{(1)} = \sum\limits_{j=1}^{i} \mathrm{A}_{1,i}\left(\theta_{1,j} u_j^{(6)} + \tau_{1,j} y_j\right) & i \in [1, \eta] \\ R_i^{(1)} = \sum\limits_{j=1}^{i} \mathrm{B}_{1,i}\left(\theta_{1,j} u_j^{(6)} + \tau_{1,j} y_j\right) & i \in [\eta, 2\eta] \end{cases} \tag{10}$$

where $A_{1,i}$, $\theta_{1,j}$ and $\tau_{1,j}$ are the weight matrices of the first layer.

Layer l in the relationship layer can be expressed as:

$$\begin{cases} R_i^{(l)} = \sum\limits_{j=1}^{i} A_{l,i}(\theta_{l,j} R_j^{(l-1)} + \tau_{l,j} R_{j+\eta}^{(l-1)}) & i \in [1, \eta] \\ R_i^{(l)} = \sum\limits_{j=1}^{i} \mathrm{B}_{l,i}(\theta_{l,j} R_j^{(l-1)} + \tau_{l,j} R_{j-\eta}^{(l-1)}) & i \in [\eta, 2\eta] \end{cases} \tag{11}$$

where $A_{,i}$, $\tau_{l,j}$ and $\mathrm{B}_{l,i}$ are the weight matrices of layer l. Here, $l \leq 2$. This is because there are only three features, and the number of combinations between them is not much.

In the last layer, Softmax is used to amplify the differentiation of each classification result and realize the classification of fatigue and non-fatigue. ψ represents the output of the Softmax layer as follows:

$$\psi_i = \sum_{j=1}^{2\eta} D_{i,j} R_j^{(2)} \tag{12}$$

where $D_{i,j}$ is the weight matrix and will be automatically adjusted by the ADAM algorithm.

L2 regularization is used to prevent overfitting. We define a loss function based on cross-entropy cost. $\hat{\psi}$ is defined as the measured value on the Karolinska sleepiness scale (KSS) table corresponding to this feature sequence. Since the KSS value is a number used to measure fatigue, we set $\hat{\psi} = \left\{\left(\frac{KSS}{10}\right), 1 - \left(\frac{KSS}{10}\right)\right\}$. Thus, the loss function f of the whole MFRNN model is as follows:

$$f = \sum_{i=1}^{n} \hat{\psi}_i log\left(\frac{1}{\psi_i}\right) + c\left\|\hat{\psi} - \psi\right\|_2^2 \tag{13}$$

where c is a constant. We chose $c = 0.01$ as the regularization coefficient of L2 regularization.

V. Experiments

To validate the performance of the proposed method, a series of fatigue recognition experiments were performed. Three types of experiments were set up: namely, cross-subject experiments (simulation driving), special conditions experiments (simulation driving), and field experiments.

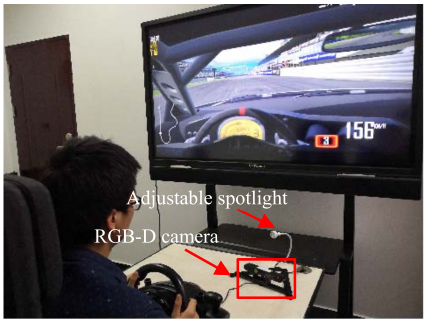


Fig. 4. An RGB-D camera is placed in front of the volunteer and fatigue recognition is performed by reading the facial video data from the volunteers.

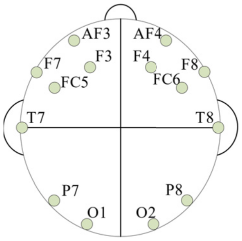


Fig. 5. The relative position between the electrodes.

A. Simulation Driving Experiment Design

Similar to [36]–[39], we also used sleep deprivation to create fatigue. Twenty volunteers, 14 males and six females participated in our investigation, and their ages ranged from 20 to 34 years old. These volunteers were divided into two groups: the nonsleep-deprived and sleep-deprived groups. In the nonsleep-deprived group, the volunteers were required to sleep 9 hours a day before the test, from 22:00 pm to 7:00 am; those in the sleep-deprived group were told to sleep for only 5 hours, from 1:00 am to 6:00 am. Actually, references [48]–[51] have pointed out that 5-hour sleep deprivation can cause subjects to fatigue.

For the nonsleep-deprived or sleep-deprived group, each volunteer played a driving simulation game for three hours, from 9:00 am to 12:00 am. Fig. 4 shows that during the driving simulation process, an RGB-D camera was placed approximately 0.5 m in front of the volunteer to record the volunteer's face, and the videos were divided into 5 seconds segments. From these segments, we extracted the three fatigue-driving features, namely, heart rate and degrees of eye openness and mouth openness. Here, each frame of each video is 1920 × 1080 pixels.

To obtain more accurate data labels for the proposed method, we need to assess the fatigue level of each volunteer in the nonsleep-deprived or sleep-deprived group. Similar to references [19], [41], and [42], we adopted the Karolinska sleepiness scale (KSS) mapping procedure for fatigue assessment. Specifically, the KSS mapping procedure contains two steps: (i) objective assessment and (ii) self-assessment. The objective assessment used the EEG data of the subjects to assess their fatigue level [43]. The self-assessment used questionnaires to assess their fatigue level. The two assessments are demonstrated as follows.

First, the objective assessment is employed to assess the subjects' fatigue status before self-assessment. The EMOTIV EPOC + device was used to consistently record the EEG signals of the subjects when executing a driving simulation game. Those EEG signals were used to evaluate and determine driver's fatigue states. Here, EEG signals were recorded from 14 channels. The 14 channels are called AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4. This is an internationally used 10-20 electrode placement method. Fig. 5 shows the relative position between the electrodes. A band-pass filter with a frequency range of 0.16 Hz to 43 Hz was used to filter out noise in EEG data.

Particularly, we borrowed the RSEFNN model proposed in [43] to assess the fatigue level of the subjects. The RSEFNN used all the EEG data from the driving simulation as the input training data. To label the EEG data, we divided the collected EEG data into segments every 3 seconds. Since the sampling frequency of EEG data is 128 Hz, each segment was composed of EEG signals of 384 samples (128 × 3). Note that each segment of the EEG data from the sleep deprivation group was labeled as fatigue, however, that from the non-deprived sleep group was non-fatigue. The RSEFNN model has executed a total of 110 iterative training. The objective assessment value is defined as the output of the RESFNN, which is a probability value and its ranges from 0 to 1.

Second, during the self-assessment, according to his/her own current physical, physiological and psychological situation, every subject adopted a questionnaire (see Appendix 2) to evaluate his/her own fatigue state. In other words, for each subject, each question in the questionnaire was queried by dialogue. Each subject's self-assessment was used as the fatigue level assessment on a scale of 1-10 with 10 being most fatigued. When the score is 9 or more, the subject is considered to be in a fatigue driving state. Here, we consider using a probability value to describe the subject's self-assessed fatigue level. Note that each self-assessed value is based on a scale of 1-10 with its range between 1 and 10. Therefore, each value is divided by 10 to implement its normalization with its range between 0 and 1. In other words, the purpose of dividing each value by 10 is to change the value to a probability value between 0 and 1.

Particularly, for each subject, we conducted an objective assessment every five minutes and then did a self-assessment process based on the questionnaire. Finally, by combining the objective assessment value and self-assessment value, the final KSS value can be expressed as the final KSS value = 0.5× objective assessment value + 0.5× self-assessment value. That is, for each final KSS value, the self-assessment score and the objective assessment score each accounted for half. Each final KSS value has ten anchored levels. When the final KSS value >= 0.9, the driver is considered to be in the state of fatigue driving.

In addition, the facial video of each subject was recorded for three hours during the driving simulation period. The fatigue features were extracted from each subject's facial video. More precisely, each set of features can be extracted

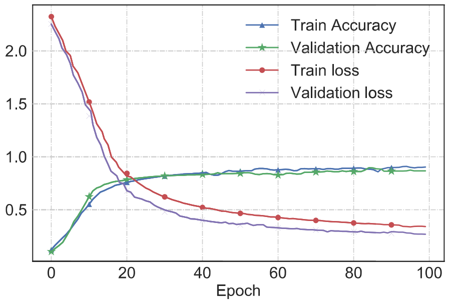


Fig. 6. Number of epochs on the horizontal axis of the coordinate refers to the number of iteration training in the MFRNN with the training and validation data set. When the MFRNN is trained 110 times, both its loss function value and accuracy tend to stabilize on the train and validation set. The accuracy and loss values of the training set and the validation set are consistent, indicating that there is no overfitting in the MFRNN.

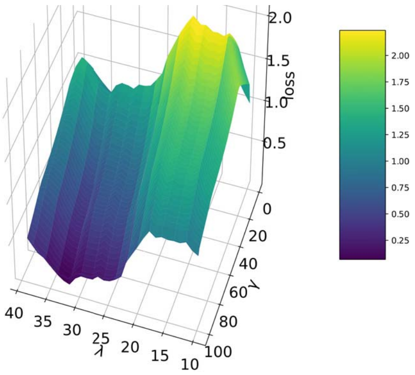


Fig. 7. Different values of λ and γ correspond to different loss function values.

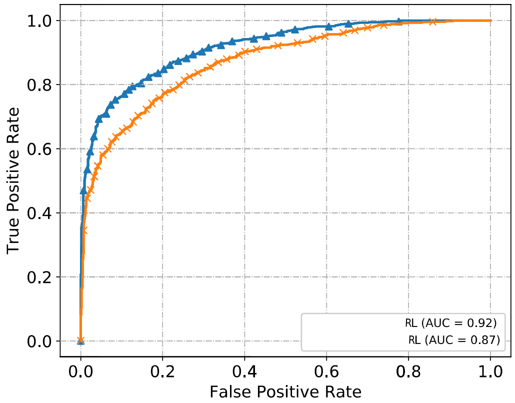


Fig. 8. ROC curves with and without a multilayer relationship layer (RL).

from every 3 seconds for each subject's facial video. Each set of features contained three types of features, namely, the heart rate, eye openness level and mouth openness level. Since a final KSS value is obtained every five minutes, the label of the features in the five minutes is set to this KSS value. For each subject, a total of 30 final KSS values was recorded over the 3-hour driving simulation. There are a total of 20 subjects. Therefore, we can extract a total of 60,000 sets of features (30 KSS×5 minutes ×60 seconds ×20subjects/ 3 seconds), including 21862 non-fatigue driving feature sets and 38138 fatigue driving feature sets. The extracted fatigue features can be used to train and evaluate the MFRNN model based on the simulation experiment. The final KSS values were used as the target values (Fatigue Feature Label), that is, the proposed MFRNN model outputs the probability of fatigue.

B. Training and Evaluation of the Proposed MFRNN

To train the MFRNN and achieve the optimal parameters, we split the fatigue-driving feature dataset into three sub-datasets: a training set (60%), a validation set (10%), and a test set (30%). The three sub-datasets played an important role in avoiding data leakage between the different phases. Similar to references [46] and [47], we only use the training set data for iterative training, which aims to achieve the effective model parameters of the proposed MFRNN. According to evaluation criteria including validation loss (see equation 13) and validation accuracy, the validation set was then used to evaluate the performance of the proposed MFRNN under different model parameters, which can determine the optimal parameters of the proposed MFRNN and avoid its overfitting. Note that the validation set can limit the number of iterations of the training set. Finally, the test set was used to verify the accuracy of the proposed MFRNN model. Particularly, the test set data set does not participate in model training.

An experiment was performed to verify the effectiveness of the above training process of MFRNN. Fig. 6 shows that the accuracy and loss values of MFRNN both on the training set and on the validation set are consistent, indicating that there is no overfitting in the MFRNN. More specifically, if the proposed MFRNN performs well on both the training set and the validation set, this indicates that it does not exhibit overfitting. If the proposed MFRNN performs well in the training set and does not perform well on the validation set, this indicates the overfitting. Therefore, the above process can not only ensure the accuracy of the MFRNN but also ensure that it does not appear overfitting.

Furthermore, the optimal lengths of the fatigue-driving feature sequences E and X were determined by minimizing the loss function. Specifically, different lengths of fatigue feature sequences correspond to the different values of the loss function. The segment length of the feature sequences with the lowest loss function value is optimal. The input sequence length of the heart rate statistic feature is γ . Note that the input sequence of the eye openness level and mouth openness level are the same length λ because they used the same subneural-network.

Fig. 7 shows that different values of λ and γ correspond to different loss function values; when λ is unchanged and γ increases, the value of the loss function gradually decreases. However, when γ is constant and λ increases, the loss value still tends to decrease overall. As the sequence length of the heart rate statistical feature γ increases within a certain range, the value of the loss function gradually decreases. This suggests that the statistical characteristics of the heart rate reflect fatigue. In particular, when $\gamma = 97$ and $\lambda = 29$, the loss value is minimized.

In addition, Fig. 8 shows two receiver operating characteristic (ROC) curves for the MFRNN with and without the

TABLE I

PERFORMANCE COMPARISON OF THE MFRNN BASED ON FEATURE COMBINATIONS FROM DIFFERENT SOURCES

Feature selection	Accuracy	$\frac{\text{False Negative}}{\text{True Negative + False Negative}}$
Heart rate(HR)	92.3%	3.4%
Eye openness level	90.1%	3.9%
Eye openness level+ Mouth openness level	91.1%	3.7%
HR+Eye+Mouth	94.74%	2.2%

False Negative is the number of fatigue samples misidentified as non-fatigue samples; True Negative is the number of fatigue state; **True Negative + False Negative** is the total number of samples whose model output is non-fatigue.

TABLE II

THE FEATURES AND CLASSIFIERS USED IN THE COMPARISON METHODS

Classifier	Feature
Spectral regression [11]	Eye openness level
PCA+LBP+SVM [13]	Eye openness level
Two-level SVM [19]	Eye openness level+the Mouth openness level+HR

relationship layer (RL) among the three features: heart rate and eye and mouth openness. Note that a higher area under the curve (AUC) indicates a higher the ability of the corresponding neural network to distinguish the fatigue-driving state. Interestingly, it can be seen from Fig. 8 that the AUC increases to 0.92 from 0.87 by introducing RL into MFRNN. This indicates that RL can effectively improve the fatigue recognition performance of MFRNN. This may be due to temporal or quantitative relationships among the heart rate, and eye and mouth openness linking these three features with fatigue identification. Because the MFRNN can learn from the relationships among the three features by the relationship layer (Subneural-network-3), the performance of the MFRNN to distinguish fatigue or non-fatigue is substantially improved.

Finally, we verified the impact of using different fatigue features or their combinations on the fatigue recognition performance of the MFRNN. Table I shows that the utilization of the heart rate feature can enhance the reliability of the MFRNN compared with only using eye openness or the combination of eye and mouth openness. More interestingly, Table I shows that the combination of heart rate and eye and mouth openness level can better improve the performance of the MFRNN relative to the other three cases. This implies that the multiple source fuse between feature information can effectively improve the accuracy of fatigue-driving recognition.

C. Performance Evaluation of Different Methods in Cross-Subject Situations

To verify the performance of the proposed MFRNN, it was compared with three methods including spectral regression (SR) [11], PCA+LBP+SVM [13], and Two-level SVM [19] under cross-subject testing by conducting an experiment in which one subject was left out. Reference [11] adopted SR to estimate the continuous level of eye openness; the eye was detected by Opencv eye detector; the fatigue score was measured by PERCLOS. In [13], the eyes were detected either using principal component analysis (PCA) during daytime or using the block local-binary-pattern (LBP) features during nighttime; from the eye detection, the support vector machine (SVM) divided the eye states into the open or closed state; fatigue score was also measured by PERCLOS. Reference [19] used two-level with the combination of the eye openness level, the Mouth openness level, and heart rate to detect the fatigue state. These methods are briefly described in Table II.

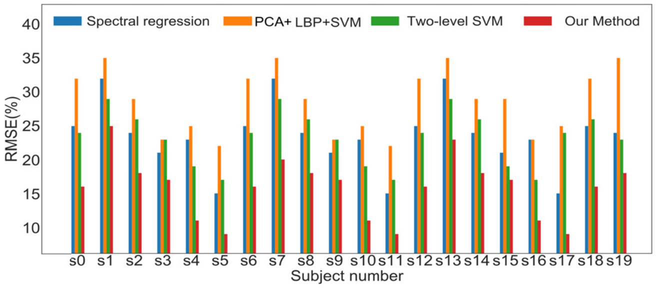


Fig. 9. The performance of each method was compared when one subject was used as the test set and the other subjects were used to construct the training sets. “s0” refers to an experimenter with the number “0”.

For each compared method, the data for one subject was selected as the test set, whereas the data from the remaining subjects were employed to construct the training data set. Suppose we take the data of subject “0” as the testing data, denoted “s0”. The data from the rest of the subjects were then used as the training sets to train each model. We used root mean square error (RMSE) as a performance metric to assess the stability of the compared methods.

Fig. 9 shows the boxplots of the test RMSEs for all 20 subjects using the different methods. Fig. 10 shows that the test RMSEs of SR, PCA+LBP+SVM, Two-level SVM, and the proposed MFRNN were 0.235 ± 0.085, 0.285 ± 0.065, 0.23 ± 0.06, and 0.17 ± 0.08, respectively. Interestingly, the proposed MFRNN is more stable than the other three methods under cross-subject testing. This may be because MFRNN contains different subneural-networks for data from different feature sources. More specifically, MFRNN effectively integrates RNN, fuzzy reasoning, and a relationship layer to extract temporal information and fuzzy information features from multiple information sources, and establishes the relationships among the features to improve the fatigue-driving recognition.

D. Performance Evaluation of Different Methods Under Different Conditions

To verify the fusion performance of each method considering noise, we increased the noise in the features by having volunteers wear glasses and use bright light. We conducted

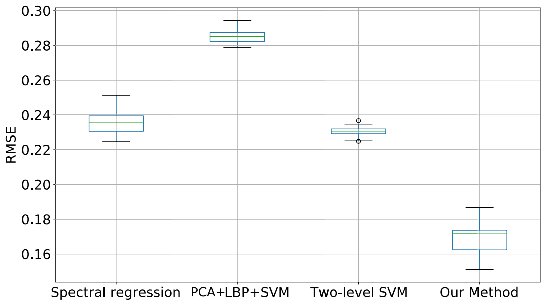
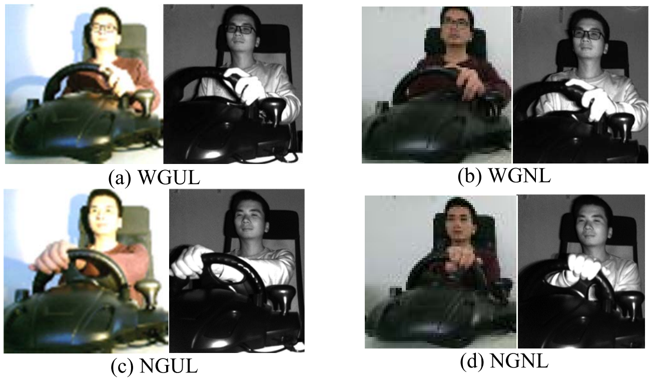


Fig. 10. Boxplots of the test RMSEs for four compared models on cross-subject testing.

TABLE III
ABBREVIATED LIST OF FOUR EXPERIMENTAL CONDITIONS

Acronym	Meaning
WGUL	Wearing glasses and using bright light.
WGNL	Wearing glasses and not using bright light.
NGUL	Not wearing glasses and using bright light.
NGNL	Not wearing glasses and not using bright light.


(a) WGUL

(b) WGNL

(c) NGUL

(d) NGNL

Fig. 11. RGB images (left) and infrared images (right) captured by RGB-D camera under four experimental conditions.

TABLE IV
THE ACCURACY(%) OF SEVERAL METHODS UNDER FOUR EXPERIMENTAL CONDITIONS

Methods	Experiment condition			
	WGUL	**WGNL**	**NGUL**	**NGNL**
Spectral regression [11]	78.89	80.95	79.40	84.21
PCA+LBP+SVM [13]	62.63	67.89	68.42	70.68
Two-level SVM [19]	79.63	82.16	80.68	87.21
Our Method	**83.95**	**89.21**	**85.47**	**94.74**

comparisons between the proposed MFRNN model, SR [11], PCA+LBP+SVM [13], and Two-level SVM [19] on four test conditions: WGUL, WGNL, NGUL, and NGNL. TABLE III lists the full names of the four conditions that denote wearing glasses and using bright light to illuminate the face (WGUL), wearing glasses and not using bright light (WGNL), not wearing glasses and using bright light (NGUL), and not wearing glasses and not using bright light (NGNL). Fig. 11 clearly shows the four different conditions.

It can be seen from Table IV that each compared method has the best fatigue recognition accuracy under the NGNL of the four test conditions. This is because NGNL has no glasses and bright light that interfere with the fatigue features that were collected by the RGB-D camera. However, under conditions WGUL, WGNL, and NGUL, each compared method has significant performance attenuation because glasses or bright light can affect the fatigue features extracted by the RGB-D camera. Particularly, the proposed method, SR [11], PCA+LBP+SVM [13], and Two-level SVM [19] have the lowest recognition accuracy under the WGUL of the four conditions because both glasses and bright light can affect the performance of the RGB-D camera.

Furthermore, the proposed method, SR [11], and Two-level SVM [19] under WGNL have better recognition accuracy than those under NGUL, however, PCA+LBP+SVM [13] has worse accuracy. The reasons behind this are explained as follows. Note that heart rate features are extracted through the frequency of color changes of the face image so that exposure to light on the face can make more noise into the heart rate features. Both the proposed method and Two-level SVM [19] used the combination of the heart rate features, eye openness level, and the mouth openness level. Therefore, these two methods greatly affected by bright light. Different from the previous two methods, SR [11] first matches the current picture with images of different light intensities in the training set and then eliminates the effects of light based on similar light intensity images in the training set. Therefore, the recognition accuracy of SR [11] is obviously affected when the light intensity of the face during the test is different from the light intensity of the training data. PCA+LBP+SVM [13] uses both PCA and LBP to extract eye features. Particularly, PCA can still extract facial features under the influence of bright light. However, this method is susceptible to obstructions such as wearing glasses.

Furthermore, an interesting observation from III is that the proposed MFRNN consistently provides the best fatigue recognition accuracy: 83.95%, 89.21%, 85.47%, and 94.74% among all compared methods under the four different conditions, respectively. Some reasons behind this are given as follows. First, the proposed method considers both the temporal information of fatigue features and the relationship between the features. Simultaneously, the proposed MFRNN can provide the best fusion performance of the three fatigue features relative to the other three methods. This also further demonstrates such a fact that there is an effective combination of RNN, fuzzy reasoning, and the relationship layer in MFRNN.

E. Performance Evaluation of Drowsy Driving Dataset

In order to further verify the performance of the proposed method, we compared the proposed MFRNN with Spectral regression (SR) [11], PCA+LBP+SVM [13], and Two-level SVM [19] on another existing public drowsy driving dataset [33]. This data set consists of the subjects' face RGB video data from simulated driving. In the data set,

TABLE V
THE PERFORMANCE OF SEVERAL METHODS ON DROWSY DRIVING DATASET

Methods	Accuracy (%)
Spectral regression [11]	80.13
PCA+LBP+SVM [13]	67.89
Two-level SVM [19]	79.23
Our Method	**86.75**

there were 23 subjects; their age ranges from 26 to 55 years old; male and female ratios are 55% and 52%, respectively. Each subject provided 30 minutes of facial video. The collected video resolution is 720 × 1280 per frame. Subjects' fatigue levels were quantified by filling out questionnaires. Note that the dataset consists of an RGB video. To obtain the eye openness level and mouth openness level of the video data set, we use a method from [34]. To obtain the independent quantity matrix of the three channels of RGB, the face video was analyzed by independent quantity using the joint approximate diagonalization of the eigen matrices (JADE) algorithm. Then, a Fourier transform was used to extract the image change frequency to obtain heart rate (HR).

The performance of several methods on the drowsy driving dataset is illustrated in Table V. It can be seen from Table V that the accuracy of the proposed MFRNN model is higher than those of other methods. This suggests that the proposed method can identify fatigue more accurately.

F. Field Experiment

To evaluate the fatigue recognition performance of the proposed method, we compared it with two-level SVM [19] and long short-term memory (LSTM) [41] in a field environment. Since sequence signals are usually processed using LSTM, we also use LSTM as a baseline for comparison. Note that LSTM uses 30 recurrent units. Note that all the compared methods fuse the three fatigue features: heart rate, eye openness level, and mouth openness level. In this experiment, one subject did the actual driving and drove for a total of two and a half hours. Fig. 12(a) shows our driving experiment route on Google maps.

A total of 24 questionnaire evaluations (see Appendix) were made during the field experiment. Via such evaluations for the subject, we can obtain 24 fatigue measurement values, their ranges from 0 to 1. The measurement values are considered as the reference value of the fatigue probability for comparing different methods. Subsequently, we use three different methods to output their corresponding fatigue prediction values based on the features extracted by the facial video. Figs. 12(b) and 12(c) show the captured facial image and the RGB-D camera for capturing the facial image, respectively. Those fatigue features involving heart rate, eye openness level, and mouth openness level can be extracted using the same extraction method as before. The subject is seen as fatigue when the output of these methods is greater than or equal to 0.9. If the prediction value of each method for each time is consistent with the corresponding evaluation value of the questionnaire, we believe that the method's prediction is correct this time. The accuracy rate of each compared method is equal to C/T, where C and T indicate the number of correct predictions and the total number of predictions, respectively.

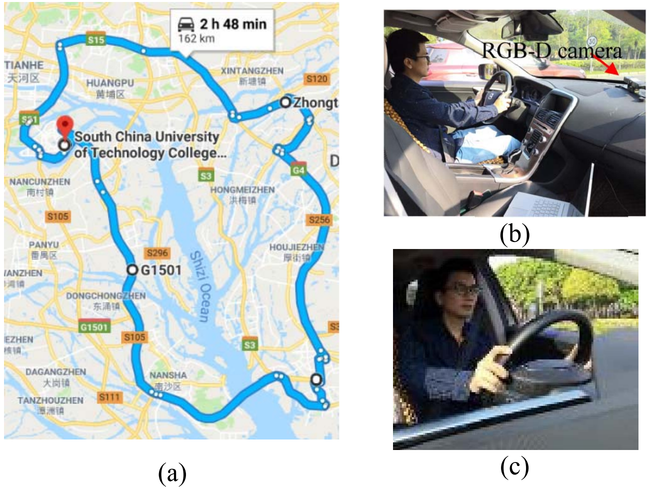


Fig. 12. (a) Experiment route shown on Google maps. (b) The position of the RGB-D camera. (c) RGB image of the driver's face captured by the RGB-D camera.

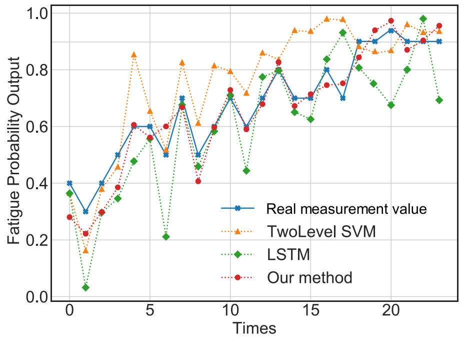


Fig. 13. Comparison of performances in the real road test.

TABLE VI
COMPARISON OF ACCURACY IN THE REAL ROAD TEST

Methods	Accuracy (%)
Two-level SVM [19]	70.83
LSTM [41]	75.00
Our Method	91.67

The experimental results are given in Fig. 13, where the 24 real measurements are shown as a solid blue line. The Two-Level SVM generally provides the 24 higher prediction values compared with the real measurements. Therefore, the Two-level SVM easily misjudge a non-fatigue state as a fatigue state. The 24 prediction values of the LSTM have a large fluctuation range, which makes the LSTM impossible to accurately and stably recognize the driver's fatigue state when the driver is in a fatigue state. An interesting observation from Fig. 13 that the output prediction values of the proposed method can better match the real measured values compared with LSTM and Two-level SVM. Therefore, the proposed method can better predict the fatigue level of the driver.

Table VI shows the fatigue prediction accuracy of the three compared methods in a field environment. The proposed MFRNN obtains the highest prediction accuracy among the three methods, further indicating that the proposed method can identify fatigue more accurately by integrating both the temporal information of fatigue features and the relationship between fatigue features for driver fatigue detection.

VI. Conclusion

In this paper, we propose a novel multimodal fusion recurrent neural network (MFRNN) to detect driver fatigue by integrating heart rate, eye openness level, and mouth openness level. A nonintrusive approach is used to estimate the statistical characteristics of the heart rate. We designed different time information extraction networks for different source features. Since it is found that a certain correlation exists between the characteristics and fatigue, an extraction neural network layer was designed to identify the relationships among the features. Experiment results show that the system has good robustness and accuracy under various conditions.

Nevertheless, the performance of the presented method is affected when the driver's head moves abruptly or the RGB-D camera shakes. In the future, we will be looking at this problem as well as the contextual features.

Appendix

Below is the content of the questionnaire for KSS self-assessment.

1) Are you fatigue now? If yes, to what degree you are feeling fatigued? (Scale: 1–10)
2) To what degree your fatigue may affect your ability to work? (Scale: 1–10)
3) To what degree you are feeling sleepy now? (Scale: 1–10)
4) To what degree you are feeling able to walk normally? (Scale: 1–10)
5) To what degree you are feeling energetic? (Scale: 1–10)
6) To what degree you are feeling able to concentrate? (Scale: 1–10)
7) To what degree you are feeling able to think clearly? (Scale: 1–10)
8) Chance of dozing:
    a) If allowed to lie down for rest (Scale: 1–10).
    b) If allowed to listen to music (Scale: 1–10).
    c) If allowed to drive in a long and monotonous road (Scale: 1–10).

References

[1] M. I. Chacon-Murguia and C. Prieto-Resendiz, "Detecting driver drowsiness: A survey of system designs and technology," *IEEE Consum. Electron. Mag.*, vol. 4, no. 4, pp. 107–119, Oct. 2015.

[2] S. Kaplan, M. A. Guvensan, A. G. Yavuz, and Y. Karalurt, "Driver behavior analysis for safe driving: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 3017–3032, Dec. 2015.

[3] Y. Liang, M. L. Reyes, and J. D. Lee, "Real-time detection of driver cognitive distraction using support vector machines," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 2, pp. 340–350, Jun. 2007.

[4] J. Wang, S. Zhu, and Y. Gong, "Driving safety monitoring using semisupervised learning on time series data," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 3, pp. 728–737, Sep. 2010.

[5] B.-F. Wu, Y.-H. Chen, C.-H. Yeh, and Y.-F. Li, "Reasoning-based framework for driving safety monitoring using driving event recognition," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 3, pp. 1231–1241, Sep. 2013.

[6] R. Fu and H. Wang, "Detection of driving fatigue by using noncontact EMG and ECG signals measurement system," *Int. J. Neural Syst.*, vol. 24, no. 03, May 2014, Art. no. 1450006.

[7] A. K. Kokonozi, E. M. Michail, I. C. Chouvarda, and N. M. Maglaveras, "A study of heart rate and brain system complexity and their interaction in sleep-deprived subjects," in *Proc. Comput. Cardiol.*, Sep. 2008, pp. 969–971.

[8] J. Vicente, P. Laguna, A. Bartra, and R. Bailón, "Drowsiness detection using heart rate variability," *Med. Biol. Eng. Comput.*, vol. 54, no. 6, pp. 927–937, Jun. 2016.

[9] C. Zhang, H. Wang, and R. Fu, "Automated detection of driver fatigue based on entropy and complexity measures," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 1, pp. 168–177, Feb. 2014.

[10] E. Vural, M. Cetin, A. Ercil, G. Littlewort, M. Bartlett, and J. Movellan, "Drowsy driver detection through facial movement analysis," in *Proc. ICCV Workshop HCI*, 2007, pp. 6–8.

[11] B. Mandal, L. Li, G. S. Wang, and J. Lin, "Towards detection of bus driver fatigue based on robust visual analysis of eye state," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 3, pp. 545–557, Mar. 2017.

[12] Q. Ji, Z. Zhu, and P. Lan, "Real-time nonintrusive monitoring and prediction of driver fatigue," *IEEE Trans. Veh. Technol.*, vol. 53, no. 4, pp. 1052–1068, Jul. 2004.

[13] A. Dasgupta, A. George, S. L. Happy, and A. Routray, "A vision-based system for monitoring the loss of attention in automotive drivers," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1825–1838, Dec. 2013.

[14] R. O. Mbouna, S. G. Kong, and M.-G. Chun, "Visual analysis of eye state and head pose for driver alertness monitoring," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 3, pp. 1462–1469, Sep. 2013.

[15] B. Cyganek and S. Gruszczyński, "Hybrid computer vision system for drivers' eye recognition and fatigue monitoring," *Neurocomputing*, vol. 126, pp. 78–94, Feb. 2014.

[16] G. M. Bhandari, A. Durge, A. Bidwai, and U. Aware, "Yawning analysis for driver drowsiness detection," *Int. J. Res. Eng. Technol.*, vol. 3, no. 2, pp. 502–505, Feb. 2014.

[17] X. Fan, B.-C. Yin, and Y.-F. Sun, "Yawning detection for monitoring driver fatigue," in *Proc. Int. Conf. IEEE Mach. Learn. Cybern.*, Aug. 2007, pp. 19–22.

[18] S. Abtahi, B. Hariri, and S. Shirmohammadi, "Driver drowsiness monitoring based on yawning detection," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf.*, Binjiang, China, May 2011, pp. 1–4.

[19] W. Sun, X. Zhang, S. Peeta, X. He, and Y. Li, "A real-time fatigue driving recognition method incorporating contextual features and two fusion levels," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 12, pp. 3408–3420, Dec. 2017.

[20] G. Yang, Y. Lin, and P. Bhattacharya, "A driver fatigue recognition model based on information fusion and dynamic Bayesian network," *Inf. Sci.*, vol. 180, no. 10, pp. 1942–1954, May 2010.

[21] B.-G. Lee and W.-Y. Chung, "Driver alertness monitoring using fusion of facial features and bio-signals," *IEEE Sensors J.*, vol. 12, no. 7, pp. 2416–2422, Jul. 2012.

[22] L. Zhao, Z. Wang, X. Wang, and Q. Liu, "Driver drowsiness detection using facial dynamic fusion information and a DBN," *IET Intell. Transp. Syst.*, vol. 12, no. 2, pp. 127–133, Mar. 2018.

[23] X. Li, J. Chen, G. Zhao, and M. Pietikainen, "Remote heart rate measurement from face videos under realistic situations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 4264–4271.

[24] L. Tarassenko, M. Villarroel, A. Guazzi, J. Jorge, D. A. Clifton, and C. Pugh, "Non-contact video-based vital sign monitoring using ambient light and auto-regressive models," *Physiol. Meas.*, vol. 35, no. 5, pp. 807–831, May 2014.

[25] C. S. Dhir and S.-Y. Lee, "Discriminant independent component analysis," *IEEE Trans. Neural Netw.*, vol. 22, no. 6, pp. 845–857, Jun. 2011.

[26] F. Palsson, M. O. Ulfarsson, and J. R. Sveinsson, "Sparse Gaussian noisy independent component analysis," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2014, pp. 4224–4228.

[27] K. Fujiwara *et al.*, "Heart rate variability-based driver drowsiness detection and its validation with EEG," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 6, pp. 1769–1778, Jun. 2019.

[28] M. Malik, "Task force of the European society of cardiology and the north American society of pacing and electrophysiology. Heart rate variability. Standards of measurement, physiological interpretation, and clinical use," *Eur Heart J.*, vol. 17, no. 3, pp. 354–381, 1996.

[29] A. J. Camm *et al.*, "Guidelines heart rate variability—Standards of measurement, physiological interpretation, and clinical use," *Eur. Heart J.*, vol. 115, no. 5, pp. 354–381, 1996.

[30] C. Papazov, T. K. Marks, and M. Jones, "Real-time 3D head pose and facial landmark estimation from depth images using triangular surface patch features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 4722–4730.

[31] A. Acıoğlu and E. Erçelebi, "Real time eye detection algorithm for PERCLOS calculation," in *Proc. 24th Signal Process. Commun. Appl. Conf. (SIU)*, Zonguldak, Turkey, 2016, pp. 1641–1644.

[32] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 1–41.

[33] A. Byrnes and C. Sturton, "On using drivers' eyes to predict accident-causing drowsiness levels," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 2092–2097.

[34] X. Dong, S.-I. Yu, X. Weng, S.-E. Wei, Y. Yang, and Y. Sheikh, "Supervision-by-registration: An unsupervised approach to improve the precision of facial landmark detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 360–368.

[35] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980

[36] H. Martensson, O. Keelan, and C. Ahlström, "Driver sleepiness classification based on physiological data and driving performance from real road driving," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 2, pp. 421–430, Feb. 2019.

[37] W. Sun, X. Zhang, S. Peeta, X. He, Y. Li, and S. Zhu, "A self adaptive dynamic recognition model for fatigue driving based on multisource information and two levels of fusion," *Sensors*, vol. 15, no. 9, pp. 24191–24213, Sep. 2015.

[38] M. Jirina, P. Bouchner, and S. Novotny, "Identification of driver's drowsiness using driving information and heart rate," *Neural Netw. World*, vol. 20, no. 6, pp. 773–791, 2010.

[39] J. Berg, G. Neely, U. Wiklund, and U. Landstrom, "Heart rate variability during sedentary work and sleep in normal and sleep-deprived states," *Clin. Physiol. Funct. Imag.*, vol. 25, no. 1, pp. 51–57, Jan. 2005.

[40] T. Åkerstedt and M. Gillberg, "Subjective and objective sleepiness in the active individual," *Int. J. Neurosci.*, vol. 52, nos. 1–2, pp. 29–37, Jan. 1990.

[41] W. Sun, X. Zhang, S. Peeta, X. He, Y. Li, and S. Zhu, "A self-adaptive dynamic recognition model for fatigue driving based on multi-source information and two levels of fusion," *Sensors*, vol. 15, no. 9, pp. 24191–24213, Sep. 2015.

[42] J. Krajewski, S. Schnieder, D. Sommer, A. Batliner, and B. Schuller, "Applying multiple classifiers and non-linear dynamics features for detecting sleepiness from speech," *Neurocomputing*, vol. 84, pp. 65–75, May 2012.

[43] Y.-T. Liu, Y.-Y. Lin, S.-L. Wu, C.-H. Chuang, and C.-T. Lin, "Brain dynamics in predicting driving fatigue using a recurrent self-evolving fuzzy neural network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 2, pp. 347–360, Feb. 2016.

[44] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, 2011. [Online]. Available: http://www.csie.ntu.edu.tw/ cjlin/libsvm

[45] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[46] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[47] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[48] N. Petrovsky *et al.*, "Sleep deprivation disrupts prepulse inhibition and induces psychosis-like symptoms in healthy humans," *J. Neurosci.*, vol. 34, no. 27, pp. 9134–9140, Jul. 2014.

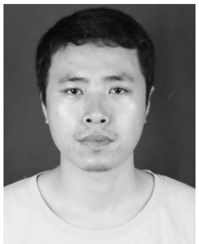
[49] J. T. Balkin, T. Rupp, D. Picchioni, and N. J. Wesensten, "Sleep loss and sleepiness: Current issues," *Chest*, vol. 134, no. 3, pp. 653–660, 2008.

[50] D. F. Dinges *et al.*, "Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4-5 hours per night," *Sleep*, vol. 20, no. 4, pp. 267–277, 1997.

[51] J. P. Nilsson *et al.*, "Less effective executive functioning after one night's sleep deprivation," *J. Sleep Res.*, vol. 14, no. 1, pp. 1–6, Mar. 2005.

Guanglong Du received the Ph.D. degree in computer application technology from the South China University of Technology, Guangzhou, China, in 2013. He is currently an Associate Professor with the School of Computer Science and Engineering, South China University of Technology. His research interests include intelligent robotics, human–computer interaction, artificial intelligence, and machine vision.

Tao Li received the B.S. degree in software engineering from the Wuhan University of Technology, Wuhan, China, in 2018. He is currently pursuing the Ph.D. degree in computer science and technology with the School of Computer Science and Engineering, South China University of Technology, Guangzhou, China. His research interests include the analysis of physiological electrical signals and human–robot interaction.



Chunquan Li received the B.Sc., M.Sc., and Ph.D. degrees from Nanchang University, Nanchang, China, in 2002, 2007, and 2015, respectively.

Since 2002, he has been with the School of Information Engineering, Nanchang University, where he is currently a Professor and a Young Scholar of Ganjiang River. He is also a Visiting Professor with the Department of Systems and Computer Engineering, Carleton University, Ottawa, ON, Canada. He has published over 30 research articles. His current research interests include computing intelligence, haptics, virtual surgery simulation, robotics, and their applications to biomedical engineering.

Peter X. Liu (Fellow, IEEE) received the B.Sc. and M.Sc. degrees from Northern Jiaotong University, China, in 1992 and 1995, respectively, and the Ph.D. degree from the University of Alberta, Canada, in 2002. Since 2002, he has been with the Department of Systems and Computer Engineering, Carleton University, Canada, where he is currently a Professor and the Canada Research Chair. Also, he has an appointment with the School of Information Engineering, Nanchang University, as an Adjunct Professor. He has published more than 280 research articles. His research interests include interactive networked systems and teleoperation, haptics, micromanipulation, robotics, intelligent systems, context-aware intelligent networks, and their applications to biomedical engineering. He is also a Licensed Member of the Professional Engineers of Ontario (P.Eng.) and a fellow of the Engineering Institute of Canada. He was a recipient of the 2007 Carleton Research Achievement Award, the 2006 Province of Ontario Early Researcher Award, the 2006 Carty Research Fellowship, the Best Conference Paper Award from the 2006 IEEE International Conference on Mechatronics and Automation, and the 2003 Province of Ontario Distinguished Researcher Award. He serves as an Associate Editor for several journals, including the IEEE TRANSACTIONS ON CYBERNETICS, the IEEE/ASME TRANSACTIONS ON MECHATRONICS, the IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING, and IEEE ACCESS.

Di Li received the B.Sc. and M.Sc. degrees from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 1985 and 1988, respectively, and the Ph.D. degree in control from the South China University of Technology, Guangzhou, China, in 1993. She is currently a Professor with the School of Mechanical and Automotive Engineering and the Head of the Institute of Optical, Mechanical, and Electronic Integration, South China University of Technology. Her research interests include control systems, embedded systems, computerized numerical control systems, and machine vision.