Summary of Research on Facial Expression Recognition

Qizheng Chen*

Department of Computing Science, University of Connecticut Storrs, Connecticut, 06269, United States

*Corresponding Author Email: Qizheng.Chen@uconn.edu

Abstract. Facial expression recognition allows computers to understand expressions according to human thinking. It is an important component of human-computer interaction. However, with the rapid development of deep learning, the research of deep learning technology in the field of facial expressions has also become a research hotspot, so is deep learning. It is necessary to analyze the application of technology in facial expression recognition and the results obtained. Firstly, several commonly used expression data sets are summarized; then the expression recognition methods based on deep learning are classified from the aspects of feature extraction and feature classification, and several network improvements in expression recognition based on deep learning are analyzed from the aspect of network improvement. Method; Finally, the challenges and future development in the field of facial expression recognition are described.

Keywords: Facial expression, human thinking, human-computer interaction, feature extraction

1. Introduction

Facial expression implies the people's emotions beyond people's words in interpersonal communication. Human emotional response can be roughly expressed through expressions, sounds, texts, gestures, and physiological signals [1]. Selecting appropriate emotional features is an important basis for emotional recognition research. Psychologist Mehrabian [2] pointed out that 55% of information in human communication is conveyed through facial expressions. Facial expression is an important part of judging a person's emotional state, intention, and personality. It is one of the most natural, common and intuitive signals for humans to convey the emotional state and intention.

Facial expression recognition is a hotspot of interdisciplinary research, mainly including psychology, biology, and computer science. As early as the 1970s, Ekman and Friesen defined six basic emotions based on the conclusions of cross-cultural research: anger, disgust, fear, happiness, sadness, and surprise. The way of perception is the same [3]. There are other emotion description models such as the Facial Action Coding System (FACS), which can describe and classify a wider range of facial expressions. The early exploration of face recognition technology began with SUWAM, etc., by marking the feature points on the continuous face image sequence, and then tracking and identifying the changes of these feature points and comparing them with the original model to obtain expression information [4]. Facial expression recognition technology based on traditional machine learning relies on artificial features. Local Binary Algorithm (LBP), Gabor feature extraction, Active Shape Model (ASM), etc. are commonly used methods for extracting features in machine learning algorithms, and these methods are often used for small data sets. With the improvement of computer hardware and the emergence of high-performance GPUs, deep learning networks have become a hotspot again. In 2012, the appearance of the AlexNet model based on CNN has greatly improved the accuracy of expression recognition, far exceeding the accuracy of traditional machine learning. [5]. This article reviews the research progress of facial expression recognition in machine learning and deep learning at home and abroad. First, it introduces the commonly used facial expression data sets; then, the main algorithms of machine learning and deep learning in facial expression recognition are respectively carried out. Summary; Finally, looking forward to the challenges and opportunities of facial expression recognition technology in the future, the future development trend of facial expression recognition technology is analyzed. With the advent of the data age, resource sharing has become more convenient, and the establishment of databases has gradually improved and diversified. The blowout of advanced algorithms inspires researchers to

explore in depth, and the field of facial expression recognition is changing with each passing day. The needs of the market and the scientific research community indicate that facial expression research will become an indispensable chapter in the research process of human-computer interaction.

In the foreseeable and intelligent future, facial expression recognition will inevitably penetrate into many life scenes, such as:

(1) Smart medical care. During the operation, the patient's micro-expression changes are monitored in real time, combined with parameters such as heart rate and blood oxygen monitored by medical devices to assist doctors in monitoring, and combined with big data medical case analysis to give diagnosis and treatment opinions. During the stay in the ward, the patient's expression can be monitored through electronic cameras to prevent the failure to detect changes in the condition at night or due to the negligence of family members and nurses.

(2) Service robots. In service industries, such as banking, restaurant robots, etc., robots are given a "seeing" function, combined with information such as the voice and gestures of the demanders, to provide higher-quality problem processing capabilities, and reduce the pressure of manual labor during busy business hours. At the same time, with the aging phenomenon on the stage of history, when the need for companionship cannot be fully satisfied by young people, the service robot will show warmth and charm, which can not only calm the mood of the elderly, but also transmit information to children and grandchildren in real time with the monitoring function. Its peace of mind.

(3) Intelligent driving. With the development of intelligent driving technology in recent years, human beings are expected to free their hands, but human guidance needs will never be outdated. In the process of road safety driving, intelligent facial expression monitoring will promptly remind the driver. At the same time, it can also be developed to adjust the speed of the car in conjunction with the automatic driving system, effectively preventing car accidents caused by fatigue driving, and the combination of man and machine to construct a safer driving system.



Figure 1. Expression recognition in a real driving environment. To construct the dataset, a baseline dataset sequence was captured in a real vehicle driving environment using a near-infrared camera.

2. Introduction to facial expression recognition

Humans are often accompanied by various emotional changes in the process of communication, and emotional changes will inevitably lead to certain physical changes, such as changes in facial expressions, body posture, and physiological signals such as pulse and heartbeat. Emotion recognition is to speculate and judge a person's basic emotions based on the changes in these different patterns of people in different emotional states. At present, common emotion recognition is based on facial expressions, voice [6-7], physiological signals [8-10], text [11-12], posture [13], etc., and facial expressions are the most intuitive and easy to obtain. This paper mainly focuses on the research methods of facial expression recognition. Emotion recognition based on facial expressions is to allow computers to understand expressions according to human thinking. When a person has a certain basic emotion, the face will show different expression patterns, which have different characteristics and

distribution scales. Facial expression recognition is to detect the face part from the static image or video, then extract the facial expression features from the face part, and finally classify the facial expressions into happy, sad, angry, surprised, fearful, disgusted, etc. Basic expressions.

3. Face expression data set

Datasets play a very important role in model training and selecting a dataset that provides accurate and sufficient information is one of the important conditions for training a sufficiently satisfactory quality model. In this section, some public datasets on facial expression detection are presented.

3.1. CK+

This is a facial expression data set released by a team from the University of Pittsburgh. It always seems as the initial dataset in facial expression researches. The CK+ contains 123 objects and 593 image sequences. But there are only 327 images have labels which they set as anger, neutral, contempt, disgust, fear, happy, sadness and surprise.

3.2. FER2013

FER2013 is a Kaggle official dataset. This dataset consists of 48*48pixel grayscale images. The faces have been automatically registered so that the face is more or less centered and occupies about the same amount of space in each image. There are 7 types of emotions in total, and 35887 images are available. Facial expression images in the database have a large amount of complete expression data, including real facial images.

3.3. AffectNet

AffectNet includes more than 1,000,000 images about facial expressions, it use six different ways and 1,250 key words to collect data on three net engine. The data set provides two emotion models: classification and dimension, of which 450000 are manually labeled with eight basic expression labels.

3.4. JAFFE

Jaffe database was released in 1998. The database was released early and the data samples were small. This database includes all kinds of expressions made by 10 Japanese women under the experimental environment according to the instructions. Each person is a group. Each group includes 7 kinds of expressions. Each expression has about 3 or 4 sample images, a total of 213 images. These images include different emotions with different intensities, have a high recognition rate, and the database is completely open, so they have a high utilization rate.

3.5. WebFace260M

Webface260m, composed of 4 million identities and 260 million faces images, provides a good resource for millions of deep face cleaning and recognition. It was published by Tsinghua University and it plays very well in some popular models.

4. Facial expression recognition based on machine learning

Facial expression uses machine learning algorithms as a common algorithm for image recognition tasks, with high recognition accuracy. The main processing steps of the facial expression recognition algorithm based on machine learning, include image preprocessing, feature extraction and expression classification.

4.1. Image preprocessing

Image processing is divided into two methods, analog image processing and digital image processing, and this paper focuses on digital image processing and its computer algorithms. Digital

image processing has two aims and purposes. The first method is to improve image data features by removing unnecessary noise that may weaken the intensity of image features and distortion in the data image, and the other method is to enhance important features, thus helping the model to process exposure and enhancement data more effectively later. In the example, for the facial expression task under image processing, components such as backgrounds that are not relevant to the face can be weakened or even completely removed in the pre-processing stage to highlight the face part of the data image.

Besides, other detailed handlings like light normalization also helps improve quality of the extracted facial image features. Since the fluctuations of illumination affect the accuracy of image feature extractions, the texture features extracted from facial expression features are more susceptible to the impact of uneven light intensity. Light normalization can effectively remove the influence of light [14]. Moreover, there is processing related to image size that needs to be considered when key features are not lost and then the image size is appropriately reduced.

4.2. Image feature extraction

Feature extraction is the core problem in classification tasks which directly affects the accuracy results.

Here is a list of several well-known and commonly used feature extraction algorithms used in face extraction:

Local Binary Algorithm (LBP) uses binary codes to describe the neighborhood of image elements. LBP is commonly used to study characteristics (low calculation cost and resistant to fluctuations in image gray scale values) of the local and individual parts within an image. It calculates the relative values for comparing each pixel with its neighboring pixels. Because LBP has good gray-scale invariance and rotation invariance, according to this feature, LBP is often used in texture classification, face recognition, and Gesture recognition and other fields [15]. Based on the lack of completeness of the features extracted by the basic LBP operator, Wang Xian et al. proposed a block-based complete local binary model (CLBP). The segmented CLBP histogram sequence is connected to obtain the CLBP feature of the face [16]; Jabid proposed to enhance the robustness of the algorithm by calculating the edge response based on the local direction mode (LDP) [17].

The response of Gabor wavelet is very similar to the visual stimulus response of simple cells in the human visual system. Gabor wavelet has good sensitivity to image edges and robustness to illumination. Gabor wavelet is widely used in the field of visual processing [18].

Optical flow method is a moving target detection algorithm based on optical flow, which is applied to pedestrian recognition and target detection. Optical flow method is applied to sequence-based expression recognition in facial expression recognition. Wu Xingen et al. proposed an improved instantaneous position velocity estimation method, which overcomes the drift problem of the optical flow algorithm and speeds up the convergence speed [19].

Facial expression feature extraction is a key issue for facial expression recognition. The features that meet the task scene are conducive to the completion of facial expression recognition. The above-mentioned feature extraction algorithm is a commonly used algorithm for facial expression recognition. In addition, there are color features, shape features, HOG features, etc. Used to portray facial expressions.

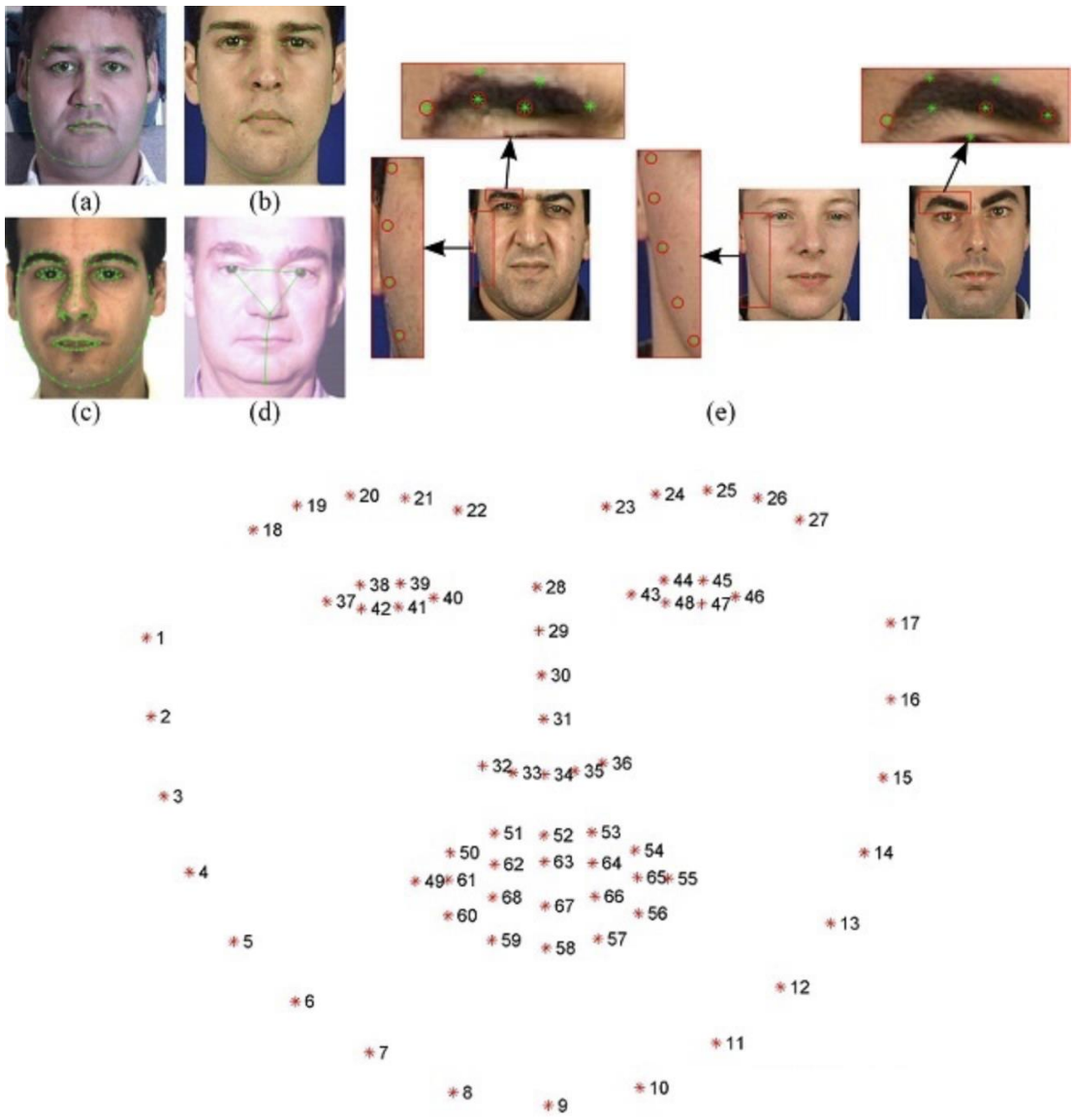


Figure 2. The 300-W dataset, for face key point localization, is mainly collected and labeled with 68 points for faces by color features and shape features of each pixel.

4.3. Expression classification algorithm

After feature extraction, the classifier is used to determine the category of the expression. Appropriate classifiers play an important role in facial expression classification. How to design high-performance classifiers is also an important research direction for facial expression recognition tasks. Introduction of several classifiers for facial expression recognition: Support Vector Machine (SVM) is based on statistics The machine learning method based on the VC theory and the principle of structural risk minimization has a good ability to solve the classification problems of small sample, nonlinear and high-level pattern recognition problems. The essence of the standard SVM problem is to solve a constrained quadratic programming. Through feature fusion, and then using SVM to recognize facial expressions, 95% accuracy is achieved on the CK data set [20]; Combining Gabor wavelet and SVM, reducing the amount of data processing, and achieving good results [21].

Bayesian classifier refers to the general name of the algorithm with Bayes theorem as the core. Bayesian decision-making uses the loss of misjudgment to select the best category classification through the situation where the relevant probabilities are known. Improved Bayesian classifier and combined with traditional facial features improve the recognition rate and have better robustness to

illumination and expression changes [22]. Traditional machine learning classification methods often face various problems for large-scale data processing. Most traditional machine learning algorithms only learn a classification boundary and do not have the characteristics of multi-dimensional classification.

Facial expression recognition has gradually developed into the field of deep learning, but the application of traditional machine learning in expression recognition is still worth exploring.

5. Facial expression recognition based on deep learning

Research on deep learning has made great progress in the field of image recognition, and the recognition rate in the ImageNet dataset is close to the human level. In the field of facial expression recognition, deep learning-related algorithms have taken the lead and are widely used in facial expression recognition based on static images and dynamic sequences. Researchers improve the accuracy of facial expression recognition from all angles, from feature optimization to data enhancement, from network structure changes to the proposal of new classifiers.

5.1. Data preprocessing

Deep learning also needs to pre-process the data to remove the differences in illumination and head posture in the background and image that have nothing to do with facial expressions. These factors will affect the feature learning of the deep network, and it is necessary to preprocess the data before training.

5.2. Deep network structure

The deep learning network uses the combination of low-level features to form more abstract high-dimensional representation features, and uses these high-dimensional features to discover the distribution of data, thereby achieving data classification. The deep learning network has a rich structure to achieve end-to-end classification tasks. Common network structures for facial expression recognition:

5.2.1 Convolutional Neural Network (CNN)

Because CNN has strong robustness to face position changes and scale changes, CNN networks are widely used. Using CNN to solve the problems of subject independence, translation, rotation, and scale invariance in facial expression recognition [23]. CNN is composed of convolutional layer, pooling layer, and fully connected layer. The convolution layer mainly uses the convolution kernel to calculate. The convolution kernel refers to the weighted average processing for a part of the input image. The weight of this process is defined by a function, and this function is the convolution kernel. The input image is convolved through the kernel function, and the feature map is obtained after the activation function is linearized; the pooling layer is connected to the convolution layer, and its function is equivalent to feature extraction of the feature map, reducing the size and size of the input feature map. Data volume, while acquiring more abstract features, preventing over-fitting and improving generalization; The fully connected layer converts high-dimensional feature maps into one-dimensional feature maps to achieve the task of classification. Common CNN network architectures include AlexNet, VGGNet, GoogleNet, Resnet, etc.

5.2.2 Recurrent Neural Network (RNN)

Recurrent Neural Networks (RNN) are often used to mine time series information in data and the deep expression of semantic information, and have achieved breakthroughs in speech recognition, language models, machine translation, and time series analysis. In the network architecture, the cyclic neural network will memorize the previous information and affect the subsequent information. Long-short-term memory network (LSTM) solves the problem that RNN loses learning ability when the information interval is relatively large, or the performance of RNN is reduced when the information interval is different and the length is different [24]. LSTM uses some "gate" structures to allow

information to selectively affect the state of the recurrent neural network at each moment. The gates of LSTM generally include forgetting gates, input gates and output gates. The forget gate in LSTM controls whether to forget the hidden cell state of the previous layer with a certain probability. A kind of nested LSTM, internal time dynamic modeling, external output, integrated acquisition of multi-level features [25].

5.3. Expression recognition based on deep network

With the development of deep learning in the field of facial expression recognition, a large number of expression recognition algorithms based on deep networks have emerged. According to the type of data processed, they can be divided into expression recognition algorithms for static images and expression recognition algorithms for dynamic sequences.

5.3.1 Static image

The static image expression recognition algorithm improves the accuracy of expression recognition through strategies such as pre-training and fine-tuning, adding functional modules, and building a cascade network. Yao et al. proposed a novel CNN architecture HoloNet, which combines the CReLU activation function with the residual network to increase the depth of the network without reducing efficiency, and designed a residual block to make the facial expression recognition system learn Multi-scale features are used to capture changes in facial expressions and achieve good results in facial expression recognition [26]; good results can be achieved based on a single deep network, and researchers have turned to the direction of multi-network integration. The integrated network should consider the integration effect of each network and the complementarity of the network [27].

5.3.2 Dynamic sequence

Expression recognition based on dynamic sequences obtains information from the temporal correlation of consecutive frames. Common strategies include frame aggregation, expression intensity network, and deep spatiotemporal network. The frame aggregation strategy uses the characteristics of the different expression intensity of the frames in the video clips to extract the features, and aggregates the output of the frames through various algorithms [28]. Frame aggregation is achieved by concatenating the mean, variance, minimum and maximum values of the features of all frames [29]; for most sequence-based expression recognition algorithms, subtle expressions are often ignored. The expression intensity takes training samples of different intensities as input. The inherent correlation between expressions in different sequences of expression intensity is used to optimize feature extraction [29]; although frame aggregation technology can integrate frames in video sequences, time dependence is not utilized. Using the deep spatiotemporal network under the premise of unknown expression intensity, a series of frames in the time window are used as a single input, and the texture and time information are used to encode more subtle expressions [30]. In real scenes, human facial expressions are a dynamic process, from subtle to obvious. The analysis of facial expressions through sequences is an important direction for future research.

A well-built deep expression recognition system mainly has the following problems: Lack of rich and diverse training data and expression-independent change factors, such as lighting, head posture, and ethnicity. In order to solve the problem of insufficient training data and over-fitting, pre-training and fine-tuning, network cascade, network collection, etc. have become the mainstream of facial expression recognition. The light normalization, posture normalization and image cropping operations in the image preprocessing stage can well remove the influence of irrelevant factors. Faced with changes in head posture, most deep expression recognition networks still fail to handle this problem well.

6. Conclusion

Facial expressions, as the main way to express emotional information during communication, have a wide range of applications in education, companion robotics, reconnaissance, medical care and

other fields that apply to human-computer interaction. Therefore, facial expression recognition has far-reaching research significance. Although deep learning methods have achieved certain results in facial expression recognition, in order to enable computers to better understand human emotions, the challenges and future developments in facial expression recognition research are as follows.

(1) The expression recognition database is often obtained in a laboratory environment, and the data set is relatively single. And some dirty data can occur;

(2) The accuracy of facial expression recognition on the SFEW dataset is relatively low. The reason is that the data in the field static facial dataset comes from the natural environment and contains many influencing factors such as illumination and posture, highlighting the lack of robustness of the facial expression recognition algorithm;

(3) The lack of data sets is also a key issue for face recognition. Faced with the differences in facial expressions, skin colors, and appearances, data plays an important role in deep learning methods. Using a richer data set can train a facial expression recognition network with better generalization;

(4) There are relatively few traditional facial expression classifications. Human emotions are more complicated, and the traditional expressions cannot adapt to the complexity of human-computer interaction, and their practicability may be greatly compromised.

Facial expression recognition based on the combination of images and other media has also made great progress. The facial expression recognition algorithm of voice and image, and the facial recognition of 3D expression have made great breakthroughs in solving the occlusion problem; the thermal data of the face is obtained through infrared rays, and the thermal characteristics of the learning are fine-tuned for facial expression recognition, which effectively avoids the effect of light. Facial expression recognition is also a popular research direction in face occlusion. At present, occlusion and non-frontal recognition are two major problems in face recognition and expression recognition.

References

[1] Shen Lidi. Research on facial expression recognition method combining deep learning and sparse representation [J]. Techniques of Automation and Applications, 2019, 38 (10): 93-97.

[2] Mehrabian A. Communication without words [J]. Psychology Today, 1968, 2(4): 53-56.

[3] EKMAN P, FRIESENWV. Constants across cultures in the face and emotion [J]. Journal of personality and social psychology, 1971, 17(2): 124-129.

[4] SUW A M, SUGIE N, FUJIMORA K. A preliminary note on pattern recognition of human emotion expression [C] //Proceeding of the 4th International Joint Conference on Pattern Recognition, 1978: 408-410

[5] LIU Dong, LI Su, CAO Zhidong. Overview of deep learning and its application in image object classification and detection [J]. Computer science, 2016, 43(12): 13-23.

[6] Wang Zhongmin, Liu Ge, Song Hui. Speech emotion recognition based on multi-core learning feature fusion [J]. Computer Engineering, 2019, 45(8): 248-254.

[7] Cao N T, Ton-That A H. Speech emotion recognition using a fuzzy approach [J]. Journal of Intelligent and Fuzzy Systems: Applications in Engineering and Technology, 2019, 36(2): 1587-1597

[8] Chen Jingxia, Wang Liyan, Jia Xiaoyun, et al. Emotion recognition of EEG signals based on deep convolution neural network [J]. Computer Engineering and Application, 2019, 55 (18): 103-110.

[9] Du Zhaohui, Si Yujuan. Pulse signal emotion recognition based on hierarchical support vector machine[J]. Journal of Jilin University (Information Science Edition), 2017, 35 (1): 37-42.

[10] Momennezhad A. EEG-based emotion recognition utilizing wavelet coefficients [J]. Multimedia Tools and Applications, 2018, 77 (20): 27089-27106.

[11] Zhang Yang, Liu Xiaoxia, Sun Kailong, et al. Research on text tendency recognition based on emotional description [J]. Computer Engineering and Application, 2015, 51(4): 158-161, 195.

[12] Li Yang, Pan Quan, Yang Tao. Sensitive information recognition based on short text sentiment analysis[J]. Journal of Xi'an Jiaotong University, 2016, 50(9): 80-84.

[13] Yan Jingjie, Zheng Wenming, Xin Minghai, et al. Bimodal emotion recognition of expression and gesture [J]. Journal of Image and Graphics, 2013, 18 (9): 1101-1106

[14] ZHOU Wei, CHENG Yong, CAO Xuehong. Face recognition based on illumination normalization and segment-complete LBP feature [J]. Computer engineering and application, 2015, (11): 145-149.

[15] OJALAT, PIETIKINEN M, HARWOODD. A Comparative Study of Texture Measures with Classification Based on Feature Distributions [J]. Pattern Recognition, 1996, 29: 51-59.

[16] WANG Xian, ZHANG Yan, MU Xin, et al. Improved LBP face recognition algorithm [J]. Optoelectronic engineering, 2012, 39 (7): 12-13.

[17] JABIDT, KABIRMH, CHAEO. Robust facial expression recognition based on local directional pattern [J]. ETRI journal, 2010, 32(5): 784-794.

[18] LEETS. Image representation using 2D Gabor wavelets [J]. IEEE transactions on pattern analysis and machine intelligence, 1996, 18(10): 959-971.

[19] WU X G, LUO L M. An improved optical flow field calculation method [J]. Acta electronica sinica, 2000(1): 130-131.

[20] LIEW C F, YAIRI T. Facial expression recognition and analysis: a comparison study of feature descriptors [J]. IPSJ transactions on computer vision and applications, 2015, 7: 104-120.

[21] WANG Xiaohua, XIA Chen, Hu Min, et al. Video sequence facial expression recognition with spatio-temporal features [J]. Journal of electronics and information, 2018, 40(3): 626-632.

[22] FENG Xiaorong, HUI Kang-hua, LIU Zhendong. Face recognition based on convolution feature and Bayesian classifier [J]. Journal of intelligent systems, 2018, 13(5): 769-775.

[23] XIA Tian, ZHANG Yifeng, LIU Yuan. Expression recognition based on feature points and multi-network training [J]. Journal of computer aided design and graphics, 2019, 31(4): 552-559.

[24] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735-1780.

[25] YU Z, LIU G, LIU Q, et al. Deng, Spatio - temporal convolutional features with nested lstm for facial expression recognition [J]. Neuro computing, 2018, 317: 50-57.

[26] YAO A, CAI D, HU P, et al. HoloNet: towards robust emotion recognition in the wild [C]//The 18th ACM International Conference. ACM, 2016: 472-478.

[27] ZHANG Z, LUO P, LOY C C, et al. From Facial Expression Recognition to Interpersonal Relation Prediction [J]. International Journal of Computer Vision, 2018, 126(5): 1-20.

[28] KAHOU S E, BOUTHILLIER X, LAMBLIN P, et al. Emonets: Multimodal deep learning approaches for emotion recognition in video [J]. Journal on Multimodal User Interfaces, 2016, 10(2): 99-111.

[29] YU Z, LIU Q, LIU G. Deeper cascaded peak - piloted network forweak expression recognition [J]. The Visual Computer, 2017 (6-8): 1-9.

[30] SUN Ning, et al. Deep spatial-temporal feature fusion for facial expression recognition in static images [J]. Pattern Recognition Letters, 2017, (119): 49-61.