



Recognizing unsafe behaviors of workers by frequency domain features of facial motion information

Ming Zhu¹ · Guohui Li¹ · Qin Huang¹

Received: 15 December 2021 / Revised: 20 February 2023 / Accepted: 6 June 2023 /
Published online: 14 June 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Extracting time domain features of facial motion information to recognize unsafe driving behaviors such as fatigue and distraction is helpful to reduce traffic accidents. However, it is uncertain whether facial motion can recognize unsafe behaviors of workers. And whether it can improve recognition accuracy by introducing frequency domain features needs further research. This paper proposes a recognition method of workers' unsafe behavior based on frequency domain features of facial motion information. Firstly, the facial video of workers is obtained. And Gabor, histogram of oriented gradient (HOG) and local binary patterns (LBP) motion information are extracted. Then frequency domain features of these information are calculated, and finally sent to machine learning classifiers to recognize unsafe behavior. The results show that: (1) Compared with safe behavior, the complexity of facial motions during unsafe behavior is higher, especially in nose and mouth areas. (2) Wavelet entropy frequency domain features of Gabor motion information can better describe the complexity and have higher recognition accuracy (AUC=0.766); (3) The proposed method can recognize unsafe behaviors of workers and is effective for operation errors (AUC=0.818). The results can be used as a new idea to recognize unsafe behaviors of workers and provide technical support for on-site safety management.

Keywords Unsafe behavior · Behavior recognition · Facial motion · Frequency domain feature · Computer vision

1 Introduction

Unsafe behavior of workers is the main reason for accidents, resulting in a large number of casualties and property losses. Heinrich investigated 75,000 industrial accidents and found that more than 88% of accidents were caused by unsafe behavior [19]. Shin [31] counted 147

✉ Guohui Li
guohuili@sicnu.edu.cn

Ming Zhu
mingzhusicnu@hotmail.com

Qin Huang
sicnuhuangqin@hotmail.com

¹ School of Engineering, Sichuan Normal University, Chengdu 610101, China

major industrial accidents and found that 85% of the accidents were caused by unsafe behaviors. And these unsafe behaviors are mainly operation errors such as violation of safety work permit and safety operation procedure violation/failure. How to reduce accidents caused by unsafe behavior has become an important issue of on-site safety management.

Using computer vision technologies to recognize unsafe behaviors is helpful to reduce such accidents. Researchers use computer vision technologies to collect motion information of face, eyes and joint data, and calculate features of motion information (motion features) to recognize behavior. The result of this method is objective and the process is non-invasive. It is expected to replace safety observers as an important means of on-site safety management in the future [39]. Among them, facial motions are the nonverbal expression of human have the greatest variability [21], which widely used in the field of driving safety to recognize fatigue and distraction. Different from driving operation, the environment and operation types faced by workers in the production process are more complex. And compared with fatigue and distraction, operation error is a more frequent unsafe behavior of workers. If facial motions can be used to recognize unsafe behaviors of workers, it can provide technical support for on-site safety management.

Selecting reasonable motion features to describe facial motion information is important for recognition [36]. When using facial motion recognition behavior, researchers extract motion information of different facial regions such as eyes and mouth. And time domain features (such as mean and standard deviation) are calculated to describe motion information. They focus on finding motion information that is conducive to recognition, ignoring the influence of motion features selection on recognition accuracy. In addition to time domain features, frequency domain features can also describe motion information. And frequency domain features are considered to be suitable for analyzing complex and changeable biological signals [12].

When dealing with various operation conditions, worker's face will intentionally or unconsciously generate micro expression and other motions [32]. Coupled with the great variability of facial motions, complexity can be considered to describe facial motion information. Complexity can be defined as the difficulty of describing or predicting a signal. It is often used to analyze biological signals such as functional magnetic resonance imaging and electrocardiogram [3, 15]. Complexity can be divided into time domain complexity and frequency domain complexity. Time domain complexity can be measured by features such as first-order autocorrelation coefficient (FOA) and Hurst exponent (HE). Frequency domain complexity can be measured by features such as wavelet entropy. This paper attempts to use the frequency domain complexity to describe facial motion information in order to improve the recognition accuracy.

To reveal relationships between facial motions and unsafe behaviors, find facial motion features with higher recognition accuracy and use these to recognize unsafe behaviors. The contributions of this paper are as follows: (1) Describing facial motions by frequency domain complexity, and finding that facial motions during unsafe behavior are more complex; (2) Through experiments verification, finding that the recognition accuracy of frequency domain features is higher than that of time domain features; (3) Proposing an unsafe behavior recognition method based on frequency domain complexity, which can recognize operation errors, fatigue and distraction. The innovations of this paper: (1) A new clue related to unsafe behavior recognition is found, namely complexity of facial motions; (2) Frequency domain features of facial motion information are introduced, which improves the recognition accuracy compared with the time domain feature; (3) The proposed method expands the application range of facial motions in unsafe behavior recognition, especially can recognize operation errors recognized difficultly and happened frequently in work.

The remainder of this paper is structured as follows: In Section 2, we introduce research status of unsafe behavior recognition and face-based behavior recognition. In Section 3,

we describe the overall architecture of the proposed method. In Section 4, we introduce experiments designed to evaluate performance of the method and describe the experimental results. In Section 5, we give the conclusion and clarify the next work.

2 Related works

2.1 Recognition of unsafe behavior

It is believed that methods of unsafe behavior recognition can be divided into the following 3 types according to the means of obtaining motion information: observation-based methods, wearable equipment-based methods, and vision-based methods. The means of obtaining behavior information is one of the keys to unsafe behavior recognition, which will affect the recognition accuracy and application scope.

In an observation-based method, safety observers use on-site observation to obtain motion information of workers. And the motion information is qualitatively compared with the unsafe behavior in the behavior checklist to recognize. Zhang et al. [38] developed a list of 20 typical unsafe behaviors according to safety management manual provided by construction contractors. Eighteen safety observers were recruited to collect motion information of workers through on-site observation to complete the recognition. Guo et al. [17] and Yue et al. [37] used similar methods to recognize unsafe behaviors. The former designed a behavior checklist containing 9 primary elements and 49 secondary elements for tunnel construction. The latter designed a behavior checklist containing 5 primary elements and 30 secondary elements for coal mining. The methods are simple and convenient and are widely used in on-site safety management at present. But the methods have high labor costs and are difficult to cover the whole workplace and workers [35]. And more importantly, recognition results of the methods are obtained by safety observers qualitatively, which are easily affected by subjective factors such as personal ability and distraction.

In a wearable equipment-based method, researchers use wearable devices (such as acceleration sensors, pressure sensors and electroencephalographs) to obtain motion information (such as ankle acceleration, plantar pressure, electroencephalogram (EEG)) of workers. And a model between motion information and unsafe behavior is built by means of statistical analysis to recognize unsafe behavior. Jebelli et al. [22] used a tri-axial accelerometer to extract curves of ankle accelerations changing with time when subjects walk through an I-beam. Then maximum Lyapunov exponent of each curve is calculated to recognize unsafe behaviors such as carrying a load and faster speed walking. Antwi-Afari et al. [4] used a wearable insole pressure system to extract 26 plantar pressure data of subjects and got curves of the data changing within 1~4 seconds. Loss of balance behavior was recognized by calculating 5 statistic parameters of each curve, such as mean pressure, peak pressure and pressure-time integral. Wang et al. [33] used a wireless electroencephalograph to extract EEG of workers. And 30 vigilance ratio indices of the EEG were calculated to recognize non-vigilance state. The methods can obtain quantitative models between motion information and behavior types. But wearable equipment is easy to affect operation of workers. And when the number of workers is large, time and economic cost of installing wearable equipment will increase greatly [39].

In a vision-based method, researchers use visual sensors and computer vision algorithms to obtain motion information of face, eye and joint. And a model between motion information and unsafe behavior is built by means of a machine learning classifier or statistical

analysis to recognize unsafe behavior. Jeelani et al. [23] used a remote eye tracker to extract fixation positions of the subjects, and calculated fixation duration, fixation count and other data on each area of interest. Then subjects with high hazard recognition performance were found to have more fixation duration and fixation count. Han et al. [18] used a Kinect camera to extract rotation angles and 3D locations of 12 joints and 21 joint angles of workers. Then the data was mapped to 3-dimensional space by gaussian process dynamics model. Finally, unsafe behaviors such as climbing with articles, backward-facing climbing and climbing on the side away from the ladder were recognized by support vector machine (SVM). Chen et al. [10] used a monocular camera to extract 2D locations of 17 joints of workers. And the motion phase feature of joint data was found to be able to recognize climbing without three-point contact.

With its objective results and non-invasive process [39], the vision-based method is expected to become an important means of on-site safety management in the future. In particular, starting from the inner mechanism of behavior formation, recognizing unsafe behaviors by motion information of face [24] and eyes [5], which is roadworthy of exploration.

2.2 Recognizing behaviors by face

Face can provide rich nonverbal behavior clues. And it avoids disadvantages that the eyes can only be used for screen operation (remote eye trackers can only extract the gaze position on the matching screen), which is more suitable for on-site safety management. According to whether motion information is considered in recognition, methods of recognizing behaviors by face can be divided into image-based and video-based.

In the image-based method, researchers locate the facial area in the image to extract facial features, and use machine learning classifiers to recognize behaviors. Navarathna et al. [29] and Zhang et al. [40] recognize audience and student engagement by facial images. The former extracts HOG features of facial images and uses the random forest classifier for recognition. The latter extracts facial features through local gray code patterns (LGCP) descriptor and uses the fast sparse representation classifier for recognition. Li et al. [25] and Abdallah [1] extract facial features through VGG-16 network. The former uses SoftMax to directly recognize engagement and non-engagement. The latter uses SoftMax to recognize facial expressions to judge whether students understand the subject content. Hu et al. [20] designed a semi cascaded network (SCN) with lightweight structure to extract facial features and recognize driver distraction. The method is easy to implement and computationally small, but it ignores the facial motion information.

In the video-based method, researchers locate the facial area in the video and extract curve of the facial data changing with time to form motion information. Then motion features are calculated to describe motion information. Finally, machine learning classifiers are used to recognize behavior. Zhao et al. [42] extracted facial data from facial region through deep 3D convolution neural network and output sleepiness probability value of each video frame. Then the state probability vector of the video is obtained by cascading the sleepiness probability values of all frames. Finally, the recurrent neural network is used to recognize driver sleepiness with an accuracy of 0.886. Monkaresi et al. [28] extracted curves of 6 facial animation units changing with time. Then motion features (such as mean, median and standard deviation) of the curve were calculated. At the same time, local binary pattern (LBP-TOP) on 3 orthogonal planes is used to extract facial motion features of the video. Finally, the Updatable Naïve Bayes, Bayes Net, Logistic Regression and clustering

algorithms were used to recognize student engagement. Yuce et al. [36] extracted the curve of 14 facial action units and their relationships of the driver with time in 8.5 seconds. Then 4 motion features of the curve were calculated. Finally, the features are sent into SVM and random forest classifier to recognize distracted driving, with an accuracy of 0.681 for different subjects. In addition, researchers also recognize deception [6, 27], fatigue [13, 41], and distraction [16] through facial motion.

In short, researchers extract facial motion features through computer vision technologies and recognize fatigue, distraction, engagement and deception. However, there are still 2 problems in this kind of research. (1) During driving and learning, the types of environments and operations are single, which is difficult to be directly used in the field of production safety. Whether facial motions can recognize unsafe behaviors of workers remains to be verified. (2) Current researches rarely describe motion information from frequency domain and complexity. And they focus on finding motion information that facilitates recognition [36, 41], ignoring the influence of motion features selection.

3 Method

This paper proposes a recognition method of workers' unsafe behavior based on frequency domain features of facial motion information. The framework of the method is shown in Fig. 1. This method is based on the following assumption: workers have different facial motions when performing safe and unsafe behaviors.

The method consists of 3 parts: motion information extraction, motion feature calculation, and machine learning classification. Firstly, curves of facial data changing with time are extracted from facial video of workers to form facial motion information. Then motion features of the information are calculated from frequency domain. Finally, the features are sent into machine learning classifiers for training and classification to recognize unsafe behavior.

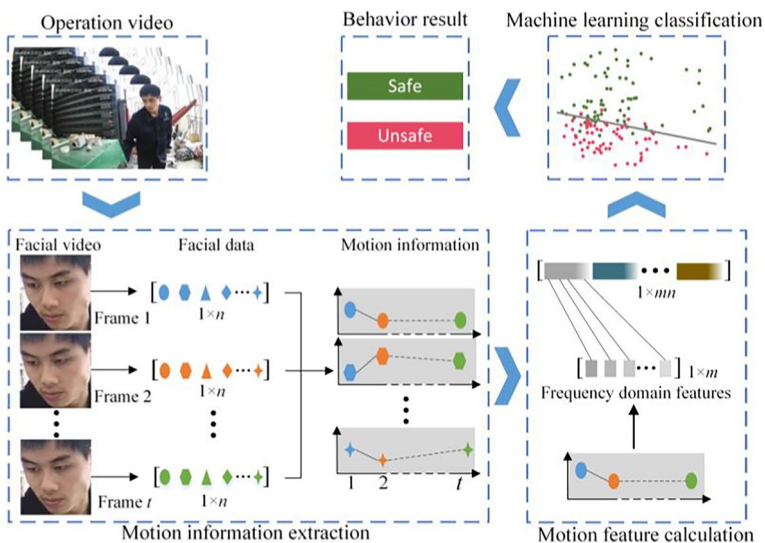


Fig. 1 Framework of recognizing workers' unsafe behavior

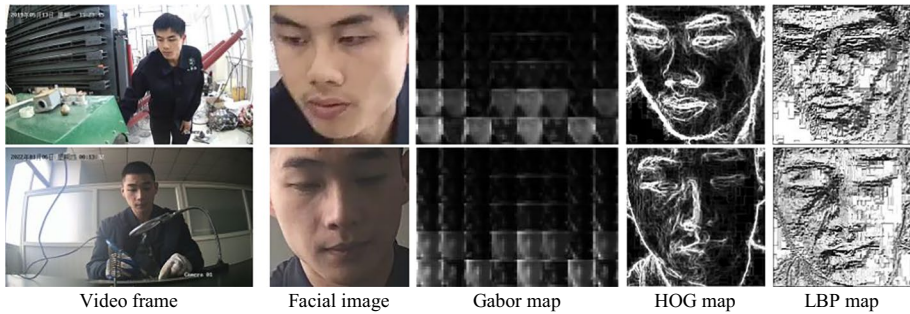


Fig. 2 Facial images and facial feature maps

3.1 Motion information extraction

Motion information extraction includes face location, facial data acquisition and motion information composition, as shown in Algorithm 1. Firstly, the face image X is intercepted from the operation video V by the face location algorithm. Then the facial data $D_{1 \times n}$ is extracted by feature operators. Finally, motion information $M_{n \times t}$ is constituted by the facial data of t frame images.

Algorithm 1 Motion information extraction

Input: Operation video $V = \{F_1, F_2, \dots, F_t\}$.

1: **for** $F_i \in V$ **do**

2: $X_i = \text{Face location}(F_i)$.

3: $D_i [d_1^i, d_2^i, \dots, d_n^i] = \text{Feature operator}(X_i)$.

4: **end for**

5: $M = [D_1^T, D_2^T, \dots, D_t^T]$.

Output: Motion information $M_{n \times t}$.

Face++ artificial intelligence open platform (<https://www.faceplusplus.com>) is used to locate face. Coordinates of face area obtained by the platform are used to intercept facial images from video frames, as shown in Fig. 2.

Gabor operator is used to extract facial motion information. It can describe structure of facial images under different scales, localizations and orientations [26]. Facial images are transformed into feature maps by the operator, as shown in Fig. 2. Then the feature map is transformed into a vector as facial data by down-sampling and frequency statistics. To study the influence of facial motion information on recognition accuracy, this paper also extracts HOG, LBP and combination of the three (Gabor, HOG and LBP) motion information. HOG operator can describe edge strength and gradient direction of facial images [11], namely facial contour. LBP operator can describe local contrast of facial images [30], namely facial texture.

Four types of facial data are extracted by using 3 operators, namely Gabor data, HOG data, LBP data and combined data. Forty (5 scales \times 8 directions) feature maps obtained by Gabor filtering are down sampled as 11×10 size. The pixel values of these feature maps are concatenated to obtain Gabor data with 1×4400 . The HOG feature map is divided into 4×4 cells, and using the block composed of 2×2 cells scans the image in

steps of one cell. Each block is extracted 36 bin values, and these bin values are concatenated to obtain HOG data with 1×324 . Frequency of each texture unit with a value of 0~255 in the LBP feature map is counted. And these frequencies are concatenated to obtain LBP data with 1×256 . Gabor, HOG and LBP data are concatenated to obtain combined data with 1×4980 . Then, curves of each facial data value changing with time are extracted to constitute 4 motion information.

3.2 Motion feature calculation

Frequency domain features of facial motion information are calculated to recognize unsafe behaviors. The process includes motion feature calculation and feature dimension reduction, as shown in Algorithm 2. Firstly, m frequency domain features are calculated for each row of motion information $M_{n \times t}$. Then the frequency domain features of each row are connected in series into the motion feature $W_{1 \times mn}$. Finally, the feature dimension of $W_{1 \times mn}$ is reduced to obtain W' for subsequent machine learning classification.

Algorithm 2 Motion feature calculation

Input: Motion information $M = [m_1, m_2, \dots, m_n]^T$.
 1: **for** $m_i \in M$ **do**
 2: $w_i [w^i_1, w^i_2, \dots, w^i_m] = \text{Feature calculation}(m_i)$.
 3: **end for**
 4: $W_{1 \times mn} = [w_1, w_2, \dots, w_n]$.
 5: $W' = \text{Dimension reduction}(W_{1 \times mn})$.
Output: Motion feature W' .

Frequency domain features represent frequency structure of the facial motion information. Wavelet packet transform is used to calculate frequency domain features. It can measure the frequency components of a signal and the time when these components appear, which is suitable for analyzing nonstationary signals. In this paper, 10 frequency domain features are calculated, including wavelet energy entropy (WEE), 8 wavelet scale entropy (WSE) and wavelet singular spectrum entropy (WSSE). These wavelet entropies can represent frequency domain complexity of facial motions.

To study the influence of motion features on recognition accuracy, time domain features and combination of the two (time and frequency domain) are also calculated. Time domain features represent the change law of facial information. FOA and HE are used to represent time domain complexity. In addition, other 6 time domain features are also calculated to discrete concentrated trend, discrete trend and distribution pattern of the signal. They are mean, median, standard deviation (SD), range, kurtosis, skewness.

Finally, 3 types of motion features (time domain, frequency domain, and their combination) are calculated for 4 types of motion information. The names and compositions of facial motion features are shown in Table 1.

The purpose of feature dimension reduction is to remove redundant and irrelevant features, in turn improving recognition accuracy [41]. Classical PCA method needs to calculate covariance matrix. When feature dimension is too large, the running memory of the ordinary computer cannot meet calculation requirements. Therefore, PCA based on singular value decomposition is used to reduce the dimension of features. For each facial

motion feature, principal components are selected according to the contribution rate from large to small, until the cumulative contribution rate of the selected principal components reaches 95%. The results of feature dimension reduction are shown in Table 1.

3.3 Machine learning classification

Machine learning classification is used to determine behavior types. Compared with the statistical analysis method, the machine learning classifier is more suitable for problems with complex data [8]. Using machine learning classifiers to recognize unsafe behavior includes 2 steps: training and classifying.

During training, the classifier receives facial motion features and labels to build a classification model of unsafe behavior. During classifying, the classification model receives the facial motion features of recognition object and outputs the corresponding decision value. The decision value is a final description of the classification model to the recognition object. The classification model compares the decision value with a threshold value given artificially to realize unsafe behavior recognition. The object whose decision value is higher than the threshold value is determined as safe behavior, otherwise as unsafe behavior, to realize the recognition of unsafe behavior.

Two classical machine learning classifiers are used to build classification models. They are SVM [9] that is suitable for solving classification problems to small-scale train sets, and back propagation neural network (BPNN) that has been widely used in pattern recognition. Different classifiers have different models, so, multiple classifiers are used to explore the greater potential of facial motion features in recognition. Deep learning classifiers are not used in this paper. Because unsafe samples are difficult to collect. And the needs of deep learning classifiers for sample size [41] are difficult to meet.

Receiver operating characteristic (ROC) curve and area under curve (AUC) [14] are taken as performance criteria of the classification model. ROC curve takes both false positive rate and true positive rate of recognition result into account. Horizontal axis of ROC curve is false positive rate, which is proportion of safe samples misjudged as

Table 1 Names and compositions of facial motion features

Names of facial motion features	Constitution		Raw feature dimension	Feature dimension after reduction
	Motion information	Motion features		
GF	Gabor	Frequency domain	44,000	208
HF	HOG		3240	130
LF	LBP		2400	153
CF	Combination		49,640	209
GT	Gabor	Time domain	35,200	211
HT	HOG		2592	124
LT	LBP		1957	148
CT	Combination		39,749	214
GTF	Gabor	Time domain and frequency domain	79,200	214
HTF	HOG		5832	131
LTF	LBP		4357	152
CTF	Combination		89,389	217

*The dimensions containing NaN in facial motion features is removed

unsafe samples by the classification model. Vertical axis of ROC curve is true positive rate, which is correct proportion of the classification model in safe samples. Diagonal represents ROC curves in the case of random guesses, which can be used as a reference for whether the classification model has value. AUC represents the area under ROC curve and is used to measure the degree of ROC curve upper left corner. The larger AUC, the more ROC curve protrudes to upper left corner, and the higher accuracy of the classification model is.

4 Experiment and results

To investigate the feasibility of recognition unsafe behavior by facial motion, and find motion features with high recognition performance. We designed an experiment aim to collecting the facial motions of workers during safe and unsafe behaviors. Because unsafe behaviors are random and easy to cause casualties, which is difficult to collect in practice. This paper builds a dataset containing various types of work (carpentry, cutting, electrician and drive) and various unsafe behaviors (operation error, fatigue and distraction).

In similar simulation experiments, subjects are recruited to simulate safe and unsafe behaviors during various types of work in laboratory. The scenes of operations are shown in Fig. 3. Twenty-four students (16 males and 8 females) aged 20~24 years simulate safe and unsafe behaviors of carpentry, cutting and electrician. Safe behaviors are to operate according to operation regulations. Unsafe behavior is operation error, which mainly considers 2 factors: (1) behaviors that often occurs during operation; (2) behaviors that is easy to cause common accidents. Operation videos (resolution 1920×1080 , time 3~10 seconds, 24 frames per second) are collected by a common camera. Each operation video is cut to a 2 seconds sample, then 55 safe samples and 67 unsafe samples are collected.

In online collection, 86 safe samples and 72 unsafe samples with 2 seconds are collected through video websites and public datasets (YawDD [2]). Among them, 85 safe samples, 35 fatigue samples and 35 distraction samples are selected in YawDD dataset. Finally, 141 safe samples and 139 unsafe samples are collected and are assigned a label with 1 and -1 respectively. Information of samples is shown in Table 2.

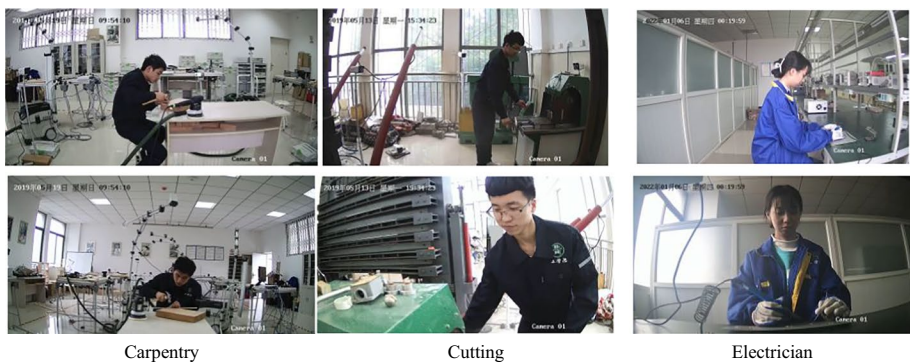


Fig. 3 Similar simulation experiment process. The first line is a scene of each operation, the second line is a frame of samples

Table 2 Information of samples

Behavior types	Label	Work types	Collecting methods	Number
Safe	1	Carpentry, cutting, electrician and drive	Similar simulation experiments and online collection	141
Operation error	-1	Carpentry, cutting, electrician and drive	Similar simulation experiments and online collection	69
Fatigue	-1	Drive	Online collection	35
Distraction	-1	Drive	Online collection	35

4.1 Relationships between complexity of facial motions and unsafe behavior

Revealing laws of facial motions during unsafe behaviors can provide clues for recognizing unsafe behavior. And it also can provide a reference for safety managers to recognize unsafe behaviors. We first show facial image sequence, motion information, and motion features under different behaviors. Then, the face is divided into different areas according to the five senses. The complexity of motion information of total face and facial areas are calculated. And reveal the laws by comparing the complexity during safe and unsafe behaviors.

Facial image sequence, motion information, and motion features under different behaviors are shown in Fig. 4. The sample of 2 s long is sampled every 7 frames from

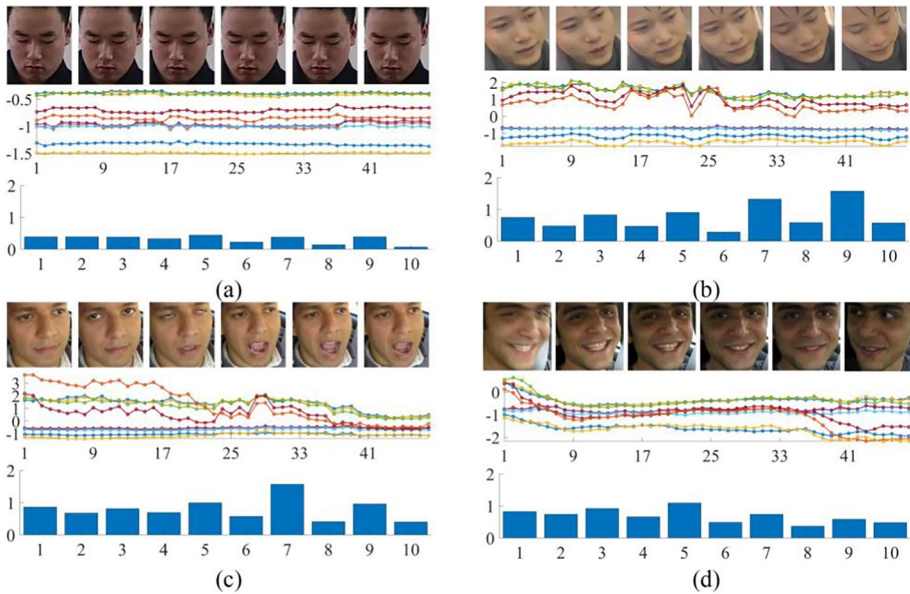


Fig. 4 Facial image sequence, motion information, and motion features under different behaviors. **a** is safe, **b** is operation error, **c** is fatigue, and **d** is distraction. For each subgraph, the first row is facial image sequence. The second row is the curve of 10 Gabor data values changing with time. Horizontal axis is number of frames, and vertical axis is value of Gabor data. The third row is the WSSE of these 10 curves. Horizontal axis is number of Gabor data, and vertical axis is value of WSSE

the first frame to obtain a facial image sequence. Curves of 10 Gabor data values changing with time were drawn to represent motion information. And WSSE of these 10 curves is calculated to represent motion features.

The facial motion under safe behavior is different from that under unsafe behavior. It can be seen from Fig. 4, compared with safe behavior, facial images and Gabor data at different times under unsafe behavior have more changes, and WSSE under unsafe behavior is higher. In operation error, fatigue and distraction, the mouth and nose of subjects have obvious motion. In fatigue and distraction, the eyes of subjects have obvious motion.

To quantitatively analyze the difference of facial motion under different behaviors, the face is divided into 3 areas: eye and brow, nose and mouth, as shown in Fig. 5a. WSSE of Gabor motion information is calculated to represent complexity. The higher the WSSE, the higher the complexity of facial motions. Mean histogram of WSSE in safe and unsafe samples is drawn, as shown in Fig. 5b.

Compared with safe behavior, the complexity of facial motion during unsafe behaviors is higher, especially in nose and mouth areas. It can be seen from Fig. 5, WSSE of total face in unsafe samples is higher than that of safe samples. And these differences are mainly reflected in the nose and mouth areas. WSSE of the eye and brow areas in unsafe samples is higher than that of safe samples, but no significant difference is found. This may be because unsafe behavior is usually manifested as deviation in action (such as too fast, too slow, etc.). When there are deviations in the action of workers, facial motions of workers may have corresponding deviations. And making facial motions are more complex during unsafe behaviors. Revealing the theoretical mechanism of this phenomenon may need the help of EEG and other professional equipment, which needs to be studied by physiological and psychological researchers.

The complexity of facial motions can be used as a new clue to recognize unsafe behaviors. Compared with the eye and brow area, nose and mouth areas are found to have more significant differences in safe and unsafe behavior. When recognizing unsafe behaviors, safety managers can attempt to pay more attention to nose and mouth areas of workers.

4.2 Influence of frequency domain features on recognition accuracy

Finding facial motion information and feature with high accuracy in recognizing unsafe behaviors can provide a reference for building classification models. We extract time domain and frequency domain features of various facial motion information. The influence

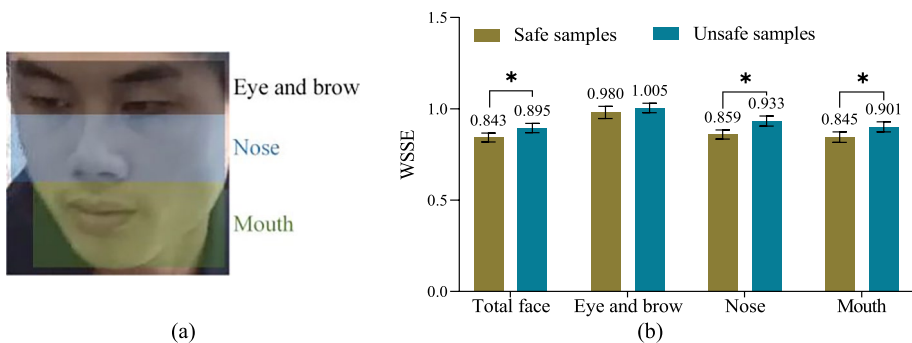


Fig. 5 WSEE of facial Gabor motion information during safe and unsafe samples. **a** Schematic of facial area. **b** Mean histogram (mean with 95% confidence interval) of WSSE. Asterisk represents that there are significant differences (significance of Mann Whitney U test <0.05) between the 2 groups of data

of motion features on recognition accuracy is explored by recognition results of machine learning classification model.

SVM and BPCNN classifiers are used to train 4 frequency domain features, 4 time domain features and 4 time and frequency domain features. Mean AUC of 10 times of 5-fold cross validation of the 24 models is compared, as shown in Fig. 6.

Classification models built by frequency domain features of facial Gabor motion information have higher accuracy. It can be seen from Fig. 6, mean AUC of the classification models built by GF feature is higher than that of the other 11 facial motion features, under SVM classifiers. And among all models, mean AUC of the classification model built by GF feature and SVM classifier is the highest, which is 0.766.

For motion features, frequency domain is more conducive to recognition. It can be seen from Fig. 6, mean AUC of classification models built by frequency domain features is higher than other motion features when motion information is the same. And the influence of motion features on recognition accuracy is greater than that of motion information. For motion information, Gabor is more conducive to recognition. It can be seen from Fig. 6, under SVM classifiers, mean AUC of classification models built by Gabor motion information is higher than other motion information when motion features are the same. And among all models, the model with the highest mean AUC is also built by Gabor motion information.

To further explore the relationship between motion features and recognition, the relevance between motion features and behavior labels is calculated. Taking GTF features as an example, mutual information entropies between each feature and behavior labels are

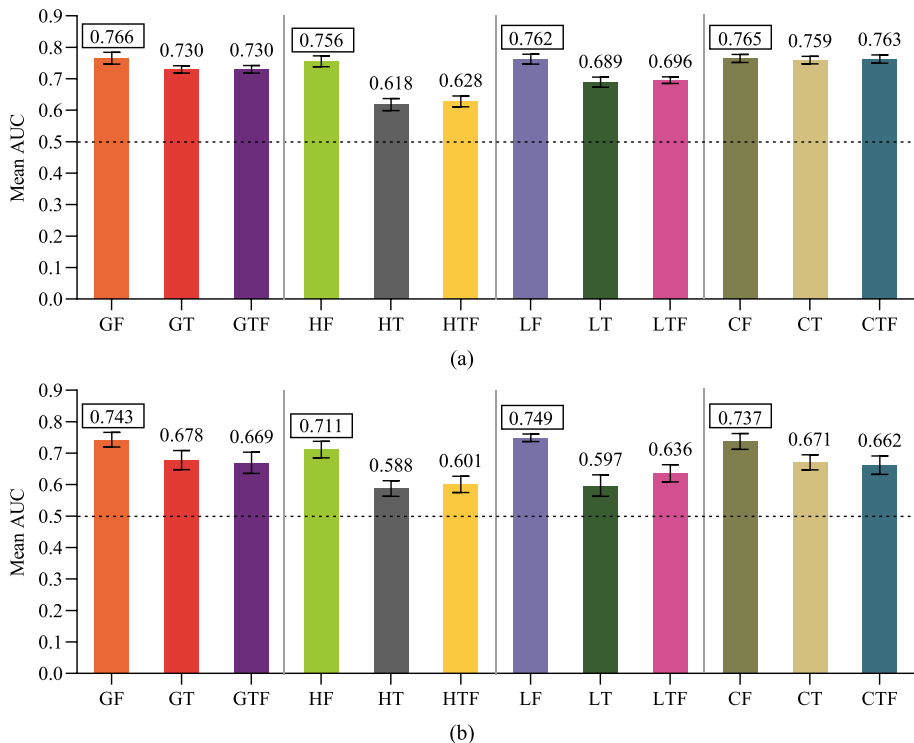


Fig. 6 Histogram of mean AUC (\pm standard deviation) of classification models. **a** Histogram of mean AUC under SVM classifier. **b** Histogram of mean AUC under BPCNN classifier

calculated to represent the relevance between them. The percentage of various motion features among the top 1% (792) most relevant features is calculated, as shown in Fig. 7.

Compared with time domain features, frequency domain features are more relevant to recognition. Compared with other features, features describing the complexity of facial motions are more relevant to recognition. It can be seen from Fig. 7, among the top 1% most relevant features, percentage of frequency domain features (65.783%) is higher than time domain features (34.217%). And the percentage of features describing facial motion complexity (96.97%) is much higher than other features (3.03%).

Calculating frequency domain features of facial Gabor motion information can obtain higher accuracy, when recognizing unsafe behaviors through facial motions. The influence of motion features on recognition accuracy is greater than that of motion information. And describing facial motion features from frequency domain and complexity is helpful to improve the recognition accuracy.

4.3 Unsafe behavior recognition model with higher applicability

To improve applicability and accuracy of the model, we describe facial motions by the frequency domain complexity to build an unsafe behavior recognition model. To analyze advantages of the model, current researches that have performed well in recognizing unsafe behaviors by facial motions are compared with us.

A classification model is built by GF feature and SVM classifier. Mean ROC curve of 10 times of 5-fold cross validation of the model is used for performance evaluation, as shown in Fig. 8.

Facial motions can recognize unsafe behaviors of workers such as operation errors, fatigue and distraction. It can be seen from Fig. 8, mean ROC curve of the model is above the diagonal, and mean AUC of the model is 0.766. This shows that prediction results of the model have a higher true positive rate under any false positive rate, which has predictive value. Mean AUC of the model for 3 types of unsafe behaviors is higher than 0.5. And the model has higher performance (AUC=0.818) in recognizing operation errors.

Recognizable unsafe behavior types, applicable work types and recognition accuracy are used to analyze the model. Researches on recognizing fatigue [7] and distraction [34] by facial motions are selected to compare with ours. Fatigue and distraction samples used in these 2 researches are also from the YawDD dataset. The comparison information is shown in Table 3.

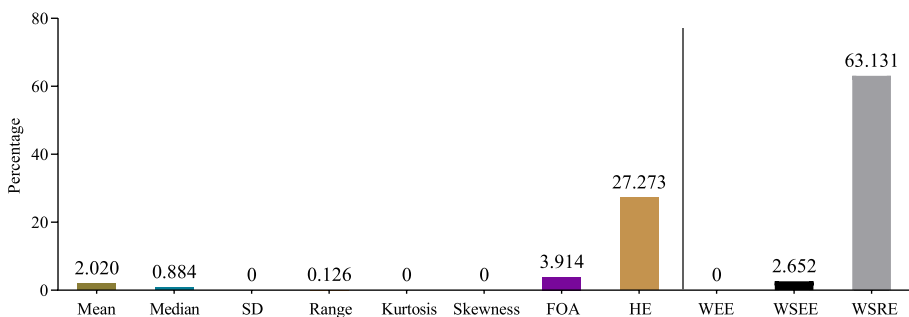
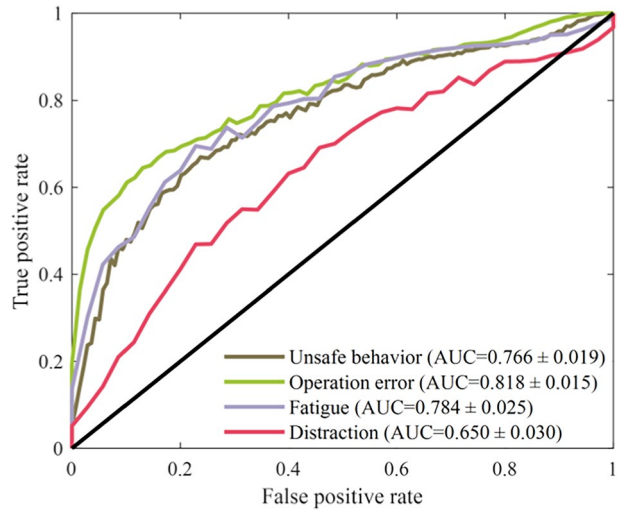


Fig. 7 Percentage of various motion features among the top 1% most relevant features

Fig. 8 Mean ROC curves of the classification model built by GF feature and SVM classifier



The model built in this paper can recognize operation errors, fatigue and distraction. In particular, the model has achieved good performance (AUC = 0.818, F-score = 0.686, ACC = 0.736) on operational errors not involved in other researches. The recognition accuracy of the model for fatigue and distraction is lower than the deep learning model of Bai et al. and Yang et al. If the model of our, Bai et al. and Yang et al. can be fused by multi-model fusion algorithms, the high recognition accuracy of operational errors, fatigue and distraction can be achieved.

The model built in this paper is suitable for various work types. In addition to driving, the model can recognize operational errors in carpenters, cutters and electricians. These types of work include sitting posture, standing posture and mixed posture. It leads to various angles of faces in our dataset, and making it more challenging to recognize unsafe behaviors.

The model built in this paper has a good performance (AUC = 0.766, F-score = 0.734, ACC = 0.721) in recognizing various unsafe behaviors. This may benefit from describing facial motion information from frequency domain and complexity. In addition, we

Table 3 Comparison of researches on recognizing behaviors by facial motions

Researcher	Recognition accuracy			
	Operation error in carpentry, cutting, and electrician (our dataset)	Fatigue in driving (YawDD dataset)	Distraction in driving (YawDD dataset)	Mixture
Bai et al. [7]	–	AUC = – F-score = 0.895 ACC = 0.934	–	–
Yang et al. [34]	–	–	AUC = – F-score = – ACC = 0.812	–
Ours	AUC = 0.818 F-score = 0.686 ACC = 0.736	AUC = 0.7843 F-score = 0.527 ACC = 0.7193	AUC = 0.643 F-score = 0.415 ACC = 0.628	AUC = 0.766 F-score = 0.734 ACC = 0.721

compared the calculation time of frequency domain features (WEE, 8 WSE and WSSE) and time domain features (mean, variance, maximum and minimum) of Gabor motion information for 48 frames. Due to the large dimension of Gabor motion information, the calculation time of frequency domain feature (16.15 s) is higher than that of time domain feature (0.261 s). Considering the real-time performance of the method, LBP motion information with small dimension can be used, and its frequency domain characteristic calculation time is 1.35 s. And LF feature also has a high recognition accuracy (AUC = 0.762).

5 Conclusions

To recognize unsafe behavior of workers, we describe facial motion information through frequency domain, build classification models of unsafe behaviors, and find facial motion features with high recognition accuracy. The following conclusions are obtained: (1) Compared with safe behavior, the complexity of facial motions during unsafe behavior is higher, especially in nose and mouth areas. (2) Wavelet entropy frequency domain features of Gabor motion information can better describe the complexity and have higher recognition accuracy (AUC = 0.766); (3) The proposed method can recognize unsafe behaviors of workers and is effective for operation errors (AUC = 0.818).

In the next step, we will build a model with high recognition accuracy for operation errors, fatigue and distraction through the multi model fusion algorithm. And the mathematical model between facial motions and unsafe behavior will be built by using feature coding to improve the interpretability of recognition.

Funding This work was supported by the MOE (Ministry of Education in China) Project of Humanities and Social Sciences [grant numbers 17XJC630004]; the Sichuan Science and Technology Program [grant numbers 2018SZ0351].

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Abdallah TB, Elleuch I, Guermazi R (2021) Student behavior recognition in classroom using deep transfer learning with VGG-16. *Proced Comput Sci* 192:951–960. <https://doi.org/10.1016/j.procs.2021.08.098>
2. Abtahi S, Omidyeganeh M, Shirmohammadi S, Hariri B (2014) YawDD: a yawning detection dataset. In proceedings of the 5th ACM multimedia systems conference on (pp. 24–28). <https://doi.org/10.1145/2557642.2563678>
3. Alexeenko V, Howlett PJ, Fraser JA, Abasolo D, Han TS, Fluck DS, Fry CH, Jabr RI (2021) Prediction of paroxysmal atrial fibrillation from complexity analysis of the sinus rhythm ECG: a retrospective case/control pilot study. *Front Physiol* 12:570705. <https://doi.org/10.3389/fphys.2021.570705>
4. Antwi-Afari MF, Li H (2018) Fall risk assessment of construction workers based on biomechanical gait stability parameters using wearable insole pressure system. *Adv Eng Inform* 38:683–694. <https://doi.org/10.1016/j.aei.2018.10.002>
5. Avila F, Delrieux C, Gasaneo G (2019) Complexity analysis of eye-tracking trajectories: permutation entropy may unravel cognitive styles. *Eur Phys J B* 92(12):273. <https://doi.org/10.1140/epjib/e2019-100437-4>

6. Avola D, Cascio M, Cinque L, Fagioli A, Foresti GL (2021) LieToMe: an ensemble approach for deception detection from facial cues. *Int J Neural Syst* 31(02):2050068. <https://doi.org/10.1142/S0129065720500689>
7. Bai J, Yu W, Xiao Z, Havyarimana V, Regan AC, Jiang H, Jiao L (2021) Two-stream spatial-temporal graph convolutional networks for driver drowsiness detection. *IEEE T Cybern* 52(12):13821–13833. <https://doi.org/10.1109/TCYB.2021.3110813>
8. Breiman L (2001) Statistical modeling: the two cultures (with comments and a rejoinder by the author). *Stat Sci* 16(3):199–231. <https://doi.org/10.1214/ss/1009213726>
9. Chang CC, Lin CJ (2011) LIBSVM: a library for support vector machines. *ACM T Intel Syst Tec* 2(3):27
10. Chen Z, Wu L, He H, Jiao Z, Wu L (2022) Vision-based skeleton motion phase to evaluate working behavior: case study of ladder climbing safety. *Hum-Cent Comput Info* 12:1–18. <https://doi.org/10.22967/HCCIS.2022.12.001>
11. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. <https://doi.org/10.1109/CVPR.2005.177>
12. Daud SNSS, Sudirman R (2022) Wavelet based filters for artifact elimination in electroencephalography signal: a review. *Ann Biomed Eng* 50(10):1271–1291. <https://doi.org/10.1007/s10439-022-03053-5>
13. Du Y, Wang Y, Huang X, Hu Q (2018) Driver state analysis based on imperfect multi-view evidence support. *Neural Process Lett* 48(1):195–217. <https://doi.org/10.1007/s11063-017-9698-z>
14. Fawcett T (2006) An introduction to ROC analysis. *Pattern Recogn Lett* 27(8):861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
15. Figueroa-Jimenez MD, Carbó-Carreté M, Cañete-Massé C, Zarabozo-Hurtado D, Perú-Cebollero M, Salazar-Estrada JG, Guàrdia-Olmos J (2021) Complexity analysis of the default mode network using resting-state fMRI in down syndrome: relationships highlighted by a neuropsychological assessment. *Brain Sci* 11(3):311. <https://doi.org/10.3390/brainsci11030311>
16. Gjoreski M, Gams MZ, Luštrek M, Genc P, Garbas JU, Hassan T (2020) Machine learning and end-to-end deep learning for monitoring driver distractions from physiological and visual signals. *IEEE Access* 8:70590–77060. <https://doi.org/10.1109/ACCESS.2020.2986810>
17. Guo BHW, Goh YM, Wong KLX (2018) A system dynamics view of a behavior-based safety program in the construction industry. *Saf Sci* 104:202–215. <https://doi.org/10.1016/j.ssci.2018.01.014>
18. Han S, Lee S, Peña-Mora F (2014) Comparative study of motion features for similarity-based modeling and classification of unsafe actions in construction. *J Comput Civ Eng* 28(5):A4014005. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000339](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000339)
19. Heinrich HW (1959) Industrial accident prevention: a safety management approach. McGraw-Hill Book Company, New York
20. Hu J, Liu W, Kang J, Yang W, Zhao H (2019) Semi-cascade network for driver’s distraction recognition. *P I Mech Eng D-J Aut* 233(9):2323–2332. <https://doi.org/10.1177/0954407019857408>
21. Jabon ME, Ahn SJ, Bailenson JN (2011) Automatically analyzing facial-feature movements to identify human errors. *IEEE Intell Syst* 26(2):54–63. <https://doi.org/10.1109/MIS.2009.106>
22. Jebelli H, Ahn CR, Stentz TL (2016) Comprehensive fall-risk assessment of construction workers using inertial measurement units: validation of the gait-stability metric to assess the fall risk of iron workers. *J Comput Civ Eng* 30(3):4015034. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000511](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000511)
23. Jeelani I, Albert A, Han K, Azevedo R (2019) Are visual search patterns predictive of hazard recognition performance? Empirical investigation using eye-tracking technology. *J Constr Eng M* 145(1):04018115. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0001589](https://doi.org/10.1061/(ASCE)CO.1943-7862.0001589)
24. Kerdawy ME, Halaby ME, Hassan A, Maher M, Fayed HA, Shawky D, Badawi AH (2020) The automatic detection of cognition using EEG and facial expressions. *Sensors* 20(12):3516. <https://doi.org/10.1109/ICME.2014.6890301>
25. Li Z, Zhan Z (2020) Integrated infrared imaging techniques and multi-model information via convolutional neural network for learning engagement evaluation. *Infrared Phys Technol* 109:103430. <https://doi.org/10.1016/j.infrared.2020.103430>
26. Liu C, Wechsler H (2002) Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *IEEE T Image Process* 11(4):467–476. <https://doi.org/10.1109/TIP.2002.999679>
27. Monaro M, Maldera S, Scarpazza C, Sartori G, Navarin N (2022) Detecting deception through facial expressions in a dataset of videotaped interviews: a comparison between human judges and machine learning models. *Comput Hum Behav* 127:107063. <https://doi.org/10.1016/j.chb.2021.107063>

28. Monkareisi H, Bosch N, Calvo RA, D’Mello SK (2016) Automated detection of engagement using video-based estimation of facial expressions and heart rate. *IEEE T Affect Comput* 8(1):15–28. <https://doi.org/10.1109/TAFFC.2016.2515084>
29. Navarathna R, Carr P, Lucey P et al (2017) Estimating audience engagement to predict movie ratings. *IEEE T Affect Comput* 10(1):48–59. <https://doi.org/10.1109/TAFFC.2017.2723011>
30. Ojala T, Pietikäinen M, Harwood D (1996) A comparative study of texture measures with classification based on featured distributions. *Pattern Recogn* 29(1):51–59. [https://doi.org/10.1016/0031-3203\(95\)00067-4](https://doi.org/10.1016/0031-3203(95)00067-4)
31. Shin IJ (2013) Major industrial accidents in Korea: the characteristics and implication of statistics 1996–2011. *Process Saf Prog* 32(1):90–95. <https://doi.org/10.1002/prs.11551>
32. Takalkar M, Xu M, Wu Q, Chaczko Z (2018) A survey: facial micro-expression recognition. *Multimed Tools Appl* 77(15):19301–19325. <https://doi.org/10.1007/s11042-017-5317-2>
33. Wang D, Li H, Chen J (2019) Detecting and measuring construction workers’ vigilance through hybrid kinematic-EEG signals. *Autom Constr* 100:11–23. <https://doi.org/10.1016/j.autcon.2018.12.018>
34. Yang H, Liu L, Min W, Yang X, Xiong X (2020) Driver yawning detection based on subtle facial action recognition. *IEEE T Multimed* 23:572–583. <https://doi.org/10.1109/TMM.2020.2985536>
35. Yu Y, Guo H, Ding Q, Li H, Skitmore M (2017) An experimental study of real-time identification of construction workers’ unsafe behaviors. *Autom Constr* 82:193–206. <https://doi.org/10.1016/j.autcon.2017.05.002>
36. Yuce A, Gao H, Cuendet GL, Thiran JP (2017) Action units and their cross-correlations for prediction of cognitive load during driving. *IEEE T Affect Comput* 8(2):161–175. <https://doi.org/10.1109/TAFFC.2016.2584042>
37. Yue Y, Yanyan L (2018) Application study of BBS on unsafe behavior and psychology of coal miners. *Neuroquantology* 16(4):52–61. <https://doi.org/10.14704/nq.2018.16.4.1188>
38. Zhang P, Li N, Fang D, Wu H (2017) Supervisor-focused behavior-based safety method for the construction industry: case study in Hong Kong. *J Constr Eng M* 143(7):5017009. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0001294](https://doi.org/10.1061/(ASCE)CO.1943-7862.0001294)
39. Zhang M, Shi R, Yang Z (2020) A critical review of vision-based occupational health and safety monitoring of construction site workers. *Saf Sci* 126:104658. <https://doi.org/10.1016/j.ssci.2020.104658>
40. Zhang Z, Li Z, Liu H, Cao T, Liu S (2020) Data-driven online learning engagement detection via facial expression and mouse behavior recognition technology. *J Educ Comput Res* 58(1):63–86. <https://doi.org/10.1177/0735633119825575>
41. Zhao L, Wang Z, Wang X, Liu Q (2018) Driver drowsiness detection using facial dynamic fusion information and a DBN. *Iet Intell Transp Sy* 12(2):127–133. <https://doi.org/10.1049/iet-its.2017.0183>
42. Zhao L, Wang Z, Zhang G, Gao H (2020) Driver drowsiness recognition via transferred deep 3D convolutional network and state probability vector. *Multimed Tools Appl* 79(35):26683–26701. <https://doi.org/10.1007/s11042-020-09259-w>

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.