IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 20 June 2022, accepted 6 July 2022, date of publication 22 July 2022, date of current version 8 August 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3192349

APPLIED RESEARCH

Research on Behavior Analysis of Real-Time Online Teaching for College Students Based on Head Gesture Recognition

LIU QIANQIAN[1], WAN QIAN[2,3,4], XIE BOYA[2], LIN CHURAN[2], XIE ZHENYOU[2], PAN SHU[1], AND PENG GUOQING[2,3,4]

[1]School of Business Administration, Guangxi University of Finance and Economics, Nanning 530000, China
[2]School of Architecture and Transportation Engineering, Guilin University of Electronic Technology, Guilin 541004, China
[3]School of Transportation, Southeast University, Nanjing 211189, China
[4]Hualan Design and Consulting Group, Nanning 530011, China

Corresponding authors: Liu Qianqian (liuqianqian2018@gxufe.edu.cn), Wan Qian (1564656392@qq.com), and Peng Guoqing (243564846@qq.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 51508122, in part by the National Social Science Fund of China under Grant 21BMZ072, in part by the 2020 Guangxi Philosophy and Social Science Planning Research Project under Grant 20FGL009, in part by the scientific research project of National Ministry of Housing and Urban Rural Construction under Grant 2019-K-151, in part by the Guangxi Key Project of Research and Development under Grant Guike AB17292087, in part by the scientific research project of Guangxi Department of Housing and Urban Rural Construction in 2022 (the project name is ''Reserach on key technologies of space layout optimization and management and control evaluation of airport terminal under normalized epidemic prevention and control'').

This work involved human subjects in its research. Approval of all ethical and experimental procedures and protocols was granted by the Academic Committee of the School of Business Administration, Guangxi University of Finance and Economics.

ABSTRACT The accessibility of online teaching makes it popular in various teaching scenarios. Most of these researches about online interaction mainly focuse on the communication network stability, facial expression/gesture recognition algorithm and statistical description analysis, and its expertise mainly comes from the fields of computer technology and algorithm engineering. The evaluation and analysis of the effect of online teaching is an important test for the adaptability of such tools in the field of education. However, from the perspective of teachers, there is still a lack of literature on data interpretation after the application of this technology. An experiment based on real online teaching was carried out in this paper. This study uses image recognition technology to process video and extract five kinds of head movement data from dozens of student samples, and then develop statistical description interpretation. Some novel and interesting conclusions indicate that diversified behaviors occurred in real-time online learning. This study obtained data of five high-frequency online learning behaviors, including blinking, yawning, nodding, shaking head and leaving. These behaviors are related to learning state and time. Teaching features, students' personal characteristics and learning environment have a comprehensive impact on online learning behaviors. The result provides a basis for personalized learning and teaching scheme design in the future. It also helps to enrich online teaching evaluation methods and accelerate the construction of online education framework and rules.

INDEX TERMS Real-time online teaching, head gesture recognition, deep learning algorithm, experiment.

I. INTRODUCTION

Online teaching has become the main front of educational reform and innovation, and online learning behavior has also attracted researchers' interests. It is important to collect and identify the behavior data of online learning, and to analyze the relevance between behavior and learning effect. More researches are needed in this area to help analyze the effect of online teaching accurately.

Online learning behaviors have not been clearly defined due to the differences in research perspectives. In general, it refers to some activities in which learners regulate

The associate editor coordinating the review of this manuscript and approving it for publication was Juan A. Lara.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

their cognition and motivation, to use learning resources and content through network technology actively, so as to achieve network learning objectives. The learners have full autonomy to decide the various states of learning, and communicate under the help of the Internet to get more help. Online learning behavior can be defined as the learning behavior generated by learners in the state of network connection. The environment can be MOOC, e-learning, online management system (LMS), online teaching platform and so on. The online learning behavior studied in this paper refers to the real-time learning behavior with the help of various online management systems and network teaching platforms.

In fact, students' motivation, cognition, emotion and movements all have important impacts on learning effect, according to the research results about ITS (Intelligent Tutoring System) [1]. In online learning, students have more autonomy to adjust their behavior and can seek help through online communication. In other words, they are self-control and self-restraint [2]. It is needed to identify students' behaviors during online learning through modern technology and equipment, such as computer vision. The key step is to recognize the types of online learning behaviors.

Learning behavior recognition can be divided into three stages: data preprocessing, finding effective features and classification based on extracted features [2]. Most of the researches about learning behavior are from offline research or asynchronous e-Learning, and there are few researches on synchronous real-time online teaching [3], [4], [5], [15], [16]. Behavior recognition includes gesture recognition, voice recognition and facial expression recognition [2]. Previous researches are mainly about facial emotions and body postures [18], [19], [20]. Studies have shown that humans often combine the visual channels of face and body more than other channels [2], [21], [22], [23].

Different from previous researches, this paper focuses on real-time online synchronous learning, and the behavior data is collected and classified through gesture recognition algorithm. To extract the reasonable data of learning behavior, the authors design an experiment to satisfy real-time online synchronous learning conditions.

A. RESEARCH MOTIVATION

To analyze students' online learning behaviors using intelligent technology such as computer vision, it is necessary to recognize facial features and head gestures. The research of gesture recognition provides convenience for online learning behaviors. HOG (Histogram of Oriented Gradient), CNN (Convolutional Neural Network), Dlib, etc. is widely used in facial feature recognition [2]. The behavior features and learning effect of students' online learning are rarely studied through gesture recognition technology. In order to study the behavior of online learning, the motivation of this research is to solve the following questions:

Research question (RQ 1): How to recognize online learning behaviors through gesture recognition algorithm?

Research question (RQ 2): What are the features of students' gesture in online learning?

Research question (RQ 3): What are the influencing factors related to the gesture of online learning?

The deep learning algorithm used in this study is based on HOG (Histogram of Oriented Gradients), and it combines the facial feature recognition with the head pose recognition. The authors selected several facial expressions and head gestures that are proved to be related to the learning effect, and adjust the algorithm features to adapt to the online scene and the sample characteristics of China. In order to define the feature target, the authors carried out a preliminary experiment to collect the expressions and gestures. Finally, we eventually defined five gestures (blinking, yawning, nodding, shaking and leaving), and proposed corresponding recognition algorithm.

B. RESEARCH CONTRIBUTION

This study proposed a hybrid method to identify and extract online learning behaviors of Chinese college students in online learning scenarios. The program adopted in this study combines facial feature recognition and head gesture recognition based on HOG. It improved the existing algorithm to adapt to online environment and sample characteristics, and defined five gestures: blinking, yawning, nodding, shaking head and leaving. We analyzed the recognition data and found that:

There are some interesting results among these gestures. The occurrences of leaving and blinking are more than nodding, and are related to time. That is the result of the combined effect of teaching characteristics, individual characteristics and external environment.

This study contributes to the research of online real-time learning behavior. 1) It deepens the research of online learning behavior, and AI algorithm is used to collect online learning behavior. 2) The online learning behavior data provides a basis for personalized learning and teaching scheme design in the future. It also helps to enrich online teaching evaluation methods and indicators, and has reference significance for the construction of online education framework and norms. Moreover, the result expands the application of human-machine interaction in online teaching, which has reference significance for the technology optimization and updating. By embedding the recognition algorithm into teaching system, it is conducive to the collection, processing and analysis of online learning behavior data.

II. BACKGROUND AND RELATED WORK

A. REVIEW OF ONLINE TEACHING IN CHINA DURING COVID-19

The scale of online teaching in China has grown constantly in the past five years. During the COVID-19, 1454 Chinese universities launched online teaching. As of May 8, 2020, a total of 103 million teachers have offered 1.07 courses on online platforms, and 17.75 million university students have participated in online learning.

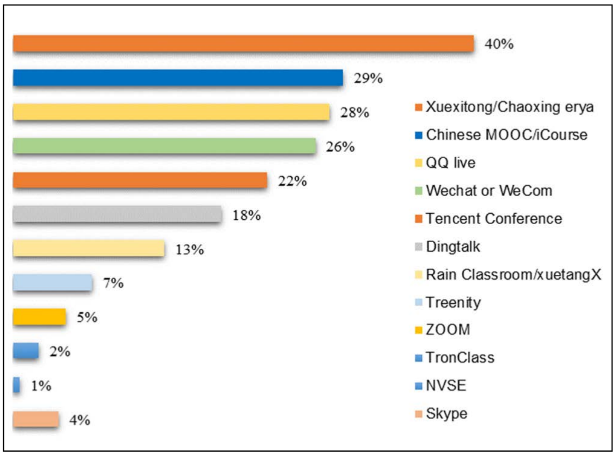

FIGURE 1. Overview of the use of online teaching platforms in Chinese universities during the 2020 pandemic (data from Wu D., 2020[5]).

Online teaching has other new changes during the COVID-19, these changes have great implications for education reform around the world. Online teaching is the most commonly method during COVID-19, such as live broadcast, video recording, MOOC, text and audio, online interactive discussion and so on. Among them, the combination of watching teaching live videos and interactive question-and-answer account for the vast majority. Online teaching is developing rapidly because of COVID-19, therefore the researchers need to summarize and discover new problems. Based on the development status of online teaching in China, we find that:

1) MULTIPLE PLATFORMS ARE USED, BUT NOT EVENLY

There are 37 online platforms offering free online courses for colleges and universities nationwide, they are all organized by China's Ministry of Education. The use of these platforms is diversified and scattered, according to the findings of the Teacher Development Center of Xiamen University. As shown in Figure 1, over 50% colleges and universities use Chaoxing erya, QQ live, Chinese MOOC, DingTalk, WeChat, and Tencent Meeting. Among them, ChaoXing erya platform is used by more than 75% of universities. The average number of online platforms used by universities is 6.9 [5]. Two international platforms called "iCourse" and "xuetangX", are provided by the Ministry of Education of China to share online education experience and results in China. In addition, many platforms offer overseas versions, such as Voov Meeting.

Some research shows that few platforms can independently meet all needs of online teaching, so teachers have to use multiple platforms simultaneously. The frequent switching between platforms affects the concentration of students and the collection of learning data.

2) TECHNICAL SERVICES MEET THE BASIC TEACHING NEEDS, BUT THE STUDENT-CENTERED NEEDS REMAIN UNMET

Technical service is the basic guarantee of the teaching platform, and the critical factor of the online learning experience.

TABLE 1. A summary of the reasons for the poor online teaching effect.

Category	Content of reasons
Teacher	insufficient teaching resources; unfamiliar with platforms or tools; unreasonable teaching methods and modes; lack of guidance and communication to students; unable to understand students' learning status and effects in time, etc.
Student	low self-control; lack of interactive and communication; poor learning environment, etc.
Platform	imperfect equipment and network; incomplete platform functions; multi-platform switching, etc.

Each platform includes many functions, such as attendance management, homework management, online speech, online discussion, etc. Most of platforms can meet basic teaching needs, but some platforms are weak in supporting functions, such as analysis of student behaviors and evaluation of teaching effect. In online teaching, teachers often concern about how to deliver teaching content smoothly, and students concern about whether they can get feedback in time [5]. Therefore, online teaching should be student-centered and meet the needs of students' feedback, however the design of teaching platforms is still lack of feedback module.

Real-time video technology makes online learning closer to offline learning, especially live video broadcast. Real-time video teaching is helpful to supervise students' learning status and effect for teachers and managers. Apply artificial intelligence technology to the field of education, would solve the lack of feedback between teachers and students. For example, deep learning algorithm would analyze a large number of learning behavior data, and provide basis for teaching decision-making. Nevertheless, relevant theoretical research and practical exploration are still lacking.

3) THE TEACHING EFFECT HAS BEEN ACCEPTED, BUT THERE ARE STILL PROBLEMS TO BE SOLVED

Online teaching has supported the courses during the pandemic. It is generally considered that online teaching has more advantages than offline teaching, as the high-quality teaching resources can be shared and repeated replay, which means that time and space restrictions are broken. But some people hold the opposite opinion, it is difficult for teachers to judge students' learning status and real-time feedback. TABLE 1 lists a summary of the reasons for the poor online teaching effect.

The effect of online teaching is closer to offline teachin due to the progress of technology. But it is hard for teachers to observe students' status and learning effect in time. Now, computer vision technology makes it possible. For example, the recognition auxiliary system records the changes of students' facial gesture and judges the learning status, and then teachers adjust the teaching process to improve the teaching effect.

TABLE 2. Existing latest work.

Title	Content	References
Online teaching	overall trend, guiding strategies for teachers, online learning ability and effect	Aristovnik et al., 2020; Mishra et al., 2020; Cao et al., 2020; Mahmood, 2020; Perez-Gaspar et al., 2016; Gunes et al., 2007
Application of AI and IOT	learning behavior, human feature resources	Crossley et al., 2017; Diana et al., 2017; Qin et al., 2020; Muzammal et al., 2020
Gesture recognition in offline learning	statue of students, gestures of teachers	R. L. L. Sie et al., 2018; Hrastinski, 2008; Gu et al., 2020;
Online learning and learning outcomes	relationship	Shimada,2015; Prior et al., 2016; Mao et al., 2019 Derkach et al., 2017; Guo, 2019;
Recognition algorithm	Image-based recognition, CNN, HMMs, ANNs	Imani et al., 2019; Cao et al.,2020; Vermun et al.,2013; Cao et al., 2017; Jiang et al., 2020

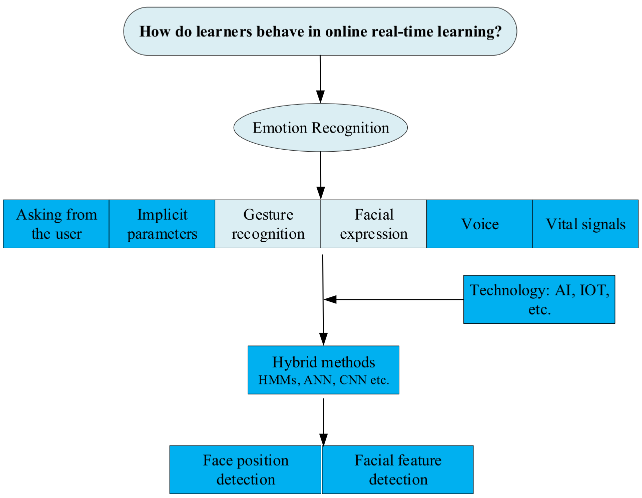

FIGURE 2. Research diagram. Existing researches usually use emotion recogniton to explore learning behaviors, and the head gesture recognition used in this study is the combination of face position detection and facial feature detection.

B. APPLICATION OF GESTURE RECOGNITION IN ONLINE TEACHING RESEARCH

Some scholars have carried out a lot of researches on online learning, focusing on the overall trend of online teaching [7]. They provided some teaching suggestions from the perspective of teachers or curriculum design. For example, Mahmood presented instructional strategies to help teachers in implementing online teaching in higher education, including maintaining a slow voice [10]. Some scholars have also conducted country-specific research on online teaching in China, India and other countries [8], [9], [34], [35], and learning ability especially autonomy, has been proved to influencing online teaching effect [11].

The application of artificial intelligence technology (AI) provides a new method for the research of learning behavior

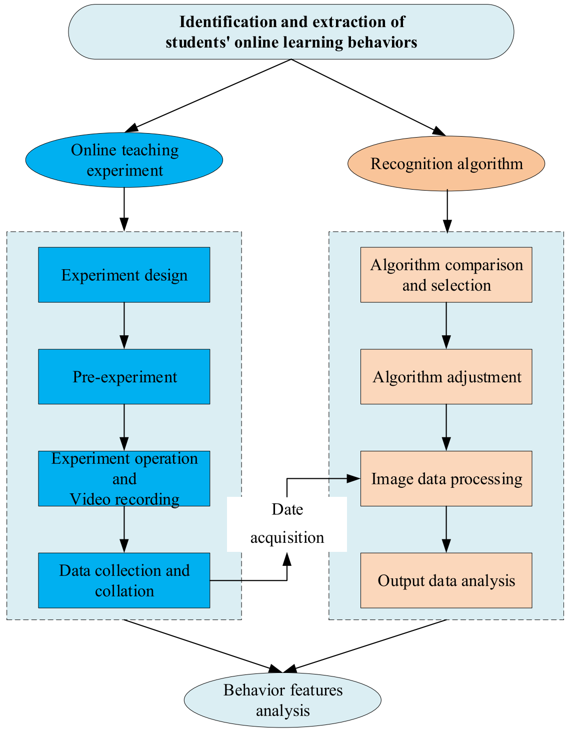

FIGURE 3. The figure illustrating of the method. The identification and extraction of students' online learning behaviors are realized by online teaching experiment and recognition algorithm.

in the field of education, but the research in online scene is insufficient. AI has been proved to be able to predict students' grades, and the application of deep learning algorithm has achieved unexpected results [12], [13].

In the previous studies, gesture recognition algorithms were applied to study the status of students and gesture of teachers in offline learning [15], [16], [17]. Some studies explored the relationship between online design and learning outcomes of students [18], as well as the relationship between online learning behavior and learning effect in MOOC [19], [20], [21], [22]. How to identify students' online learning status and behavior is a key problem. It is very important to improve the effect of online teaching. But real-time video lecture, which has been widely used during the pandemic, still lacks sufficient quantitative research. An important challenge now is how to identify students' learning state and behavior, which is of great use to improve the online teaching quality.

The application of computer vision in education has long been an important topic of research. Facial expressions, pulse frequency, breathing rate and other biological signals are jointly used to estimate learners' emotional state and interest level [23]; learner's facial expressions and movements are obtained through a web camera to estimate their interest level [24]; the composite texture and geometric features of different head gesture pictures can be used to evaluate students' attention with the aid of regression iterative algorithms [25]. But now, relevant quantitative study is lacking during the epidemic, and real-time video is widely used in the field of education.

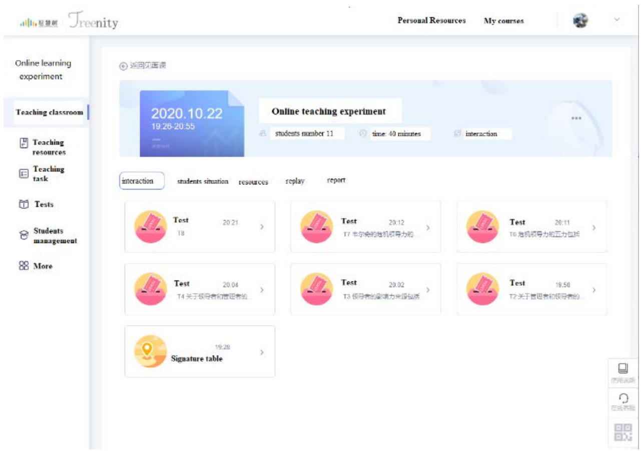

FIGURE 4. The screenshot of Treenity. Treenity is mainly for learning tests, PPT sharing, interaction and so on. Subjects' learning records and effects are also recorded in Treenity.

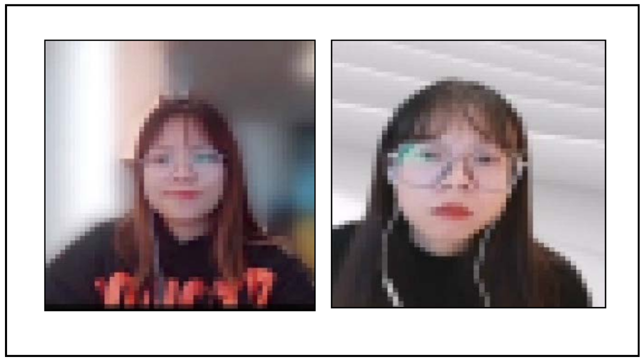

FIGURE 5. The screenshot of subjects recorded by DingTalk. DingTalk is mainly used for live broadcasting and video recording. We blurred the subjects' face for privacy.

With the development of artificial intelligence, technologies such as human-computer interaction and pattern recognition have been applied to online teaching to understand and imitate learning gestures and behaviors. It is helpful for the teaching aid systems understand learners' attitude and intentions, and predict their needs to provide active services. Gesture recognition of human bones information has been applied in remote conference, remote teaching and other scenarios to feedback the learning status [26], [27], [28] and [29]. Here we contribute to head gesture of learners in online teaching, to explore how they behave in all-remote settings.

Image-based gesture recognition algorithms help to improve the interaction between computer and students. Cao *et al.* proposed a method to recognize the 2D gesture of multiple people using a single image [30], which can achieve quasi-real-time effects under GPU acceleration. Mehta *et al.* proposed a 3D human gesture calculation method based on monocular camera using Convolutional Neural Network (CNN) [31]. Vermun *et al.* (2013) applied the gesture recognition algorithm in distance teaching to feedback the learning status of distance learners [27]. However, these image-based gesture recognition algorithms are difficult to control the error, and are easily influenced by screen background, image input analysis and other factors.

Benefit by the application of IOT and advanced sensors, human feature resources can be transformed into big data [45]. The processing of big data can be completed through the deep learning framework, and the multi-level deep structure can automatically recognize and classify features with high accuracy [44]. When the training sample size is large, the deep learning framework can be carried out in both supervised and unsupervised ways.

Gesture recognition is one of the most important methods of emotion recognition, ant the others are tracking implicit parameters, voice recognition, facial expression recognition, vital signals, and hybrid methods [2]. The main aim of gesture recognition is recognizing the human gestures in order to interact with machines [2]. Gesture recognition has been growing rapidly recently due to the development of AI technology. More and more remarkable achievements have been made in the field of new natural human-computer interaction, and new somatosensory interactive devices such as Leap Motion and Kinect have remarkable functions [32].

The effectiveness of gesture detection and tracking algorithms in gesture interaction technology is increasing due to the improvement of gesture data collection, but there are still some difficulties in education applying. Most of body gestures such as hand gesture are not feasible or feasible with a low probability in online teaching, especially in real-time online teaching. In interpersonal communication, human individuals usually use a variety of patterns for behavior recognition. Therefore, human-computer interaction should also use multiple modes, rather than one mode, such as face or gestures and body movements. HMMs and ANNs have been extensively used to address the gesture recognition problems in recent researches [2].

Single recognition algorithm has shortcomings, but the collection of different types of data is feasible, so researchers tend to use a variety of recognition technologies through information fusion [33]. A hybrid recognition framework might become an accurate and suitable method, it integrates diversified information such as facial feature, gesture, text and self-report. A multimodal emotion recognition system was proposed, it could recognize the parameters defined by HMMs and ANNs, such as the set of facial expressions and upper body gesture.

In conclusion, learners' behavior is essential for online real-time teaching. But, the learners' gesture and behavior have to be recognized at first. According to the existing latest work shown in TABLE 2, artificial intelligence and deep learning algorithm broadens the research of gesture recognition in the field of education. Based on the hybrid methods, this paper will propose a method to recognize online learning behavior through gesture recognition algorithm (as shown in Figure 2).

III. METHODOLOGY

The online teaching video experiment will be introduced in the following section. This experiment uses an open source

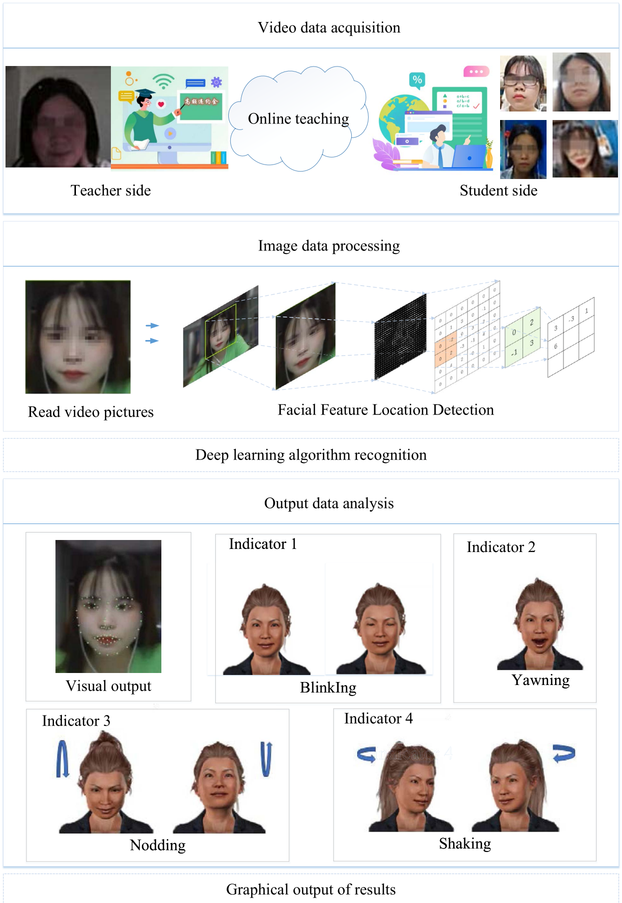

FIGURE 6. Technical flow chart, shows the head gesture recognition process.

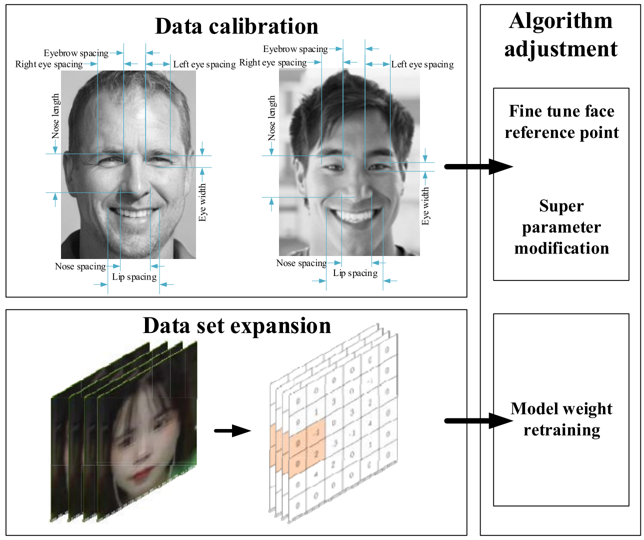

FIGURE 7. Process of algorithm optimization. The process of algorithm optimization is completed by data calibration and data set expansion.

algorithm to extract the subject's head gesture data, so as to analyze the online learning behavior characteristics. The figure illustrating of the method is shown in Figure 3.

RQ1: How to extract and recognize online learning behavior through gesture recognition algorithm?

A. ONLINE EXPERIMENT DESIGN

At present, gesture recognition algorithm is rarely used in Chinese online teaching platform, because it involves privacy and copyright issues. Online teaching experiments were designed and conducted, while online learning behavior data were collected and analyzed for scientific research purposes. Before the experiment, make sure that the subjects already know the experiment arrangement and precautions, and obtain the approval of video use. The experiment includes two parts: online teaching video recording and behavior data analysis. After the experiment, the subjects were also interviewed about their online teaching experience.

Several commonly used teaching platforms were tested in the pre-experiment, but few of them could meet the requirements of teaching, experimental recording and data extraction at the same time. After comparison, Treenity and DingTalk are finally selected. The screenshots of the platform are shown in Figure 4 and Figure 5. Four volunteers participated in the pre-experiment. Based on the research ideas in Figure 3, the whole experiment was completed and the data were extracted and analyzed. Facial expressions related to learning state were analyzed, including confusion, depression, resistance, smile, and the head gestures, including blinking, yawning, nodding, shaking head, long-term rest and leaving. After data processing, the results showed that there are significant differences in expression and gesture. The recognized facial expressions are few and not obvious, but there is more head gesture. This confirms the research results of Imani and Montazer's that online learning may not produce many emotional fluctuations [2]. The reason may be: a. Changes in the teaching scene: "Spatial-temporal separation" led to reduced participation and facial expression; b. Different in learning environments: Online teaching students are more likely to be disturbed by external factors, because they may have classes at home, dormitory, library and other places. In addition, there are more behaviors unrelated to learning, such as watching mobile phones, talking to others, leaving the screen, etc. The experimental design was optimized through pre-experiment. According to the characteristics of the subjects, the recognition algorithm is corrected, and the method of this study is finally determined.

In the first stage of the experiment, a remote live video course was conducted through an online teaching platform, in which teachers and students were located in different locations. The course content is professional content related to leadership. The subjects were 34 volunteers, mainly business administration students, who had more than two months of online learning experiences during the pandemic. In the experiment, the subjects need to turn on the camera all the time, and the experimental staff will record the course through the video recording software.

In the second stage of the experiment, facial expression and head gesture were recognized and estimated. They are important ways of emotional expression in learning state.

For example, head posture could indicate the object of communication. Nodding is usually a positive feedback to the speaker, and it suggests that the listener is listening attentively. Shaking head usually represents opposition. If students are tired, they may close their eyes, yawn and blink. Therefore, the recognition and estimation of head gesture is of great significance for analyzing students' mental states such as attention and attitude, which can be used in online teaching.to understand learners' intention, attitude and other psychological states.

In conclusion, blinking, yawning, nodding, shaking head and leaving were used as indicators of gesture recognition.

B. ACTION RECOGNITION ALGORITHM

According to Imani and Montazer, there are two types of facial recognition methods, one is judgment-based methods on message communicated with facial expression; and the other is sign based methods where occurring frequency and types of movements describe a behavior. How long a face movement lasts or how many times face moves leads to an emotion detection [2]. This paper adopts the second method.

A learner may not change his/her face frequently in online teaching, and it will be difficult to collect learners' facial emotion. Facial and head images with high precise can be obtained due to the webcams located in front of the face. So facial feature and head gesture recognition is suitable for using in e-learning context.

The head gesture recognition algorithm used in this study is the combination of facial feature recognition code and head pose recognition code based on histogram of oriented gradients. Based on the above literature review, the selection of head gesture recognition technology needs to be judged according to different situations. As an exploratory study for the first time, considering the feasibility and convenience of application in the field of education, we choose an algorithm with a definite framework, open source and easy to share.

The face position detector and facial feature position detector used are detection algorithms based on HOG implemented by Dlib library, and OpenCV is used for image processing of video interception. Among them, Dlib is a classic open source database for image processing. OpenCV is an open source cross platform computer vision and machine learning software library [37].

HOG algorithm is an image detection algorithm proposed by Dalal & Triggs in 2005 [36], which is widely used because of its good edge sensitivity. The face appearance changes modeled by HOG represent the appearance features. It should be noted that the HOG algorithm must adjust the image scale to ensure that the horizontal to vertical ratio is 1:2. After image preprocessing, it is necessary to calculate the horizontal and vertical gradients of pixels to obtain their magnitude and orientation. These data constitute the value of each cell in the gradient histogram. The cell size of image segmentation is determined by the detection requirements. After normalizing the histogram to prevent it from being affected by brightness, the detection model based on HOG algorithm can be obtained

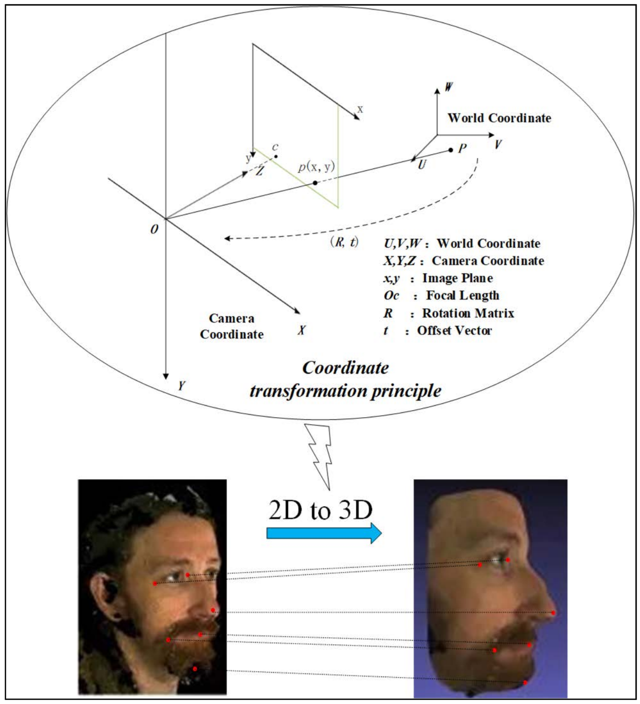

FIGURE 8. Schematic diagram of 2D point and 3D point conversion processing.

for model training. The specific recognition process is as follows in Figure 6.

As the existing face recognition algorithms are relatively mature, HOG algorithm can be calculated using OpenCV library, so this paper calls two functions of Dlib library for detection: dlib. get_ frontal_ face_ Detector () and dlib. shape_ predictor (predictor_ path). The former is the pyramid hog algorithm for facial position detection, and the HOG algorithm is the superposition of multi-layer hogs, while the latter is used for regional facial feature position detection. Most of the existing databases are European and American, in order to improve the algorithm recognition performance in Chinese samples, we add the yellow face and gesture feature data, and modify the algorithm feature parameters through training and testing. The process of algorithm optimization is completed by data calibration and data set expansion as shown in Figure 7.

After using OpenCV to process the video frame by frame, this study first detects the face position of the object in the picture. If there is no return value, it indicates that the detected object is not facing the camera.

After detecting the face position coordinate data of the object, the 2D face key points can be extracted, and the head gesture change of the object can be analyzed through algorithm transformation. The general steps of the classical head gesture algorithm are as follows:

① 2D face key point detection;

② 3D face model matching;

③ Solve the conversion relationship between 3D points and corresponding 2D points;

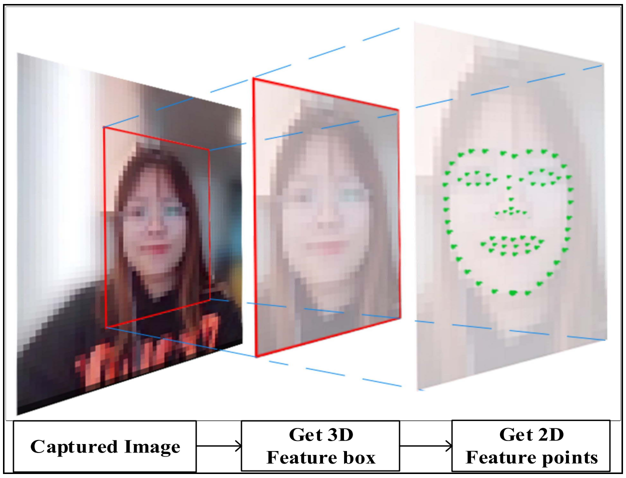

FIGURE 9. Recognition effect of facial features. The red frame line is the 3D facial feature extraction, and the green dot is the facial features extraction feature.

④ The Euler angle is solved according to the rotation matrix.

The schematic diagram of 2D point and 3D point conversion processing is as follows in Figure 8.

Face coordinate transformation needs to involve the world coordinate system, camera coordinate system, image coordinate system and pixel coordinate system. If the camera does not have distortion then the image coordinate system can be omitted. From the world coordinate system to the camera coordinate system, it involves the rotation and translation of the subject. The corresponding rotation matrix is obtained by rotating different angles around different coordinate axes, as shown in (1):

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = R \begin{pmatrix} U \\ V \\ W \end{pmatrix} + T = (R \mid T) \begin{pmatrix} U \\ V \\ W \\ 1 \end{pmatrix} \tag{1}$$

From camera coordinate system to image coordinate system, it is from 3D to 2D, which belongs to perspective projection relationship. The calculation formula is shown in (2):

$$s \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \tag{2}$$

where, f is the perspective projection relation between camera coordinate system and image coordinate system.

The pixel coordinate system and the image coordinate system are both on the imaging plane, but their origin and units of measure are different. The image coordinate system origin is the intersection point of the camera's optical axis and the imaging plane, usually it is the midpoint or principal point of the imaging plane.

The image coordinate system unit is mm, which belongs to the physical unit. The pixel coordinate system unit is pixel, and we usually describe a pixel point as several rows and columns.

TABLE 3. Part of the raw identification data of 3 random subjects.

Subject 1		Subject 2		Subject 3	
Time(s)	Behavior	Time(s)	Behavior	Time(s)	Behavior
0.843	leaving	29.710	nodding	1.572	leaving
1.085	leaving	29.712	shaking	1.814	leaving
1.325	leaving	29.878	shaking	2.0555	leaving
1.572	leaving	30.061	blinking	2.297	leaving
1.814	leaving	30.195	nodding	2.543	leaving
2.055	leaving	30.276	shaking	2.789	leaving
2.297	leaving	30.643	shaking	3.034	leaving

TABLE 4. The behavior data of one subject after data statistics.

Gesture	TI1	TI2	TI3	TI4	TI5	TI6	TI7
Yawing	55	97	103	66	49	152	123
Nodding	89	149	116	68	88	0	1
Shaking	33	23	45	2	27	10	12
Leaving	28	57	84	76	110	11	56
Blinking	30	16	39	22	33	171	24

TI=Time Interval, during the total 35 minutes online teaching class, each 5 minutes represents a time interval, denoted by i(i=1,2,3,4,5,6,7). The first-time interval represents 0-5 minutes, the second represents 6-10 minutes, and so on.

So, the conversion the unit between the image coordinate system and pixel coordinate system is shown in (3):

$$1pixel = dx\,(dy)\,mm \tag{3}$$

where dx and dy indicates how many mm each column and row represent.

After obtaining the Euler Angle through transformation relationship, head behavior and gesture can be analyzed. Where, Euler Angle refers to the rotation angle of the object around the three coordinate axes of the coordinate system, namely Pitch, Roll and Yaw. Whether the subject is shaking head or nodding can be detected based on changes in three parameters. Ideally, the emotional fluctuation of the subject can be predicted according to the size of the numerical fluctuation.

The detection of blinking and yawning requires more detailed recognition of the position of facial feature points, compared with the simple actions of nodding and shaking head. But, both sides have the same point, that is the algorithms all use double threshold judgment in this study. If the data features of leaving, nodding, shaking head, blinking and yawning appear in this frame, it needs to be judge in combination with the data of the previous and subsequent frames, and then it will be regarded as a characteristic behavior.

Taking yawning behavior as a case, the algorithm detection process of our study is as follows:

① Define the Euclidean distance of lip length and width, and calculate the aspect ratio by the number of feature points.

② If the aspect ratio in the detected picture is greater than the initially set threshold, the counter is incremented by one.

③ In the process of detection and recognition frame by frame, the value of the counter is continuously counted. Similarly, if the value is greater than the initially set threshold, it is considered to have yawned. If a certain number of frames is lower than the threshold or the counter does not change

TABLE 5. **The abnormal data.**

Gesture	TI1	TI2	TI3	TI4	TI5	TI6	TI7
Yawing	4	96	0	0	0	0	0
Nodding	4	53	0	0	0	0	0
Shaking	4	29	0	0	0	0	0
Leaving	362	0	481	509	468	602	419
Blinking	9	46	0	0	0	0	0

TABLE 6. **The descriptive statistical results.**

Gesture Type	Identification Criteria	Number of times
Blinking	Take 24 frames of normal video as 1 second, and the duration of once eye closure is less than 3 frames. When the aspect ratio is less than 0.2 and more than 3 frames, it is considered that the subject closes his eyes once.	15997
Yawning	When the aspect ratio of mouth opening is greater than 0.5 and more than 3 frames are detected within 1minute, it is considered that the subject yawn once.	12122
Nodding	When the pitch angle threshold of head up and down movement is 0.3, the subject is considered to nod once.	10061
Shaking	When the roll angle threshold of head left-right movement is 0.5, the subject is considered to shake his head once.	14328
Leaving	When the algorithm fails to capture the characteristics of the subjects' head, the subject is assumed to leave.	34270

(considering the loss problem), the counter will be cleared before it is greater than the threshold.

④ Significantly, the counter and the whole algorithm will not stop working, no matter what kind of situation of reset in process ③.

After the whole video is detected, the algorithm will output the time sequence and statistical values of five behavior features. Significantly, the algorithm needs to detect the facial position first in the flow chart, then recognize the position of facial feature point. However, in the process of behavior detection and analysis, there is no difference in the detection of yawning, blinking, shaking head and nodding. The detection is carried out at the same time, and counters are distinguished from each other. The final recognition effect is shown in Figure 9.

Multimodal systems can achieve higher recognition rate, compared with separate speech or visual systems [38]. Meza Kubo *et al.* proposed an emotion recognition method, using qualitative analysis, questionnaire survey and additional readme to improve the recognition accuracy [39]. Therefore, we take a variety of methods, for ensuring the accuracy of the attitude recognition algorithm proposed in this paper.

First, we use mature and accurate open source algorithms. The algorithm comes from the behavior recognition of the

TABLE 7. **Decription of statistical results for all samples.**

	Min	Max	Mean	St. Deviation	variance
Blinking	9	2147	648.53	530.588	281523.711
Yawning	3	3432	487.94	711.414	506109.815
Nodding	0	2071	436.91	515.183	265413.174
Shaking	4	1427	549.47	441.512	194932.499
Leaving	4	6731	1386.82	1844.677	3402831.968
Total	167	7445	3484.03	2006.956	4027871.242

cockpit, it is similar to the online teaching scene and has high mobility. Secondly, we provide enough samples for training and comparison, to improve the recognition accuracy and reduce the training cost. By the way, we added the facial feature set of yellow races to ensure the adaptability to the research content. Third, in data processing, we identify and extract all samples in the whole time period, and manually calibration to compare the identification results. In addition, in the experiment, the subjects were asked to use computers or mobile devices with higher resolution, and set a single screen background, to reduce interference factors and improve the recognition accuracy.

C. ACTION FEATURE EXTRACTION AND ANALYSIS

Blinking, yawning, nodding, shaking head and leaving are the actions that need to be recognized in this study. The parameters are as follows:

1) BLINKING

The number used to identify human eye feature points is p_ (37-42) and p_ (43-48) among 68 feature points, the equation for calculating the eye aspect ratio is shown in equation (4).

$$eye = \frac{\|p_{38} - p_{42}\| - \|p_{39} - p_{41}\|}{2\,\|p_{37} - p_{40}\|} \tag{4}$$

When eyes open, eye fluctuates up and down at a certain value; When eyes closed, eye drops rapidly and will be close to zero in theory. Therefore, the definition of one blinking is the eye lower than a certain threshold, that is, the area of pupil covered by eyelid is at least 80%. Under normal circumstances, the time to close eyes at one time is between $0.2 \sim 0.3$ seconds.

This paper considers that the detector has closed his eyes once: take 24 frames of normal video as 1 second, and the duration of one eye closure is within 3 frames. When the aspect ratio is less than 0.2 and more than 3 frames.

2) YAWNING

The amplitude of mouth opening is calculated using the same European distance formula as blinking. A yawn is defined when the aspect ratio is greater than 0.5 and more than 3 frames are detected within 1 min.

3) NODDING

In the algorithm, the Euler angle of students' facial gesture is recognized. According to the head movement process, the heading angle is approximately (−40.9 °, 36.3 °), the pitch angle is approximately (−60.4 °, 69.9 °), and the roll angle is approximately (−79.8 °, 75.3 °). [1] When students doze off or nod, the pitch angle will fluctuate. When the threshold is set to 0.3, it is considered that students have nodded once.

4) SHAKING

Like nodding, when students shake their head, the rolling angle will fluctuate. When the threshold is set at 0.5, it is considered that students shaking head once.

5) LEAVING

Students turn their heads to look out of the screen, or talk to others, leave completely, or stay too far away, which will make the facial recognition algorithm unable to capture. Therefore, when the algorithm cannot capture, it is deemed that the students leave.

IV. RESULTS

The experimental data includes two types. One is the video data recorded by DingTalk, including the complete teaching process and recorded video; The other is the online learning record on the Treenity, including attendance record, online interaction, homework release, final score and so on. This section will discuss the accuracy of the algorithm, data characteristics and analysis results in turn.

A. ALGORITHM ACCURACY

Firstly, we verify the availability and reliability of the algorithm. In order to calculate the accuracy of the online teaching scene of the recognition algorithm, we manually check the recognized data to screen out obvious abnormal data. The experiment lasted 35 minutes, taking 5 minutes as a time interval, the number of different behaviors of different students in each time interval was counted. Finally, the data were summarized and analyzed. We randomly selected three subjects, the original recognition data of the subjects are shown in Table 3. Based on the analysis of the original data, the behavior data of one subject is shown in Table 4, and the abnormal data is shown in Table 5.

Manually detect the data and calculate the availability and reliability of the algorithm. The accuracy index is used, and the specific formula is as follows:

$$Accuracy = \frac{TP + TN}{P + N} \tag{5}$$

In the data obtained in this paper, students' actions are set as Positive samples (P), and no actions are set as Negative samples (N). The action recognition true is set as True Positives sample (TP), and the action recognition error is set as False Negatives sample (FN). If there is no action and the recognition is no action, it is set as True Negatives sample (TN). If there is no action but the action is recognized, it is set as False Positives sample (FP).

According to statistical comparison of data, P is 896, N is 74, TP is 603, FN is 293, TN is 74, and FP does not exist. Through calculation, the accuracy of the behavior recognition algorithm in the application scenario of this paper is good, which is 0.698.

B. DESCRIPTIVE AND INFERENTIAL STATISTICS

Two experiments were conducted in this study, in which 34 students participated. After processing the video data in the two experiments, we found that the subjects did not recognize any behavior at some time and deleted it. The number of final analysis samples was 29. The sample information was analyzed by IBM SPSS statistical analysis software [40], [41]. TABLE 6 is the statistical results of five gestures data of 29 samples.

Among the 29 subjects, 24 were female, accounting for 82.76%, and 5 were male, accounting for 17.24%. According to the study level, junior students accounting for the majority (15, 51.72%). The data description statistical analysis results are shown in Table 7. In terms of the total number and average of each movement indicator, leaving was the most recognized among the five movements, followed by blinking and nodding the least.

C. DATA ANALYSIS RESULTS

The specific features are shown as follows (RQ2):

RQ2: Which gestures occur frequently in online teaching?

1) ALGORITHM BASED ON DEEP LEARNING IS FEASIBLE

The data results show that the head gesture actions of online learning are common and differentiated. Through the online teaching experiment, 118457 head movement data were collected. After data cleaning, a total of 86778 valid data were obtained. The attributes of the data include sample number, time, gesture type and quantity.

The type, recognition standard and recognition quantity of gesture are recorded in Table 6. It can be seen that using the algorithm proposed in this paper to recognize the head gesture in online teaching is feasible. In order to verify the rationality of the algorithm, we also manually reviewed the experimental video, and intercepted the static picture of head gesture frame by frame decomposition to calculate the occurrence frequency. Due to the large amount of manual review data, we randomly selected three gestures (leaving, nodding and shaking head) of 6 subjects for manual review. The review results show that the frequency of leaving is significantly higher than that of nodding and shaking head, and the time period of action concentration is consistent with that of algorithm recognition, that is, the results of algorithm recognition and manual review are basically consistent. Therefore, the algorithm proposed in this paper is applicable and feasible.

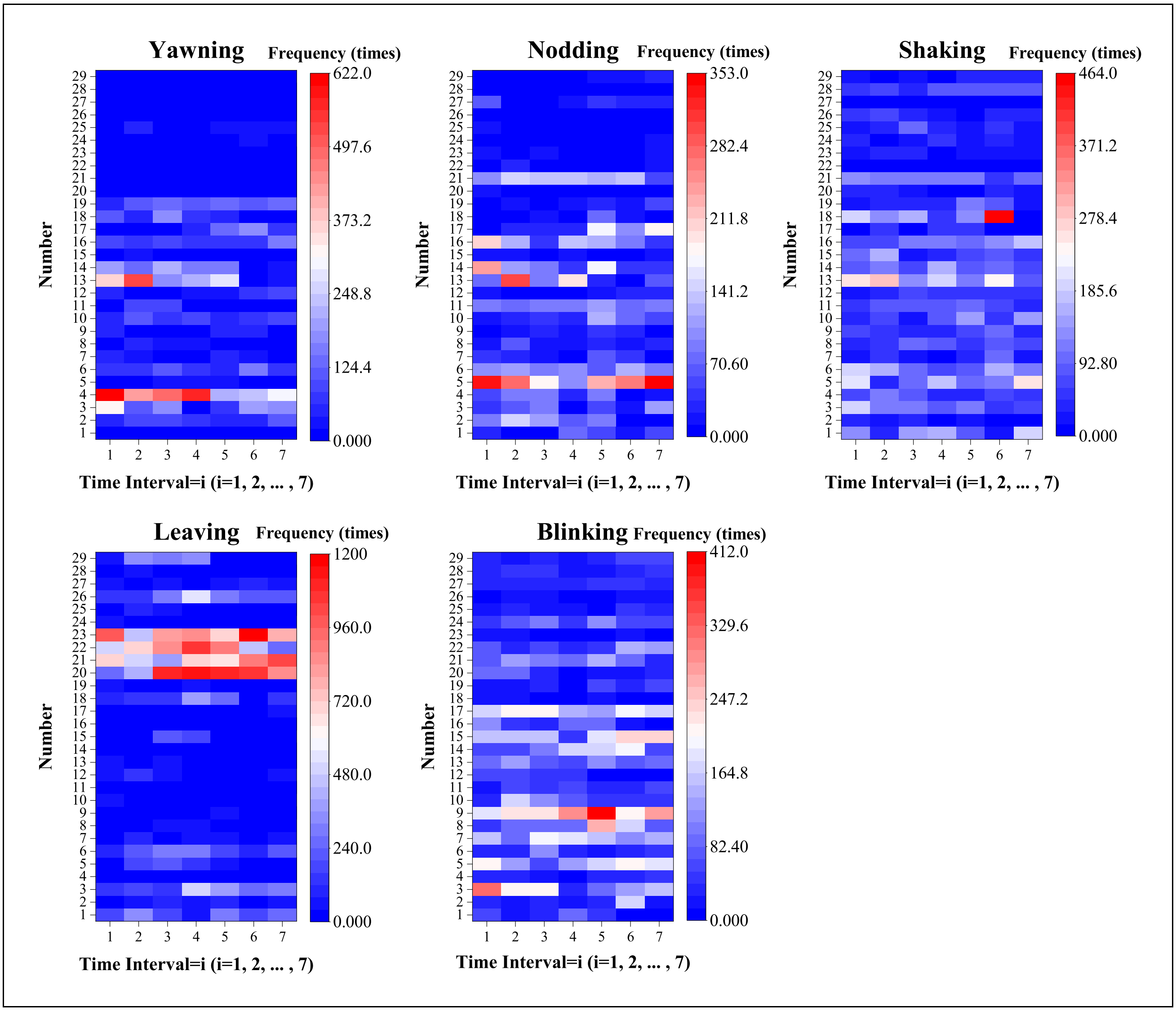

FIGURE 10. Heat maps of 5 different gestures, Time Interval, during the total 35 minutes online teaching class, each 5 minutes represents a time interval, denoted by i(i = 1, 2, 3, 4, 5, 6,7). The first-time interval represents 0-5 minutes, the second represents 6-10 minutes, and so on.

2) VOLUME OF UBIQUITIOUS HEAD GESTURES DURING ONLINE TEACHIG

All subjects can be identified with five gestures in the experiment, according to the data of Table 5. But the number of identified gestures of subjects varies greatly. According to statistics, subject NO.5 was recognized with the most actions, 7445 times in total, and subject NO.7 was recognized with the least actions, only 167 times.

The five gestures lasted the whole teaching process. Through the manual review of the video, we found the time period for recognizing more leaving actions is basically the same as that for online testing. According to the interview after the experiment, during the teaching process, after the teacher releases the online test, the students will lower their heads to operate the mobile phone or computer to complete the test. It causes their faces and heads to overstep the range set by the algorithm, so they are recognized as leaving. It can also explain why leaving is the most recognized gesture.

According to the interview results, we also found that students generally conduct online teaching learning in dormitories, homes or other places. Compared with classrooms, students in these places are more vulnerable to external interference, and their learning attention is more easily distracted. We drew the thermal diagram (Figure 10) and analyzed the five gestures. In Figure 10, the five diagrams show the distribution of the five gestures respectively. The horizontal axis represents the time interval, the vertical axis represents different subjects, red represents the high contribution value, blue represents the low contribution value, and white represents the intermediate value.

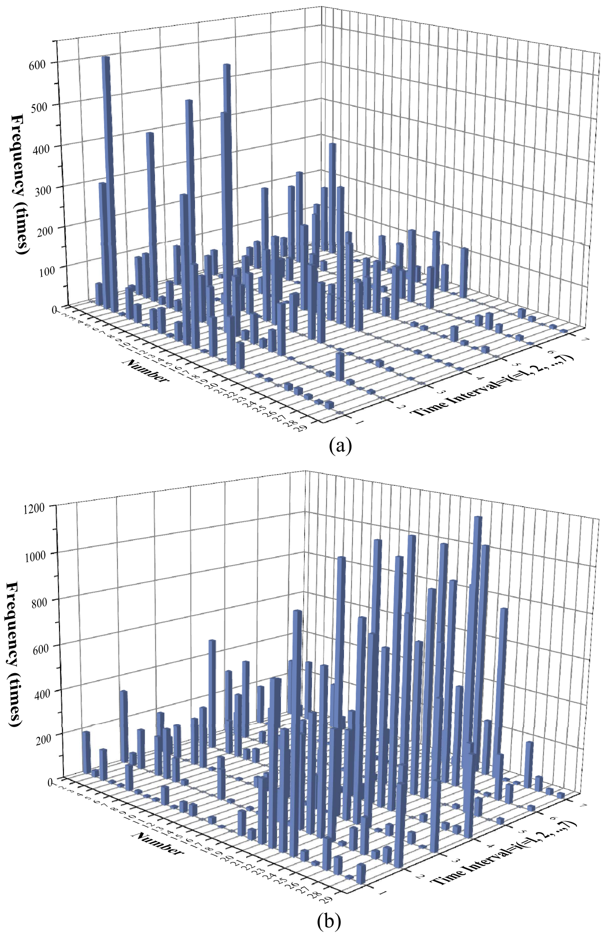

FIGURE 11. **a. The changing trends of yawning of all subjects. b. The changing trends of leaving of all subjects.**

3) EASE OF TURNING ON VIDEO MAY ENCOURAGE LESS HEAD GESTURES

In real-time video teaching, turning on the camera helps to restore the scene of offline teaching. Subjects said that opening it made them feel supervised by teachers and students, so they would pay attention to their behavior. Teachers also said that online teaching with video is better than that without video, which can reduce students' negative behavior and improve the teacher-student interaction rate. Subjects reported that the effect of video opening was more effective in small-scale classes than in large-scale classes.

According to the survey results of Jiang and Wu [5], [6], live video and interactive discussion are the most popular online teaching modes for college students. Compared with the whole video live broadcast and the whole audio live broadcast, students prefer the former [6]. Our data also verified this point. The subjects thought that live video is closer to offline teaching, they will focus more on the course and participate in more interaction. However, some subjects said that long-time video will lead to their eyesight and mental fatigue, and will be disturbed by more external factors in the learning process. This is consistent with the research results of Wu and Xie [5], [42]. Thus, online real-time video teaching helps to improve the effect of online education, but it also has its scope of application. For example, live video may be more suitable for small class teaching.

4) THE TREND OF HEAD GESTURES IS CORRELATED WITH TIME

According to the research of Casile and Giese [43], there is a correlation between gesture change and time, which is very important in emotion recognition. The experimental results of this study show that, the trend of gesture change of different subjects with time is basically the same, and this trend shows it decreases first and then increases. In teaching interaction and testing, less actions were recognized, and the subjects showed a high level of concentration. The time distribution of blinking, yawning and leaving is relatively concentrated, and the trend of gesture change with time also tends to be consistent. For example, a variety of actions occur more frequently, in the first 5 minutes and the 25th-30th minutes.

Through the analysis, we found that the time distribution of different subjects' gesture and action is consistent. In the first 5 minutes, the recognized gesture showed an upward trend and then began to decline, indicating that the students' attention was not focused at the beginning of the course. From the fifth minute, the number of recognized gestures shows a downward trend and remains stable, which means that students gradually enter the learning state. However, students' concentration inevitably declines again with the progress of the course. At the end of the course, the number of various gestures is relatively less, and the students are in a state of concentration and excitement. The results basically accord with the attention concentration curve, that is, attention shows a trend of first rising and then falling.

Taking the data of leaving and yawning as an example, the variation trend of leaving and yawning of all subjects is shown in Figure 11. Yawning occurred mainly in the early stages of the experiment, especially between 0 and 15 minutes. While leaving occurred mainly in the middle and late period, especially in the 15-25 minutes. Combined with the interview after the experiment, in the initial stage of the course, due to computer hardware failure, network delay, environmental interference and other factors, some subjects will have a delay in entering the normal learning state, and the data shows that there are many gestures and actions recognized. At 15-25 minutes, leaving showed a obvious upward trend and showed a peak, while other movements showed a downward trend. Combined with the manual video review, it was found that the subjects did not face the camera all the time when they participated in online tests, such as answering questions and voting, so they were recognized as leaving by the algorithm.

V. DISCUSSION

Based on the above research, we found that subjects' behaviors in online teaching are diversified, and also showed unique behavior characteristics and rules. In the following sections, we will continue to explore the factors influencing these characteristics, and provide suggestions for online teaching practice and teaching platform optimization (RQ3).

RQ3: What factors are associated with recognizable gestures?

Our analysis suggested that intrinsic teaching characteristics, subject's intrinsic features and external factors are correlated with online learning behaviors, as discussed in detail below.

A. INFLUENCE FACTORS

1) INTRINSIC TEACHING CHARACTERISTICS

According to the prior researches [19], [20], and [21], the teaching mode, teaching time, teacher's curriculum design, platform functions, the speed and stability of the network and so on, are all related to the effect of online teaching. Our experimental results verify the influence of online teaching mode and teaching time on subjects' online learning behavior.

Base on the former results, the subject's head gesture and behavior are related to teaching time, showing a general trend of first decline and then increase. The subjects said that it's harder to enter the learning state in online teaching than offline teaching, and they will debug equipment or software at the beginning of the course, which would affect their attention.

Among different online teaching modes, live video teaching is the most popular among students, and it is easier for subjects to focus on learning compared to the other modes. The subjects also said that live video was the closest to offline teaching, and they would control their behavior under the supervision of teachers and classmates, so as to improve their attention. However, some students still have objections to open the camera. They mentioned that they would not open the camera voluntarily unless required by the teacher.

More head gesture during online teaching means lower relevance and engagement. Subjects frequently mentioned that they behave more gestures and actions during online teaching when they find irrelevant or have lack of interest. Our experimental records show that subjects with more head gesture engage in online interactions less frequently and at a lower speed than those with less movement.

"I tend to behave more gestures in long classes online, because longer classes online make me feel tired, and it is difficult to concentrate on the classes all the time." (Subject 2)

"Sometimes subjects lose their concentration due to loss of interest or low relevance." (Subject 3)

"When I am not interested in the part of the class, I start multi-gesture." (Subject13)

2) SUBJECT'S INDIVIDUAL CHARACTERISTICS

Online learning puts forward higher requirements for autonomous learning ability. More than 60% subjects agreed that online learning needs more self-discipline and needs to develop good online learning behaviors and habits. During the survey after experiment, several subjects said "When I'm alone in my bedroom for online teaching, it's hard to concentrate all the time." "For me, it's hard not to go to other websites or not to talk to friends online".

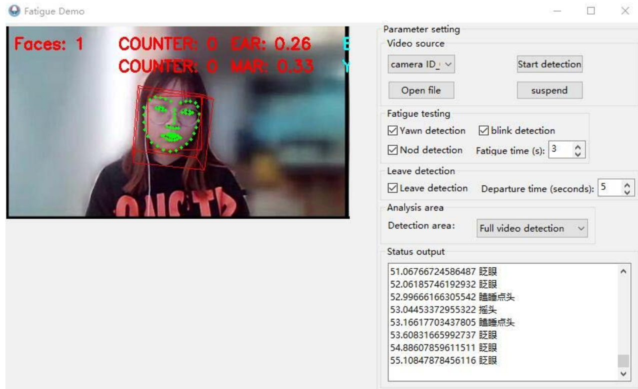

FIGURE 12. **Gesture recognition and feedback module design.**

Unfamiliarity with teaching platforms and tools leads to more invalid behavior. The head gesture identified most by the algorithm was leaving. From the survey results, part of the leaving behaviors was caused by the switch between two learning platforms and the unfamiliarity with the platform in online test. Our manual review of the video data also confirmed this.

3) EXTRINSIC FACTORS

Subjects behave much head gesture in online teaching due to external distractions. Subjects' attention gets attracted by extrinsic factors easily. For instance, two major classes of distractions are the speed and stability of network, and the studying environment. As studying moves online, subjects interact with digital tools more than they used to, and they mentioned that platform designs can be the cause of invalid behavior, especially pop-ups.

"It's hard for me to concentrate because QQ and WeChat are popping in" (Subject 5)

"During COVID19, once I studied online, too much noises interrupted me, the interruption came from my mom, my little brother." (Subject 11)

According to Wu's nationwide survey [5], more than 50% of students believed that the main problems of online teaching are the speed and stability of the network. During the period of COVID-19, large-scale online education has brought great opportunities to platform as well as great pressure, even the system platform crashed and failed to work normally. Furthermore, Hofer (2021) pointed out that students' knowledge, skills, and attitudes, together with their learning activities all influence the online learning.

According to the above, the multi-gesture behavior in online teaching is the result of the comprehensive effect of various factors. Therefore, in order to reduce students' ineffective behaviors and improve the teaching effect, these factors need to be considered and optimized. In the following section 5.2 and 5.3, we will present several practice guidelines for online teaching and design implications for better support of online teaching software.

B. PRACTICE GUIDELINES FOR ONLINE TEACHING

Offline teaching scenarios can be quickly and truly reproduced online, but there are still many problems to be solved. Based on the researches above, the following suggestions are intended for teachers to develop online teaching.

Firstly, choose the most suitable teaching mode according to the needs of the course. If the course is important and has high requirements on the content, live broadcast or video mode will be a better choice.

Secondly, the course design should consider the needs of most learners. The design of interaction and test should attract the attention of learners and maintain the continuity of their attention. Teachers also need to balance interaction and testing, because moderate interaction and testing are conducive to maintain learners' attention, but too much may also cause too much unfocused behavior.

Thirdly, let students prepare for class in advance, and attract their attention through course introduction setting to accelerate the speed of entering the learning state. Try to use the same platform or tool for the same course to reduce switching frequency.

C. DESIGN IMPLICATIONS FOR ONLINE TEACHING SOFTWARE

In online teaching, the image recognition function for gestures images has not been developed widely yet, and the human-computer interaction performance is not good enough, lacking of physical interaction [3]. The use of gesture recognition and visual function can make up for the lack of human-computer interaction, which will be helpful to improve the intelligence of the online teaching platform.

Firstly, create gesture recognition and feedback module in the teaching system. Like the page in Figure 12, the function of this module is detection and output of learning state. Fatigue and leaving detection are carried out through deep learning algorithms. Feedback the learning state of subjects to teachers and students respectively. According to the feedback teachers can make teaching decision and adjust the course design. Students may receive warnings from the auxiliary system or teacher, if they are not concentrating in the course.

Secondly, optimize the response system to display students' Q&A on the screen in real-time. For example, the bullet screen function, which is convenient for teachers to collect students' opinions, providing students with an interactive learning experience.

Thirdly, optimize of the online teaching effectiveness evaluation. The existing online teaching evaluation research is still in its early stages, and evaluation indicator and system are few. Online teaching evaluation in the future should add the behavior and attitude evaluation model. These evaluation indicators can be achieved through the collection and analysis of the learner's online expressions, gestures, behaviors and other data in the human-computer interaction.

VI. CONCLUSION

The authors have rich experiences in online and offline teaching. Based on our working experiences, we found the behavior differences between online and offline learning. Therefore, we collected and analyzed online learning behavior based on gesture recognition algorithm, in order to analyze the teaching effect of online learning, and present useful method for future online learning research. Furthermore, the conclusion provided some suggestions for improving online teaching methods.

There are diversified behaviors in online learning, which are related to learning state and time. This study obtained data of five high-frequency online learning behaviors, including blinking, yawning, nodding, shaking head and leaving. Teaching features, students' personal characteristics and learning environment have a comprehensive impact on online learning behavior.

This study focused on online real-time teaching scene. It expanded the application of artificial intelligence algorithm in online teaching, and made it possible to explore the behavior and algorithm. The online learning behavior data and conclusions obtained in this paper will significantly promote the future online teaching research. Firstly, we collected behavior data and analyzed the feedback effect of online real-time learning, providing a new path for the future education research; secondly, the conclusions provide a data basis for personalized learning and online teaching design in the future. It also helps to enrich online teaching evaluation methods and indicators, and has significance for the construction of online teaching framework.

Further investigation still needs to do in the future. The research samples in this paper are Chinese college students. We modify the algorithm parameters according to the sample characteristics, but a large number of samples are still needed to train the recognition model in the future. However, it is unknown whether the conclusions and algorithms are applicable to other regions. The experiments and algorithms in this study are reproducible, therefore, the samples could be expanded in the future. The relationship between online learning behavior and learning effect, as well as the difference between online learning behavior and offline, need to be further studied. The authors are currently continuing this study, collecting interaction and test data during the experiment, to predict the online learning effect.

The facial recognition privacy debate also exists in education field. In order to avoid controversy, all the raw data and recognition results in this paper are only used for scientific research. Artificial intelligences are changing students' online learning behavior, and they are playing games constantly around privacy. Human beings always have options and initiatives in front of technology, and students could also choose to apply or resist technology [44]–[46]. Sufficient information could help teachers improve the effect of online teaching, but invasion of privacy may cause hostility from students. The head gesture recognition used in this paper, involves fewer facial features and pays more attention to the amplitude and angle of action. We also desensitize the image, which is conducive to the promotion and application in the future. According to the research of Acquisti *et al.* [47], [48], although the need for privacy is general, individuals' privacy judgment and preferences are highly context-dependence. Therefore, whether human-computer interaction based on video recognition can be widely used in online teaching remains to be further studied.

ACKNOWLEDGMENT

The authors are grateful to Huang Yanya, Cheng Bo, and Su Guowei for their help in writing this paper.

REFERENCES

[1] N. Saberi and G. A. Montazer, "A new approach for learners' modeling in e-learning environment using LMS logs analysis," in *Proc. 6th Nat. 3rd Int. Conf. E-Learn. E-Teach.*, Tehran, Iran, Feb. 2012, pp. 25–33.

[2] M. Imani and G. A. Montazer, "A survey of emotion recognition methods with emphasis on e-learning environments," *J. Netw. Comput. Appl.*, vol. 147, Dec. 2019, Art. no. 102423.

[3] M. P. A. Murphy, "COVID-19 and emergency elearning: Consequences of the securitization of higher education for post-pandemic pedagogy," *Contemp. Secur. Policy*, vol. 41, no. 3, pp. 492–505, Jul. 2020.

[4] *Introduce the Relevant Situation of Online Education in Colleges, Middle Schools and Primary Schools During the Epidemic Period and the Next Work Considerations*, Ministry of Education of the People's Republic of China, Xicheng, Beijing, China, vol. 14, May 2020. [Online]. Available: http://www.moe.gov.cn/fbh/live/2020/51987/

[5] D. G. Wu and W. Li, "The stage characteristics of large-scale online teaching in Chinese universities—An empirical study based on a questionnaire survey of students, teachers and academic staff," *J. East China Normal Univ. Educ. Sci.*, vol. 7, pp. 1–30, Jul. 2020.

[6] Q. H. Jiang, "Investigation on the effect of online teaching in colleges and universities under COVID-19 pandemic-based on the online survey data of ordinary undergraduate colleges and universities," *J. Yangzhou Univ. Higher Educ. Res. Ed.*, vol. 24, no. 130, pp. 77–83, Jun. 2020.

[7] A. Aristovnik, D. Keržič, D. Ravšelj, N. Tomaževič, and L. Umek, "Impacts of the COVID-19 pandemic on life of higher education students: A global perspective," *Sustainability*, vol. 12, no. 20, p. 8438, Oct. 2020.

[8] W. Cao, Z. Fang, G. Hou, M. Han, X. Xu, J. Dong, and J. Zheng, "The psychological impact of the COVID-19 epidemic on college students in China," *Psychiatry Res.*, vol. 287, May 2020, Art. no. 112934.

[9] L. Mishra, T. Gupta, and A. Shree, "Online teaching-learning in higher education during lockdown period of COVID-19 pandemic," *Int. J. Educ. Res. Open*, vol. 1, Jan. 2020, Art. no. 100012.

[10] S. Mahmood, "Instructional strategies for online teaching in COVID pandemic," *Hum. Behav. Emerg. Technol.*, vol. 3, no. 5, pp. 199–203, Aug. 2020.

[11] J. Eberle and J. Hobrecht, "The lonely struggle with autonomy: A case study of first-year university students' experiences during emergency online teaching," *Comput. Hum. Behav.*, vol. 121, Aug. 2021, Art. no. 106804.

[12] S. Crossley, R. Liu, and D. McNamara, "Predicting math performance using natural language processing tools," in *Proc. 7th Int. Learn. Anal. Knowl. Conf.*, Vancouver, BC, Canada, Mar. 2017, pp. 13–17.

[13] N. Diana, M. Eagle, J. Stamper, S. Grover, M. Bienkowski, and S. Basu, "An instructor dashboard for real-time analytics in interactive programming assignments," in *Proc. 7th Int. Learn. Anal. Knowl. Conf.*, Vancouver, BC, Canada, Mar. 2017, pp. 272–279.

[14] B. H. Kim, E. Vizitei, and V. Ganapathi, "GritNet: Student performance prediction with deep learning," 2018, *arXiv:1804.07405*.

[15] R. L. L. Sie, J. Delahunty, K. Bell, A. Percy, B. Rienties, T. Cao, and M. De Laat, "Artificial intelligence to enhance learning design in UOW online, a unified approach to fully online learning," in *Proc. IEEE Int. Conf. Teach., Assessment, Learn. Eng. (TALE)*, Wollongong, NSW, Australia, Dec. 2018, pp. 761–767.

[16] S. Hrastinski, "Asynchronous & synchronous e-learning," *Educause Quart.*, vol. 31, no. 4, pp. 51–55, 2008.

[17] Y. Gu, J. Hu, Y. Zhou, and L. Lu, "Online teaching gestures recognition model based on deep learning," in *Proc. Int. Conf. Netw. Netw. Appl. (NaNA)*, Haikou, China, Dec. 2020, pp. 410–416.

[18] A. Shimada, F. Okubo, C. Yin, K. Kojima, M. Yamada, and H. Ogata, "Informal learning behavior analysis using action logs and slide features in e-textbooks," in *Proc. IEEE 15th Int. Conf. Adv. Learn. Technol.*, Hualian, Taiwan, Jul. 2015, pp. 116–117.

[19] D. D. Prior, J. Mazanov, D. Meacheam, G. Heaslip, and J. Hanson, "Attitude, digital literacy and self efficacy: Flow-on effects for online learning behavior," *Internet Higher Educ.*, vol. 29, pp. 91–97, Apr. 2016.

[20] L. Mao, N. Wang, L. Wang, and Y. Chen, "Classroom micro-expression recognition algorithms based on multi-feature fusion," *IEEE Access*, vol. 7, pp. 64978–64983, 2019.

[21] D. Derkach, A. Ruiz, and F. M. Sukno, "Head pose estimation based on 3-D facial landmarks localization and regression," in *Proc. 12th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, Lille, France, May 2017, pp. 820–827.

[22] X. Guo, "Study on the analysis system of students' classroom concentration based on micro expression recognition," M.S. thesis, Dept. Comput. Eng., Yunnan Normal Univ., Kunming, China, 2019.

[23] K. Nosu and T. Kurokawa, "A multi-modal emotion-diagnosis system to support e-learning," in *Proc. 1st Int. Conf. Innov. Comput., Inf. Control (ICICIC)*, Beijing, China, Aug. 2006, pp. 274–278.

[24] K. Fujisawa and K. Aihara, "Estimation of user interest from face approaches captured by webcam," in *Virtual Mixed Reality* (Lecture Notes in Computer Science), vol. 5622, R. Shumaker, ed. Berlin, Germany: Springer, 2009.

[25] Y. Liu, "Research and application of head pose estimation method in natural environment," Ph.D. dissertation, Dept, Comput. Eng., Central China Normal Univ, Wuhan, Hubei, China, 2015.

[26] W. D. Hillis, K. H. Nishihara, S. P. Hsu, and N. Siegel, "Gesture recognition simulation system and method," U.S. Patent 7 701 439, Apr. 20, 2010.

[27] K. Vermun, M. Senapaty, A. Sankhla, P. Patnaik, and A. Routray, "Gesture-based affective and cognitive states recognition using Kinect for effective feedback during e-learning," in *Proc. IEEE 5th Int. Conf. Technol. Educ.*, Banglore, India, Dec. 2013, pp. 107–110.

[28] G. Gkioxari, B. Hariharan, R. Girshick, and J. Malik, "R-CNNs for pose estimation and action detection," Jul. 2014, *arXiv:1406.5212*.

[29] H. Y. Zhang, W. Liu, X. U. Wei, and H. Wang, "Depth image-based gesture recognition for multiple learners," *Comp Sci.*, vol. 42, no. 9, pp. 299–302, 2015.

[30] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2D pose estimation using part affinity fields," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 7291–7299.

[31] D. Mehta, S. Sridhar, O. Sotnychenko, H. Rhodin, M. Shafiei, H.-P. Seidel, W. Xu, D. Casas, and C. Theobalt, "VNect: Real-time 3D human pose estimation with a single RGB camera," *ACM Trans. Graph.*, vol. 36, no. 4, p. 44, 2017.

[32] J. Wan, "Gesture recognition and information recommendation based on machine learning and virtual reality in distance education," *J. Intell. Fuzzy Syst.*, vol. 40, no. 12, pp. 1–11, 2020.

[33] Y. Jiang, W. Li, M. S. Hossain, M. Chen, and M. Al-Hammadi, "A snapshot research and implementation of multimodal information fusion for data-driven emotion recognition," *Inf. Fusion*, vol. 53, pp. 209–221, Jan. 2020.

[34] L.-A. Perez-Gaspar, S.-O. Caballero-Morales, and F. Trujillo-Romero, "Multimodal emotion recognition with evolutionary computation for human-robot interaction," *Expert Syst. Appl.*, vol. 66, pp. 42–61, Dec. 2016.

[35] H. Gunes and M. Piccardi, "Bi-modal emotion recognition from expressive face and body gestures," *J. Netw. Comput. Appl.*, vol. 30, no. 4, pp. 1334–1345, Nov. 2007.

[36] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, San Diego, CA, USA, Jun. 2005, pp. 886–893.

[37] *OpenCV Crash Course*. Accessed: Aug. 2021. [Online]. Available: https://opencv.org/opencv-free-course/

[38] M. Song, M. You, N. Li, and C. Chen, "A robust multimodal approach for emotion recognition," *Neurocomputing*, vol. 71, nos. 10–12, pp. 1913–1920, Jun. 2008.

[39] V. Meza-Kubo, A. L. Morán, I. Carrillo, G. Galindo, and E. García-Canseco, "Assessing the user experience of older adults using a neural network trained to recognize emotions from brain signals," *J. Biomed. Informat.*, vol. 62, pp. 202–209, Aug. 2016.

[40] D. George and P. Mallery, *SPSS for Windows Step by Step: A Simple Guide and Reference 17.0 Update*, 10th ed. London, U.K.: Pearson, 2010.
[41] IBM. *IBM Support*. Accessed: Aug. 18, 2020. [Online]. Available: https://www.ibm.com/support/pages/downloading-ibm-spss-statistics-25
[42] H. M. Xie, L. C. Liu, and Y. A. Chen, "Thoughts on online teaching in universities under the background of pandemic," *China Univ. Teach.*, vol. 357, no. 5, pp. 59–62, 2020.
[43] A. Casile and M. A. Giese, "Critical features for the recognition of biological motion," *J. Vis.*, vol. 5, no. 6, pp. 348–360, 2005.
[44] Z. Qin, Y. Zhang, S. Meng, Z. Qin, and K.-K.-R. Choo, "Imaging and fusing time series for wearable sensor-based human activity recognition," *Inf. Fusion*, vol. 53, pp. 80–87, Jan. 2020.
[45] M. Muzammal, R. Talat, A. H. Sodhro, and S. Pirbhulal, "A multi-sensor data fusion enabled ensemble approach for medical data from body sensor networks," *Inf. Fusion*, vol. 53, pp. 155–164, Jan. 2020.
[46] C. Verma, V. Stoffova, Z. Illés, S. Tanwar, and N. Kumar, "Machine learning-based student's native place identification for real-time," *IEEE Access*, vol. 8, pp. 1–15, 2020.
[47] C. Verma, Z. Illes, V. Stoffova, and P. K. Singh, "Predicting attitude of Indian student's towards ICT and mobile technology for real-time: Preliminary results," *IEEE Access*, vol. 8, pp. 178022–178033, 2020.
[48] C. Verma, Z. Illés, V. Stoffová, and V. Bakonyi, "Opinion prediction of Hungarian students for real-time e-learning systems: A futuristic sustainable technology-based solution," *Sustainability*, vol. 12, no. 16, p. 6321, Aug. 2020.

LIU QIANQIAN received the Ph.D. degree from the Business School, Hohai University, China, in 2018. Since 2018, she has been a Lecturer with the Human Resource Management Department, School of Business Administration, Guangxi University of Finance and Economics. As a College Teacher, based on her rich online and offline teaching experience, she has devoted to the research on student behavior and teaching effect in online teaching. Her research interests include behavior analysis, online teaching, organization behavior, and emergency management.

WAN QIAN received the Ph.D. degree from Southeast University, China, in 2013. He is currently a Supervisor of master's student with the Guilin University of Electronic Technology. He is also the Deputy Chief Engineer of Hualan Design and Consulting Group Company Ltd., and also the Director of the Hualan Intelligent City Research Center. His research interests include transportation planning and management, ITS, and data processing and modeling.

XIE BOYA received the bachelor's degree from the Wuhan University of Science and Technology. He is currently pursuing the master's degree with the Guilin University of Electronic Technology.

LIN CHURAN received the bachelor's degree from Hohai University. He is currently pursuing the master's degree with the Guilin University of Electronic Technology. He is good at algorithm research and data analysis. He once won the first prize of the National University Provincial Transportation Science and Technology Competition.

XIE ZHENYOU received the bachelor's degree from the Guilin University of Electronic Technology, in 2019, where he is currently pursuing the master's degree in transportation engineering. His research interests include the transport planning and safety management, and deep learning algorithm mode.

PAN SHU received the B.S. degree in human recourse management from the Wuhan University of Science and Technology, China, in 2010, and the M.S. degree in administration from Guangxi University, China, in 2013. She is currently a Lecturer with the Human Resource Management Department, School of Business Administration, Guangxi University of Finance and Economics. As a College Teacher, based on her rich online and offline teaching experience, she has devoted to the research on student behavior and teaching effect. Her research interest includes the behaviors of employees in organization and how to enhance the core staff's degree of satisfaction.

PENG GUOQING received the master's degree from the Guilin University of Electronic Technology. Since 2018, he has been jointly trained as a Visiting Student in transportation engineering with Southeast University. His research interests include ITS, and data processing and modeling. He received First Class Scholarships at the University Level in 2017.