Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2024 SCPE. Volume 25, Issues 6, pp. 4950–4959, DOI 10.12694/scpe.v25i6.3312

APPLICATION OF FACIAL ANALYSIS BASED ON CONVOLUTIONAL NEURAL NETWORK AND ITERATIVE DECISION TREE FOR TEACHING EVALUATION IN SMART CLASSROOM

JIANG HUI*, FU WENTAO†, AND ZHANG JIAN‡

Abstract. The creation of intelligent classrooms has been expedited by the rapid advancement of the Internet and computer vision, and intelligent teaching has increased the interactivity and effectiveness of learning. Teachers' teaching and students' classroom learning state ultimately affect the teaching effect. Students' facial expressions during class can reflect emotional changes and the current learning state. The computer camera in the smart classroom collects students' face image data, uses texture-based information, edge-based information, geometric information, and global and local feature extraction to identify and analyze and process the students' facial expressions. Research has shown that the combination of expression recognition and an intelligent teaching classroom can accurately identify and analyze students' emotions and learning status, and can effectively evaluate the teaching effect of the intelligent classroom, which helps to improve teaching quality and learning efficiency. Therefore, applying facial expression recognition in the intelligent teaching classroom has far-reaching significance.

Key words: Machine vision, Smart Classroom, Effective Classroom, Face detection, Recognition of sb's expression

1. Introduction. The development of intelligent teaching and learning has accelerated as a result of the Internet and education coming together [1]. The goal of the smart classroom is to enhance conventional classroom teaching strategies through the use of cutting-edge technology and intelligent teaching tools. Through digital and networked means, it integrates rich teaching resources and realizes real-time interaction between teachers and students. Students can submit and answer questions online, and teachers can provide feedback and guidance immediately, enhancing the interactivity of learning and making teaching and learning more effective and individualized [2,4].

Positive learning emotions can have a significant impact on the brain's active thinking, enhancing learning ability, and human emotions are reacted to facial expressions. The change in students' emotions throughout the lesson can reflect the students' learning environment in the classroom. In the traditional method of instruction, teachers must personally observe students' facial expressions to determine how well they are understanding the lesson. But given that educators typically deal with sizable class sizes, it can be challenging for teachers to keep track of the majority of students' situations during the lesson [5][6].

And more recently, the use of computer face recognition technologies has been encouraged by advances in computer technology, image processing theory, and pattern recognition [7]. In a classroom with innovative teaching, the computer camera can track each student's facial expression in real-time and assess the students' learning status through recognition and analysis. This aids teachers in accurately and quickly understanding the students' learning situations, improving and optimizing their teaching strategies, and enhancing the quality of their instruction.

The combination of two algorithms exemplified by convolutional neural networks(CNN) and iterative decision trees applied to intelligent teaching and evaluation of face recognition has also been gradually increased to better enhance the recognition and evaluation accuracy and efficiency.

2. The relationship between human facial expressions and behavioral expressions and the methodological approach implemented by machine vision analysis. Psychologist Mehrabian states

*College of Foreign Languages, Northeast Forestry University, Harbin 150040, China
†College of Mechanical and Electrical Engineering, Northeast Forestry University, Harbin 150040, China (Corresponding author, micoc@foxmail.com)
‡College of Mechanical and Electrical Engineering, Northeast Forestry University, Harbin 150040, China

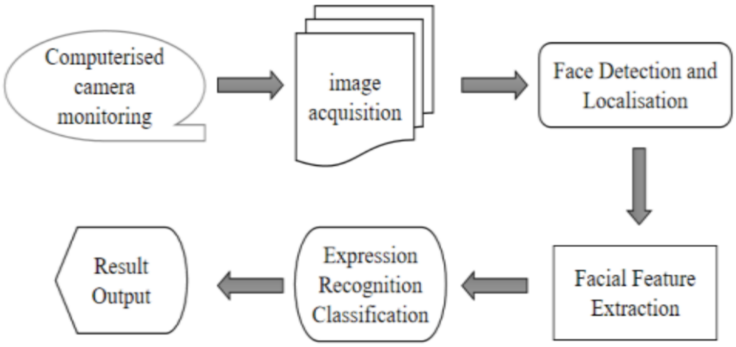

Fig. 2.1: Basic flow of computerized face expression recognition.

that emotional messages are expressed 7% in words + 38% in voice + 55% in facial expressions [8]. This demonstrates the significance of expression in evaluating human emotion. In order to ascertain a person's emotional state and expressive purpose, a system called computer face expression recognition examines and decodes that person's facial expression. It is based on pattern recognition and computer vision. By using a camera or other image acquisition device to capture a face image, this technology can then apply an algorithm to process and analyze the image, extract key features, and compare and match them with a predefined expression model to ultimately identify the category of an individual's expression, like as joyful, depressed, furious, scared, shocked, disgusted, and so forth [9]. Expression recognition is frequently employed in a variety of industries, including security surveillance, commercial promotion, psychological medical care, and fatigued driving. The field of education can also benefit from expression recognition. Similar to this, expression recognition can be used in the field of education to better understand students' psychological states and to assist teachers in capturing students' attention, understanding, and interest in knowledge and other information so that appropriate teaching control measures can be taken to enhance the quality of instruction.

Computerized face expression recognition usually consists of the following basic processes (shown in Figure 2.1):

(i) Data Acquisition: A camera or image capture device is used to acquire face image data. This data can be a live video stream or a still image.
(ii) Face Detection and Alignment: the acquired image is subjected to face detection, which identifies the presence of faces and calibrates their positions. At the same time, the detected faces are aligned so that the position and size of the face images are consistent with the model requirements.
(iii) Feature extraction: extract feature information from the aligned face image.
(iv) Expression classification: using the extracted feature information, it is fed into the pre-trained expression classification model, which uses comparison and matching to determine the category of facial expressions.
(v) Result output: determine the category of the facial expression based on the classification model's output, then show or output the conclusion.

It should be remembered that the procedure described above is only a basic framework and that the actual implementation may differ. To increase accuracy and resilience in real-world applications, additional procedures including data pretreatment, model training, and optimization could be needed.

3. Extraction and Numerical Algorithmic Approach for Machine Vision Analysis Targeting Human Facial Features. Data that can convey a learner's emotional traits are extracted via facial feature recognition. The four categories of facial feature extraction techniques are essentially comparable: (i) Feature extraction based on edge information, like directional gradient histograms, etc.; (ii) Feature extraction based on texture information, such local binary patterns, etc.; (iii) feature extraction based on geometrical information, such as local curvilinear waveform transforms; and (iv) Principal Component Analysis and feature point calibration are two methods for feature extraction that use both global and local information.

Local Binary Patterns (LBP) were originally proposed by Ojala et al [10] in 1994, can generate a binary code based on a comparison between a pixel and its surrounding pixels to represent the local texture elements of an image. LBP features have the advantages of being computationally simple and robust to light variations, and are widely used in practical applications. The traditional LBP algorithm encoding formula is:

$$S\left(g_p - g_m\right) = \left\{ \; 1, g_p \geq g_m 0, g_p < g_m \right.$$

$$LBP = \sum_{p=1}^{8} S\left(g_p - g_m\right) 2^p$$

where g_m stands for the center point's grey value and g_p stands for the surrounding eight pixel points' grey values.

For the weak intensity of facial micro-expression variations, background noise interference, and small feature differentiation, A micro-expression recognition network that combines parallel attention and LBP was proposed by Shuaichao Li et al. [11]. This network extracts RGB global features and LBP local texture features and then obtains more effective micro-expression features through the attention feature fusion module while simultaneously introducing a dense connectivity mechanism. A better LBP technique was put forth by Yu [12] that incorporates local dynamic thresholds, integrates equivalent and circular patterns, and reduces the size of the feature vectors while preserving the essential feature vectors needed for successful face recognition.

An image's local texture and edge information can be described using the feature description operator called Histogram of Oriented Gradients (HOG). The HOG feature extraction approach generates the gradient histogram by determining the gradient direction and intensity in various image regions. It primarily focuses on the distribution of gradients in a picture. HOG characteristics can assist in capturing crucial details, such as the edges and curves of the face, in the process of recognizing facial expressions. For the purpose of recognizing facial expressions, Ahmed et al. [13] suggested a brand-new local texture pattern called the Gradient Directional Pattern (GDP) and an efficient feature descriptor built utilizing GDP coding. The derived GDP features characterize the local picture primitives more steadily and maintain more information than the grey level-based techniques. Face recognition is accomplished by Xie et al. [14] using deep learning techniques, face feature extraction from HOG data, image segmentation, and convolutional neural network technology for training and coding output. The results of the experiments demonstrate that face recognition technology can accurately recognize faces with several gestures and meet application specifications.

A multi-scale transform technique for signal and image processing is the Local Curvelet Transform (LCT). To capture local features at each scale, the essential concept is to break down a signal or image into local curvilinear wave bases. Due to its superior performance in the presence of varying lighting conditions and face orientation changes, wavelet transform is widely employed in the field of face feature extraction [15]. The most important and relevant face features can be preserved while the unnecessary portions of the image information are efficiently minimized using wavelet transform. This makes the wavelet transform a trustworthy and efficient way for removing important information from face photos in challenging environmental settings for precise face verification and recognition. Ahmed et al [16] proposed a deep learning-based Gabor wavelet transform, which was used to extract features from symmetric face training data and, through testing, was found to be superior to other techniques.

In statistics and multivariate data analysis, the Principal Component Analysis (PCA) technique is frequently used for feature extraction and data degradation. PCA extracts features using global information-based approaches, focusing mainly on the direction of projection that explains the most variance in the data. Eigenfaces (Eigenfaces) are based on the idea of PCA by transforming image data to low dimensional feature space and then using these features to represent and compare face images. A face image can be represented as a vector X of size n*1 (n is the result of the image's width and height), and the number of training samples is N. At this point, using the overall scatter of the set of training samples in PCA, the maximum variance corresponds to the largest feature value, and the corresponding eigenvector defines a projection direction. By selecting principal components with large variances, we can gradually capture the main patterns and structures

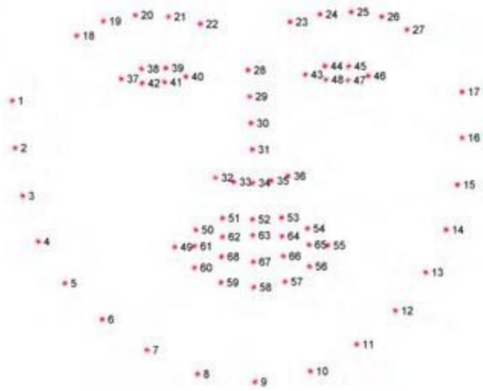

Fig. 3.1: Distribution of key points of DLIB face.

in the dataset. The cloth matrix is used as the generating matrix [17], namely:

$$C = E\left[XX^T\right] \approx \frac{1}{N}\sum_{\mathrm{k=1}}^{N} X_k X_k^T$$

Representation of N face vectors by an n *N matrix:

$$X = [X_1, X_2, X_3, \cdots, X_n]$$

Then C can be expressed as: $C \approx \frac{1}{N} X X_T$

However, PCA ignores the distribution of the data and bases its feature extraction procedure solely on the covariance matrix of the total data, disregarding the local correlations between samples. Therefore, when the data are unevenly distributed or when there is a local structure, PCA might not be able to capture crucial local information. To solve this problem, other algorithms are needed to consider local information for feature extraction. Meanwhile, to improve the accuracy and effectiveness of data representation, feature extraction methods need to be selected reasonably according to specific application scenarios and data distribution. The Active Appearance Model (AAM) model was utilized by Han et al. [18]; it labels the crucial spots in the training set to extract the average form through dimensionality reduction using Principal Component Analysis (PCA) and is used as a shape model. On this basis, it is further combined with the Constrained Local Model [19] (CLM) to achieve the extraction of multi-pose face features.

Feature extraction is based on local information, such as feature point calibration, which detects and locates key feature points in a face image, such as the position of eyes, eyebrows, mouth, etc., to achieve tasks such as face localization, pose estimation, and expression analysis. Jia et al [20] used the feature extractor provided by DLIB officials [21] to pre-train the model to obtain the key points of the face.DLIB is a C++ library containing machine learning algorithms and tools, which uses the face images that have been labeled with 68 key points as the training set to generate the model, and the acquired images are used to estimate the locations of the feature points using this model. The distribution of the 68 key points of the face in DLIB is shown in Fig. 3.1 shows.

4. Analysis and classification of correlations between human facial expression features and behavioral representations. Facial expression recognition refers to analyzing and recognizing the expressions shown by facial muscle movements through face images or video data to infer the emotional psychological state of human beings. In the classroom, where students' facial expressions can reveal a lot of information, Table 4.1 outlines some of the usual facial characteristics that are relevant to teaching and learning. The majority of the time, when students are ready to listen intently to what is being learned, they exhibit pleasing facial expressions and a forward-leaning posture that shows interest in what is being taught. When students are reluctant to

Table 4.1: Common expression features of the human face.

Expression	Brows	Eyes	Lip
pleasantly	slightly recurved	bright and lively	Slightly upturned to the sides
fury	curl up and lower the eyebrows	Eyes wide and possibly bulging	Lips tightly closed with one corner straightened or downwards
ordinarily	spontaneous outreach	spontaneous opening	No distinctive features
misgivings	pucker up	Upper eyelid lift	The corners of the mouth pull down
disdain	slightly elevated	Eyes cold, slight squinting	Tightly closed, angled slightly downwards, sometimes slightly puckered
resist	pucker up	Dodging, avoiding, or appearing indifferent	Become tightly closed, may bite their lips or pout

what they are learning, they may adopt resistant stances like slouching or spending a lot of time looking down in class. If students struggle to comprehend the course material, they may also display puzzled facial expressions like frowning. In addition to the above, there is another related concept that needs to be mentioned, namely microexpressions [22]. Microexpressions are more accurate at capturing people's genuine emotions and intentions than intentionally created facial expressions. Micro-expressions are a component of psychological stress micro-responses, which arise from human instinctual responses and are not under the conscious mind's control. As a result, they cannot be covered up or hidden because they are not under the conscious mind's control. The complexity and inherent specificity of microexpression analysis study make it difficult [23]. We can accurately understand students' emotional states and learning motivation by observing and categorizing their facial expressions during class. This allows us to inform teachers to intervene and modify teaching tactics as needed.

Facial expression classifiers include two kinds: (i) traditional machine learning algorithms, Support Vector Machine (SVM), Decision Tree Algorithm, Adaboost Algorithm, and K-Nearest Neighbor Algorithm (k-Nearest Neighbor) are the primary algorithms used for facial expression classifiers; (ii) Convolutional neural networks (CNN), recurrent neural networks (RNN), and long short-term memory neural networks (LSTN) are examples of deep learning models. Each of them has benefits and drawbacks, and in order to achieve faster and more accurate recognition, they must work in tandem.

Traditional machine learning algorithms have the advantage of better interpretability and applicability to small sample data in facial expression classifiers, but require expertise and experience in feature engineering and may be limited in their ability to process high-dimensional and complex data. SVM is a fundamental binary classification algorithm, and its objective is to determine the best hyperplane in the sample space to distinguish between various categories of samples, to maximize the interval between the two categories, i.e., to select the hyperplane division with the "maximum interval". To divide the feature space into two regions, one of which is assigned to one category and the other is assigned to another, the basic idea is to find the support vector (the closest sample point to the decision boundary) to separate the samples of different categories and maximize the distance from the support vector to the decision boundary, which can be regarded as a hyperplane. SVMs have the advantages of high classification accuracy, large data handling capacity, good performance on small sample datasets, and suitability for both linear and nonlinear classification tasks. SVMs, on the other hand, have a high computing requirement for large training sets, require empirical and experimental analysis to choose the right kernel function and parameters and perform badly when the data categories significantly overlap. Varma et al [24] used Support Vector Machines (SVMs) and Hidden Markov Models (HMMs) to classify a given face on two different datasets. Firstly, SVM was used to effectively differentiate different categories of face samples by finding the optimal hyperplane in the sample space and maximizing the spacing between the categories, while HMM modeled and classified the time series data of face samples. The system can accurately recognize the six primary emotions after integrating these two models, and the identification of face expressions is accomplished by mixing feature vectors. Sikkandar et al [25] presented the Improved Cat Swarm Optimisation (ICSO) method as a better alternative to the Applied Cat Swarm Optimisation (CSO) technique. The classification of facial expressions is performed using a Support Vector Machine (SVM) Neural Network (NN), and experimental findings indicate that ICSO performs more accurately and quickly than the current method.

An algorithm known as a Decision Tree (DT) is built on a tree structure and uses a sequence of judgment nodes and leaf nodes to generate predictions and choices. The main objective is to divide the dataset into smaller subsets recursively until all of the samples in the subset fit into one category or another. To ensure that the divided subset is as pure as possible-that is, the samples belonging to the same category are clustered as closely as possible-the division is made at each judgment node based on the value of a particular attribute. The decision tree technique has excellent interpretability, broad applicability, independence from data scalability, and the capacity to handle multi-output issues. However, it is prone to overfitting and has a high level of instability. To extract the facial features, Gupta et al. [26] proposed a feature-based method for 2D face images that uses Scale-Invariant Feature Transform (SIFT) and Speeded Up Robust Features (SURF) to extract the facial features. Following that, the expressions were categorized using the decision tree and random forest classification algorithms, with a maximum experimental recognition accuracy of 99.7%.

An integrated learning technique called Adaboost (AdaBoost, Adaptive Boosting) tries to strengthen a classifier by integrating many weak classifiers. Adaboost's main goal is to make data difficult to categorize by repeatedly training a set of simple classifiers and altering the weight of the samples based on the effectiveness of previous classification results such that difficult-to-categorize samples receive more attention. The Adaboost algorithm has the benefits of increasing classifier accuracy, avoiding the overfitting issue, being highly adaptable, and not relying on a particular classifier. However, it also has disadvantages, including being sensitive to outliers, taking longer to train, and having a tendency to misclassify a few categories of samples when there is an imbalance in the data. Hui et al [27] for the AdaBoost algorithm with the increase of learning difficulty leads to the classification efficiency of the classifier declines, stability deterioration, and other issues, combining the advantages of the two algorithms, based on the ant colony algorithm to optimize the parameters of the SVM, to improve the Adaboost_SVM cascade classification algorithm, the first to extract the haar-like rectangle features through the Adaboost classifier Firstly, haar-like rectangular features are extracted through Adaboost classifier, and then non-face regions are quickly excluded; To extract facial expression characteristics, the Gabor wavelet transform is utilized, and when paired with the Adaboost_SVM cascade classifier, the average rate of face expression identification is 94.2%, and the detection speed has been greatly improved. Lakshmi et al [28] proposed a classifier combining PCA and AdaBoost algorithm for facial expression recognition, which effectively reduces the feature redundancy of frontal Gabor features.

A popular supervised learning technique with applications in both classification and regression issues is the K-Nearest Neighbors algorithm (KNN). The fundamental concept is to categorize samples based on their proximity to one another. The benefits of KNN include its adaptability, simplicity, and lack of an explicit training step; however, it is computationally challenging, sensitive to outliers, and necessitates the choice of an acceptable K number. Zhang et al [29] proposed an expression recognition method based on Gaussian Marco Random Field (GMRF) with multiple chunking way feature combinations. The GMRF features of different chunking modalities are combined and classified with KNN. The JAFFE dataset is used for testing, and the findings demonstrate that the approach achieves an accuracy of 89.8% in recognizing facial emotions.

Choosing an algorithm that is appropriate for the task at hand and the data's properties is essential. Deep learning algorithms, in contrast, are typically better suited for processing large-scale and complicated data, but their interpretability is low. Convolutional neural network (CNN) is one of the best deep learning models for facial expression classification tasks because it has a strong ability to describe image features, can efficiently capture both local and global information of facial images, is translation invariant, reduces the risk of overfitting through parameter sharing and sparse connectivity, and exhibits a strong model generalization capability. Eventually tasks such as classification or regression are performed with a fully connected layer. In order to more accurately characterize students' facial emotions while listening in class, Zhou et al. [30] used CNN fused with Iterative Decision Tree (GBDT) to extract facial image attributes. The numerous training samples that were gathered were manually labeled in a supervised way and divided into attentive and inattentive samples based on their facial expressions. The fully-connected layer feature values are then entered into the GBDT after pre-training the CNN using the training samples. The samples are then classified using the singlelayer hidden layer MLP perceptron after the GBDT has been trained using the tree nodes as the feature values to be fused with the CNN features. In order to identify and classify the emotional expressions of physically disabled people (deaf and bedridden) as well as children with autism, Hassouneh et al. [31] developed a real-time emotion algorithm in conjunction with a CNN classifier based on facial labeling and electroencephalogram (EEG) signals. They were able to achieve the highest recognition rate of 99.81%.

A neural network model having the ability to process serial data and handle temporal correlation is called a recurrent neural network (RNN). By inserting recurrent connections and utilizing prior knowledge when processing each time step, it enables the network to simulate sequential data. Facial expressions are constantly changing during facial expression recognition, and RNN can capture the temporal information in facial expression sequences, process dynamic facial expression sequences effectively, and be able to capture long-term dependencies, increasing the accuracy and performance of facial expression classification. Kansizoglou et al [32] used an RNN architecture to accurately estimate a speaker's immediate and persistent emotional state during an interaction.

The Long Short Memory Neural Network (LSTM) is a variant of the Recurrent Neural Network (RNN) commonly used to process sequence data. Compared to traditional RNNs, LSTM introduces a gating mechanism that better captures and remembers key information in long sequences. The Memory Cell, which makes up the majority of the LSTM, is made up of three gates (Input Gate, Forget Gate, and Output Gate) and a cell state (Cell State). By using these gating mechanisms, LSTM may learn and regulate the information flow to better handle dependencies and time delays in lengthy sequences, avoiding the issue of gradient vanishing or explosion in conventional RNNs. Li et al [33] addressed the problem that most people's psychological state is in sub-health in modern society and designed a bi-directional LSTM network based on spatiotemporal attention to achieve micro-expression image recognition. A real-time micro-expression detection technique based on optical flow and LSTM was proposed by Ding et al [34]. This method extracts feature curves from the Facial Action Coding System (FACS) and utilizes the LSTM to feature curves to categorize them and identify whether or not micro-expressions are present.

5. Algorithm and implementation of classroom teaching effectiveness evaluation based on human facial expression. Evaluation of the teaching process and student learning outcomes is done in the classroom to determine how well instruction is being provided and how well students are learning. By analyzing student performance in terms of engagement, focus, and cooperative skills based on facial expression recognition, homework, and quiz assessment in smart teaching classrooms, and teacher self-assessment, classroom effectiveness may be thoroughly examined.

Han et al [18] combined traditional cognitive behaviors with students' head posture and facial expression behaviors to construct a holistic and systematic learning effect evaluation system, which evaluates the classroom teaching effect in terms of individual students and the classroom as a whole. Lastly, by comparing manual statistics with the system detection, the study confirmed the accuracy of the system in detecting the overall classroom attention, participation, difficulty, and active time. The accuracy rates of attention, participation, difficulty, and active time were 88%, 87%, 80%, and 85%, respectively, all higher than 80%, suggesting that the system can be used to teach in the classroom and can produce more accurate affective data.

Jia [20] et al. made a classroom activity analysis corresponding to the timeline by observing the changes in students' facial expressions and head posture. It mainly consists of two parts: (i) It is an analysis of the individual's activity based on the time axis. (ii) is a trend analysis of the overall change in activity. The study considers all expressions except the 'nature' state as active and engaged expression attitudes, so the overall activity index expression activity is given as follows:

$$exp_a ct = 1 - \frac{nature}{len}$$

where nature denotes the number of "usual" expressions in this frame, and len denotes the number of all expressions in this frame.

Steering activity is the left and right head bobbing of an individual. The formula for steering activity is shown below,

$$ora\_act = \frac{nun\_act}{num}$$

where num_act denotes the number of people who turned active in this frame and num denotes the total number of people in this frame.

Based on these two indicators, the change curve of the overall timeline activity of the classroom is plotted thus analyzing the quality of teaching and learning in the classroom.

By capturing and tracking image samples of students' facial expressions at various crucial moments, Tang et al.'s [35] trained network was able to determine each student's emotional state. Based on the distribution of a number of fundamental expressions in the PAD emotional state model, participation weights for various expressions are assigned. A comparison between the experimental results and the teacher's evaluation demonstrates the validity and efficacy of the approach, as well as its sensible and useful role in evaluating teaching.

Facial expression-based assessment of classroom effectiveness has many positive effects. First off, it gives teachers immediate feedback on their students' emotional states, which helps them better comprehend their engagement and emotional states. This gives teachers the ability to personalize education, make timely changes to teaching ideas and approaches, offer more help and explanations, and foster a happy learning environment. Second, judging students based on their facial expressions can improve their understanding of their learning circumstances and boost their enthusiasm and capacity for autonomous learning. Students can take the initiative to change their learning state and thus increase the effectiveness of their learning by taking note of their facial expressions to better understand their emotional state and involvement. To further develop teaching tactics and raise the level of instruction, teachers can use the evaluation method to examine the success of their instruction and discover which teaching materials or approaches are most effective with their students. In conclusion, using facial expressions to assess classroom performance gives teachers and students valuable feedback and direction, encourages individualized instruction and students' independent growth, and so significantly improves the efficiency and quality of classroom instruction.

6. Conclusions. The use of facial expression recognition in the smart classroom aids teachers in understanding students' learning responses and emotional states. At the same time, the smart teaching system can provide individualized learning materials and teaching strategies by students' emotional states and comprehension levels based on the analysis results of facial expression recognition, thereby improving the overall quality of instruction. However, there are still a lot of challenges and issues that require more thought:

(i) Privacy issues. Face expression recognition involves capturing and analyzing students' facial images, which may raise privacy issues. When applying face expression recognition technology, it is necessary to ensure the protection of students' personal information and to comply with relevant privacy regulations and policies.

(ii) Accuracy and judgment errors. Face expression recognition software could make errors in complex scenarios and is not completely accurate at capturing and interpreting students' emotional states. Because of this, attention must be taken to how the results are interpreted and applied in the application.

(iii) The balance of personalised learning. It is important to ensure that the implementation of personalised learning does not preclude traditional teaching methods, while balancing the need for personalised learning with the needs of the class as a whole.

(iv) Cultural variations. Cultural differences may cause expression recognition problems and cause human facial expressions to have different meanings in various circumstances. To prevent the issues of prejudice and misjudgment, cross-cultural applications must take into account and adapt to the cultural characteristics of the various student groups.

In addition to the above points, how two or more algorithms can be combined to further optimise the efficiency and accuracy of face recognition in the smart teaching classroom is also a key focus of future research.

With the continuous development of artificial intelligence technology, the prospect of face expression recognition in the smart teaching classroom is full of expectations. It will promote the education industry to achieve personalized teaching and improve students' learning effectiveness while providing teachers with more valuable data and feedback to help them better leverage the power of education.

REFERENCES

[1] Jabbar M A, Kantipudi M V V P, Peng S L(2022) Machine Learning Methods for Signal, Image and Speech Processing. CRC Press.
[2] Kaur A, Bhatia M, Stea G (2022) A survey of smart classroom literature. Education Sciences, 12(2): 86.
[3] Cebrián G, Palau R, Mogas J (2020) The smart classroom as a means to the development of ESD methodologies. Sustainability, 12(7): 3010.
[4] Zhou B (2016) Smart classroom and multimedia network teaching platform application in college physical education teaching. International Journal of Smart Home, 10(10): 145-156.
[5] Feng H H, Zou B. Exploration and reflection on the enhancement of students' motivation in traditional classroom. Journal of Higher Education, 2021, 7(25): 56-59.
[6] Zhang J X, Zhu L, Wu Z F (2018) Immediate evaluation method of classroom teaching based on information technology. Journal of Yangzhou University (Higher Education Research Edition), 22(02): 80-85.
[7] Fang Y N, Ge W L (2010) Intelligent video surveillance system development and application. Value Engineering, 29(17): 97-98.
[8] Mehrabian A (2008) Communication without words. Communication theory, 6: 193-200.
[9] Shi X G, Zhao X M, Zhang S Q (2014) New advances in face expression recognition research. Laboratory Research and Exploration, 33(10): 103-107.
[10] Ojala T, Pietikäinen M, Harwood D (1996) A comparative study of texture measures with classification based on featured distributions. Pattern recognition, 29(1): 51-59.
[11] Li S C, Li M Z, Sun J A (2023) A micro-expression recognition method fusing LBP and parallel attention mechanism [J/OL] Journal of Beijing University of Aeronautics and Astronautics, 1-13.
[12] Jiang Y J (2023) Research and application of improved LBP algorithm in face recognition. Journal of Hubei Normal University (Natural Science Edition), 43(02): 51-59.
[13] Ahmed F (2012) Gradient directional pattern: a robust feature descriptor for facial expression recognition. Electronics letters, 48(19): 1203-1204.
[14] Xie R Y, Hai B Z, Liu X (2023) Research on key technology of multi-gesture face recognition based on deep learning. Journal of Henan Institute of Technology, 31(01): 19-23.
[15] Cao C, Swash M R, Meng H (2020) Reliable holoscopic 3D face recognition//2020 7th International Conference on Signal Processing and Integrated Networks (SPIN). IEEE, 696-701.
[16] Ahmed S, Frikha M, Hussein T D H (2021) Optimum feature selection with particle swarm optimization to face recognition system using Gabor wavelet transform and deep learning. BioMed Research International, 2021: 1-13.
[17] Ding R, Su G D, Lin X G (2002) Comparison of feature face and elastic matching face recognition algorithms. Computer Engineering and Applications, 2002(07): 1-2+19.
[18] Han L, Li Y, Zhou Z J (2017) Analysis of teaching effect based on facial expression in classroom environment. Modern Distance Education Research, 2017(04): 97-103+112.
[19] Zhai Q R (2012) Research on face alignment methods for different poses. MsA Dissertation, Dalian Maritime University, China.
[20] Jia L Y, Zhang C H, Zhao X Y (2019) Classroom student state analysis based on artificial intelligence video processing. Modern Education Technology, 29(12): 82-88.
[21] King D E (2009) Dlib-ml: A machine learning toolkit. Journal of Machine Learning Research, 2009,(10):1755-1758.
[22] Xie H X, Lo L, Shuai H H (2022) An overview of facial micro-expression analysis: Data, methodology and challenge. IEEE Transactions on Affective Computing.
[23] Ben X, Ren Y, Zhang J (2021) Video-based facial micro-expression analysis: A survey of datasets, features and algorithms. IEEE transactions on pattern analysis and machine intelligence, 44(9): 5826-5846.
[24] Varma S, Shinde M, Chavan S S (2020) Analysis of PCA and LDA features for facial expression recognition using SVM and HMM classifiers//Techno-Societal 2018: Proceedings of the 2nd International Conference on Advanced Technologies for Societal Applications-Volume 1. Springer International Publishing, 2020: 109-119.
[25] Sikkandar H, Thiyagarajan R (2021) Deep learning based facial expression recognition using improved Cat Swarm Optimization. Journal of Ambient Intelligence and Humanized Computing, 12: 3037-3053.
[26] Gupta S, Thakur K, Kumar M (2021) 2D-human face recognition using SIFT and SURF descriptors of face's feature regions. The Visual Computer, 37: 447-456.
[27] Hui X W, Zhou J B. (2014) Face expression recognition based on improved Adaboost_SVM. Laser Journal, 35(09): 54-57.
[28] Lakshmi P V S, Akkineni H, Hanika A (2021) Adaptive convolution neural networks for facial emotion recognition//International Conference on Artificial Intelligence and Data Science. Cham: Springer Nature Switzerland, 2021: 135-143.
[29] Zhang L Z, Wang D X, Chen Y C (2020) Face expression recognition based on GMRF and KNN algorithm. Computer Application and Software, 37(10): 214-219.
[30] Zhou J G, Tang D M, Peng Z (2017) Research and implementation of classroom expression analysis software based on convolutional neural network. Journal of Chengdu Information Engineering University, 32(5): 508-512.
[31] Hassouneh A, Mutawa A M, Murugappan M (2020) Development of a real-time emotion recognition system using facial expressions and EEG based on machine learning and deep neural network methods. Informatics in Medicine Unlocked, 20: 100372.
[32] Kansizoglou I, Misirlis E, Tsintotas K (2022) Continuous emotion recognition for long-term behavior modeling through recurrent neural networks. Technologies, 10(3): 59.

[33] Li D X, Chen M S, Liu Y (2022) Spontaneous micro-expression recognition algorithm based on STA-LSTM. Journal of Jilin University (Engineering Edition), 52(04): 897-909.
[34] Ding J, Tian Z, Lyu X (2019) Real-time micro-expression detection in unlabeled long videos using optical flow and lstm neural network//Computer Analysis of Images and Patterns: 18th International Conference, CAIP 2019, Salerno, Italy, September 3-5, 2019, Proceedings, Part I 18. Springer International Publishing, 2019: 622-634.
[35] Tang X Y, Peng W Y, Liu S R (2020) Classroom teaching evaluation based on facial expression recognition//Proceedings of the 2020 9th International Conference on Educational and Information Technology. 2020: 62-67.

Edited by: Jingsha He
Special issue on: Efficient Scalable Computing based on IoT and Cloud Computing
Received: Feb 2, 2024
Accepted: Apr 2, 2024