Proceedings of the 3rd International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/6/20230902

Research of the methods on facial expression recognition

Xuanyi Chen[1, †], Yuyao Ding[2, †], Zhaoheng Li[3, 4, †]

[1]School of Artificial Intelligence, Beijing University of Posts and Telecommunications, No. 10, Xitucheng Road, Beijing, China
[2]Department of Telecommunications, Xi'an Jiaotong University, No. 28, West Xianning Road, Xi'an, China
[3]College of Physical Science and Technology, Hebei University, No. 180, Wusi East Road, Lianchi District, Baoding, China

[4]20201304003@stumail.hbu.edu.cn
[†]All authors contributed equally.

Abstract. As traditional machine learning and deep learning have developed recently, the recognition of facial expressions has been paid more attention by domestic and foreign scholars. First of all, this paper introduced the common data sets of single mode, traditional feature extraction methods and more common expression classification methods in detail, pointing out that the present single-mode expression recognition does not perform well in practical application scenarios, and cannot obtain good recognition results in complex environments. At the same time, the data sets are relatively simple and the number is small. Then, three recognition methods based on multimodality are introduced: Fusion at the feature, decision, and hybrid levels. The advantages and disadvantages of the three measures are minute described respectively. Finally, the thesis is summarized. In addition, the future development of more general and richer high-quality expression datasets is prospected and the improvement of current multimodal fusion technology are prospected.

Keywords: facial expression recognition, single mode, multi-mode, modal fusion technology

1. Introduction

As one of the ways to express people's psychological activities and emotional changes, facial expression can often express psychological activities more clearly than language and body movements, and can well convey important emotional information in face-to-face communication, while not affected by cultural background, congenital blindness and other factors [1]. It can also provide good and accurate information feedback under the condition of psychological evaluation by doctors and supervision of students' learning status by teachers. In recent years, due to the rapid growth of deep learning, identification of facial looking has begun to shine in the scene where people need to interact with machines [2], and has been applied in such fields as tired driving, intelligent security and health management [3]. However, when you leave the laboratory and come to the natural harsh environment, such as occlusion, light change and low-resolution acquisition environment [4], There are many problems, such as low recognition accuracy and large dependence on hardware requirements. By analysing the mode of facial expression and considering the linkage relationship between various key

DOI: 10.54254/2755-2721/6/20230902

local areas of the face when different emotional expressions are generated, researchers can effectively offset the negative impact of non-contributing facial areas on feature quality and help improve the accuracy of recognition in complex environments. Some scholars try to analyze facial expression in combination with modal information such as voice text and posture, and in order to still perform high-precision dynamic and static facial expression recognition in complex environments, different feature extraction models are often used to eliminate the interference of different samples. In the actual application scenario, the recognition process completed by only using a single mode: voice, posture, etc. is usually low in accuracy. To solve this problem, multimodal expression recognition gradually appears, which helps reduce the impact of external environment on recognition results. In this paper, the research status of the single mode methods is analysed and the three feature fusion methods in the multi-mode method are discussed

2. Single mode facial expression recognition

There are many similarities between facial expression recognition and face recognition. The core steps are facial image pre-processing, image recognition, judgment and classification. The specific recognition framework is as figure 1.

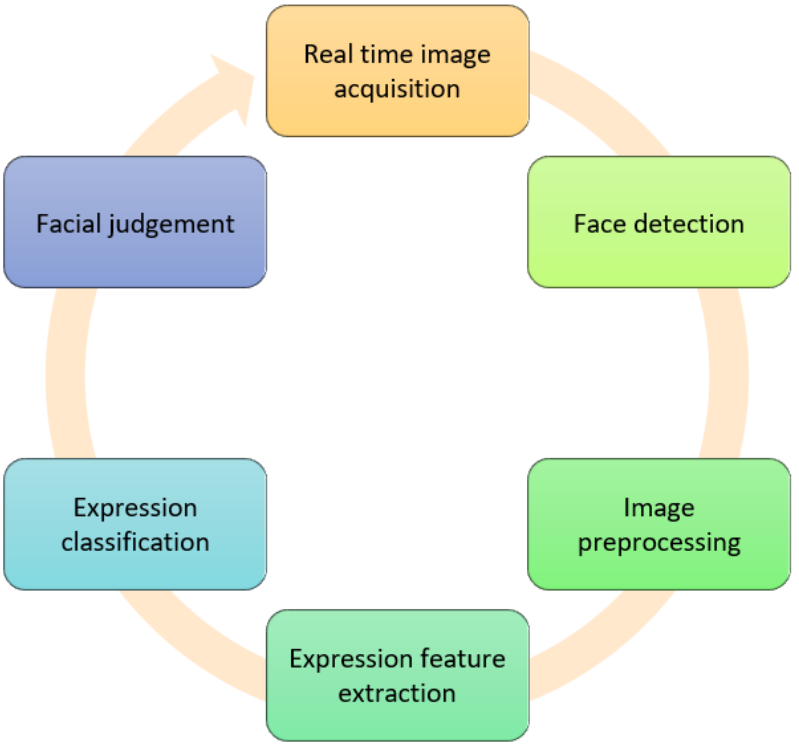

Figure 1. The facial expression recognition framework.

2.1. The data set of facial expression recognition

As early as 1971, psychologists Ekman and Friesen studied and put forward the concept of six basic human emotions, namely, wrath, pleasure, heartbreak, astonishment, disgust and fear. These basic emotions effectively summarize the types of facial expressions. The future facial expression data sets basically use these six emotions as labels for classification, which is conducive to the determination of general expression categories.

From the datasets listed in Table 1, we can find that most datasets are under the control of the laboratory, but only a few datasets, such as SFEW2.0, are collected under natural conditions, which is more consistent with real application scenarios. At the same time, such as AffectNet and FER-2013, the data is collected through network channels and reasonable filtering methods, so a large number of

DOI: 10.54254/2755-2721/6/20230902

data sets can be obtained. In addition, most data sets only capture the front of the face when collecting images, and most of them imitate basic emotional expressions rather than spontaneously, such as JAFFE, TFDO, ulu CASIA, which actually makes the structure of the database single and relatively low quality, while data sets such as Bosphorus, Multi PIE contain images under different light, different angles and even different attitudes, In fact, it can well train the deep network structure that has achieved good results in face recognition tasks. In addition, large data sets with occlusion type and head pose annotation can better meet the requirements of learning features with efficient expression recognition ability.

Table 1. Data set of common facial expressions.

Name	Static or Dynamic	Classification	Mode	Features	Sample objects	Number of data
JAFFE [5]	Static	7	Imitate	The data sources are all women, only the front of the face is included. Image expression intensity is high and easy to recognize.	10	213
CK+ [6]	Dynamic	8	Imitate spontaneous	Frontal dataset, the sample is highly included, and the sample source individuals are from 18 to 50 years old, mostly women.	210	593
TFD [7]	Static	7	Imitate	Only the front of the face is included	—	112234
Oulu-CASIA [8]	Dynamic	7	Imitate	Frontal dataset, Each video experiences 3 lighting conditions.	80	2880
BU-4DFE [9]	Dynamic	7	Imitate	The video includes the side of the head and can analyse the facial behaviour from 3D static to 3D dynamic: 3D+time.	101	606
Bosporus [10]	Dynamic	—	Imitate	Including 81 poses and occlusion conditions, 24 key points are marked for each sample	105	4666
Multi-PIE [11]	Static	6	Imitate spontaneous	15 visual angles, 19 lighting conditions	337	750000
SFEW2.0 [12]	Static	7	Film screen capture	It is not tested in the ideal environment of the laboratory, and the data source is more authentic	—	1766
AffectNet [13]	Static	8	—	It is obtained by querying different search engines with emotion related tags, and the data sources are extensive.	—	1000000
FER [14]	Static	7	Imitate spontaneous	It also includes real faces and cartoon faces, with a wide range of recognition	—	35887

DOI: 10.54254/2755-2721/6/20230902

2.2. Traditional expression recognition methods

Traditional expression recognition methods mainly use artificial design features, which requires more artificial participation. There are many expression recognition methods, which can be classified according to different image properties, and the commonly used methods are different in different situations. For dynamic images, optical flow method is often used. The static image often uses Gabor wavelet method or principal component analysis PCA, which is relatively easy to extract features. It can also be divided and applied according to whether the face is blocked before recognition, or according to the static or dynamic image. The classification of specific methods and their advantages and weakness are detailed in Table 2.

Table 2. Classification of traditional expression recognition.

Classification of methods	Main Methods	Method Description	Advantages	Disadvantages
Methods based on geometric features	Active Shape Model (ASM) [15], Active appearance model (AAM) [16]	After capturing the main structure of the face, the external contour figure of the face and the geometric feature points of the main facial key points are often extracted by point distribution model. The parameters of the statistical model are constantly optimized according to the search results, and finally the model matches the expression. AAM adds the shape and texture information on the basis of ASM to optimize the search process.	It is intuitive and suitable for traditional unshielded scenes, and can extract facial features effectively.	Need to match other algorithms to get better use effect, Angle, face size and other recognition information loss, the recognition accuracy is low
Extraction method based on frequency characteristic rate	Gabor [17]	The direction, centre frequency and base band width of Gabor wavelet filter are adjustable. It is a positive spin curve that exists in a complex domain. By adjusting different parameters of Gabor filter, information from different spatial positions, spatial frequencies and orientations in the image can be captured.	This method can get the related characteristic at separate scales and aspects in the frequency domain. It can effectively extract image features with different levels of detail	As the feature is of low level, the dimension of the feature is generally large, it is difficult to find appropriate parameters, and it is not easy to be directly used for feature matching.

Table 2. (continued).

Overall statistical characteristics	Principal component analysis (PCA) [18]	The main idea is to traverse the entire image, extracting as much as possible. The feature information of the whole image. Is to select significant feature points of the face for motion estimation. PCA uses covariance to analyse the correlation and realize the	It can eliminate the information and noise with interference in the image.	It only extracts part of the feature point information and ignores other parts of the face, which may lead to information loss. At the same time, it requires a

DOI: 10.54254/2755-2721/6/20230902

		orthogonal change of matrix to achieve dimension reduction.		large amount of storage space, and the calculation is highly complicated.
Extraction method rested on motion feature	Optical flow method [19]	Optical flow algorithm is the use of face rotation, under the same light, different parts of the face to produce different optical flow, which is different from the movement of different areas of the face picture, the optical flow effect is different, based on these differences can be used to light flow information face judgment.	It can extract the motion features of continuous sequences, and effectively extract the image features of different degrees of detail, so the interference of lighting factors is less	It's a lot of computation.
Based on texture features	Local binary method (LBP) [20]	The feature description of image usually adopts multi-region histogram and the similarity of the two images is judged by calculating the distance of the LBP histogram of the two images, which has good rotation and gray invariance.	LBP has invariance to illumination, relatively simple, small memory consumption, better extraction of local face information.	The sample is highly dependent, and the accuracy rate will decrease when the image is polluted.

Due to the complexity of face recognition itself, the impact of light and obstacles in the real environment, and the constraints of training large-scale face datasets, algorithms and computing performance, these shortcomings of traditional face recognition methods greatly reduce the accuracy of face recognition. Intermittently, many researchers give up using these traditional methods.

2.3. The expression classification methods

After feature extraction for face, it is necessary to select an appropriate classifier to classify features. Classifying the extracted features is helpful to improve the precision of facial expression recognition It can also effectively diminish the over fitting conditions caused by too few databases and deep learning. The typical traditional machine learning classification methods include K-NN, Bayesian network classifier, support vector machine, etc. The deep learning-based methods include convolutional neural network, DBN, etc. In addition, classification methods can also be divided into static and dynamic. The specific classification methods and their advantages and disadvantages will be detailed in Table 3.

Table 3. Classification of typical expressions.

Classification method	Method introduction	Advantages and disadvantages
K-NN [21]	The samples are randomized into class k, and sample tags with relevant	The algorithm can still be implemented simply when there are few data sets and no

DOI: 10.54254/2755-2721/6/20230902

	attributes are separated by adjusting the distance between the samples and the mean.	training set is needed to train. However, the classification efficiency of this algorithm is low, which is greatly affected by the actual situation. The weights of each attribute are the same when the new sample is compared with the training set, resulting in a certain reduction in classification accuracy.
SVM (Support Vector Machines) [22]	Mapping training samples to another high-dimensional space in supervised learning mode is achieved through a non-linear transformation, and to achieve expression classification by calculating the distance between samples on the support plane. At the same time, although the sample information in the data set is limited, it can find the optimal solution and obtain the best generalization ability when facing the complex model.	It can effectively deal with nonlinear, small sample, high-dimensional data and other problems, and has faster operation speed. Compared with easy over fitting of artificial neural network, SVM has better generalization ability for test samples that have not been seen before.
Adaboost [23]	It mainly solves the problem of binary classification. The best weak classifier of the weight distribution of the current training sample set is often found by constantly repeat the whole processand then adjusts the weight parameters. After meeting the number of iterations of the threshold, a strong classifier is often composed of several diverse weak classifiers.	The detection speed is fast and the detection precision is high. As the whole process continues to repeat it is not easy to have the problem of over fitting.
Bayesian network [24]	It is a probabilistic network based on statistics. The network is graphed through probabilistic reasoning, and the probability of the unknown expression class is inferred from the information of known expressions.	In the face of small sample data sets, it is insensitive to missing data, responsive, easy to train, not easy to overfit, and better handle uncertainties, but if the input variables are relevant, problems will arise.

DOI: 10.54254/2755-2721/6/20230902

Table 3. (continued).

CNN [25]	It is an improvement of artificial neural network.	In order to reduce network parameters so that the training speed becomes faster and can be applied to more real scenes, image pixels will be selected as the input of the network directly, which can also have the regularization effect. Because it can also effectively reduce the dimension of the image, which helps to reduce the workload of the computer. However, when there is too little facial expression training data, it is easy to lead to poor generalization ability of the model.
DBN [26]	The neural network is decomposed into layers of multiple RBMs, and then the layers are layered train	It can learn the abstract features of each layer from top to bottom, and use automatic learning to complete feature extraction, which can settle the problems of position and resolution in face looking recognition. But because the pixel of facial image is directly used as the input of learning, the local features of the portrait are often ignored. In complex environments, the output feature expression accuracy is poor.

3. Expression recognition based on multimodal fusion

The modes of expression recognition are mainly image, text, audio and video. At present, there are three multimodal integration techniques that are widely employed: feature fusion, decision fusion and hybrid fusion.

3.1. Feature fusion

Feature-level fusion is to extract and pre-process the features of various modes, then fuse them into a whole in a certain way, and finally put them into classifiers for classification and expression recognition. Because the feature level fusion adopts the fusion method of series direct mosaic, which is likely to cause the high feature dimensions, increase the calculation difficulty, and ignore the correlation between various features in the fusion process, leading to too much redundant information input and wastes resources. Block diagram of the feature level for fusion shown in figure 1.

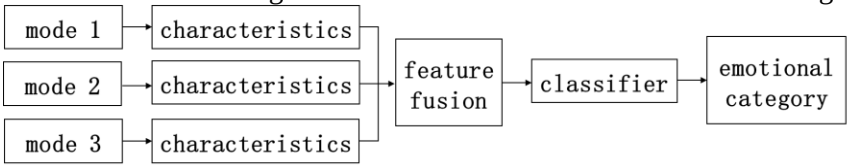

Figure 1. Block diagram of the feature level for fusion.

In reference [27], an improved expression recognition method of LBP and LDP features is used, that is, the improved LBP features and LDP features in the local region are extracted respectively. The feature level fusion method is used to concatenate these features in sequence, then put them into the classifier. Finally, experiments are carried out on JAFFE expression database to verify that this method can effectively improve the accuracy of expression recognition. The overall algorithm flow chart is shown in figure 2.

DOI: 10.54254/2755-2721/6/20230902

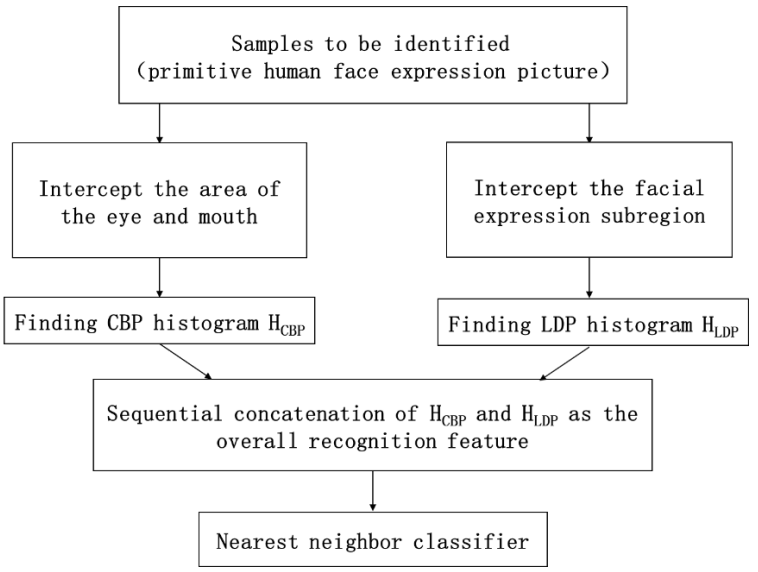

Figure 2. Overall flowchart for an algorithm.

In reference [28], according to the construction of facial expressions, LBP, LVP and CLVP algorithms are used to extract the features of different types of facial expressions. The features extracted by the three methods are combined, the cuckoo search method is used to cluster the features, and the ELM is used to classify the facial expressions. This increases sensitivity and accuracy while decreasing time usage and misclassification rates.

Because the feature level fusion is relatively simple in the fusion mode, the above two methods focus on feature extraction, both of which use the improved LBP and other local feature extraction algorithms, which are simple to calculate, have strong anti-interference and discrimination ability, and also have high recognition in the scene of light change. In the fusion method, the extracted features are simply concatenated in series.

3.2. Decision fusion

The decision level fusion is also to extract the modal features, and then use the classifiers to process the features in parallel, and then fuse the results of each classifier. The decision level fusion sets the credibility of the decision based on a certain voting mode, and selects the optimal decision. This fusion method considers the correlation between features well and reduces the amount of computation, so it has good real-time performance, but it is also because of the trade-off of features, which will lead to the decline of accuracy, a strong dependence on the previous level, and poor anti-interference ability. Block diagram of the decision level for fusion is shown in figure 3.

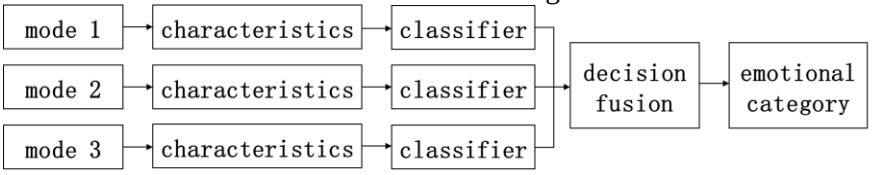

Figure 3. Block diagram of the decision level for fusion.

Reference [29] adopts decision-making level fusion. First, geometric pre-processing is carried out to normalize the scale of facial expression image, in preparation for extracting LDP features, ASM differential texture feature extraction is used to model the face and describe the facial reference point. The shape model is then used to create a gray model for each feature point, and each feature point's gray model is utilized to find the ideal position for the feature points in the target image to extract the features. At the same time, the LDP feature is also extracted, which describes the texture crying edge

DOI: 10.54254/2755-2721/6/20230902

information very well. Finally, the extracted features are combined at the decision level based on DS evidence theory. The decision-level fusion process for the two characteristics is shown in figure 4.

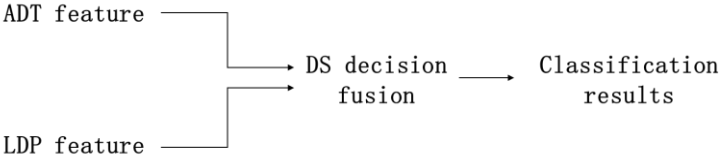

Figure 4. Fusion of two characteristics at decision level.

Reference [30] first normalizes the facial expression image, constructs the CNN model, uses the model to identify regional characteristics in the area, then sends the extracted local features to the support vector machine multiple classifiers for decision level weighted fusion, and uses the particle swarm optimization algorithm to obtain the optimal fusion weight. The trials' findings show that the approach has a high recognition accuracy and can ensure real-time recognition. Flowchart of expression recognition framework is shown in figure 5.

The important point of decision level fusion method is what kind of fusion strategy to choose. In reference [31], DS evidence theory is introduced into the decision level of the two features. The experimental results show that the recognition rate of DS decision level fusion is several percentage points higher than that of DT feature and LDP feature respectively. In reference [30], the local features extracted by CNN network are trained and recognized in the classifiers, and the posterior probabilities of three kinds of features are obtained for decision level weighted fusion. The global optimization of the three kinds of weights is carried out by particle swarm optimization algorithm. Finally, the decision matrix is then used to establish the expression category, considerably increasing recognition accuracy.

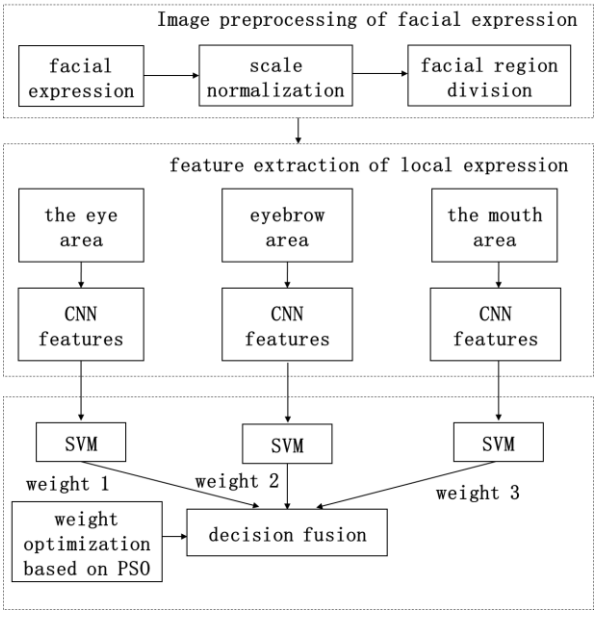

Figure 5. Flowchart of expression recognition framework.

3.3. Hybrid fusion

Hybrid fusion is the combination of feature level fusion and decision level fusion, the feature level fusion of some modal features, and the decision level fusion of some modal features. This method combines the advantages of the first two methods, loses less feature information, and can also consider the correlation between features, but this method increases the complexity of the model and the difficulty of training, so it is difficult to use. Hybrid fusion block diagram is shown in figure 6.

DOI: 10.54254/2755-2721/6/20230902

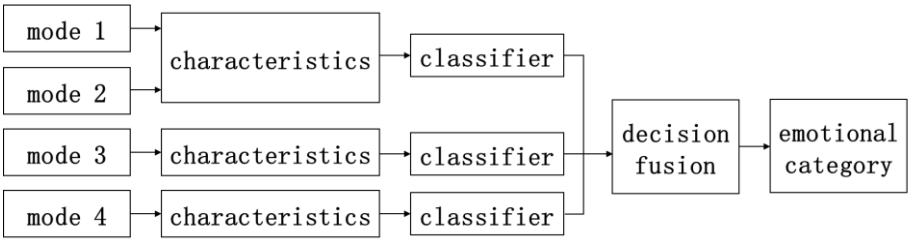

Figure 6. Hybrid fusion block diagram.

In reference [31], the modal features such as eye movement feature and PPG feature are introduced for multi-modal identification, and the multi-modal fusion method based on hybrid fusion is used for emotional analysis. Firstly, the eye movement features closely related to emotion coefficient are extracted by analysing Pierre coefficient, and the PPG features related to emotion are extracted by investigating HR-mean and the significance test index P of emotion category. Then the two features are merged at the feature level and then input into the FECNN-LSTM memory network to fully extract the deep timing information. Afterwards, the deep and shallow features of the two modes are combined at the feature level, and then the deep belief network DBN is used to fully mine and classify the correlation between the deep and shallow features.

4. Conclusion

In this study, single-modal and multi-modal fusion expression recognition are analysed. The single mode expression recognition procedure, the data set used in facial expression recognition, the conventional feature extraction methods of facial expression, and the often-employed expression classification methods are all briefly discussed in the single mode expression recognition. Three modal fusion methods—feature set fusion, decision set fusion, and mixed fusion—are examined in the recognition of multimodal fusion expressions. Both the decision set fusion approach and the feature set fusion method are very straightforward research techniques. The extraction of exterior elements like the face and limbs is the primary goal of feature set fusion. It takes into account the applicability of modes and extracts, analyses, and processes legitimate characteristics of eligible information. The fusion of decision sets is very straightforward to operate, but the data's properties are not thoroughly mined, and the outputs are just secondarily processed, making them more vulnerable to emotional influences. The benefits of the first two techniques are combined in hybrid fusion. In order for the recognized feature to receive both appearance and emotional labels, the feature sets of the physiological signal mode and the external posture signal mode are fused simultaneously, followed by the decision-making fusion.

Although network technology has continuously advanced this expression identification, there are still many flaws that need to be fixed by upcoming generation.

(1) The facial recognition expression collection lacks sufficient variety and quantity of expressions. When it comes to recognizing the same phrases, different age groups, races, and genders have variable outcomes. In the field of deep facial expression identification, facial occlusion and postural issues have gotten less attention, and there aren't enough relevant facial expression data sets. Low numbers of each modal data type, insufficient training, and insufficient modal data types for brain waves and other physiological signals all contribute to lower accuracy. Additionally, there are not enough models because it is exceedingly difficult to label the data for complicated natural scene changes. The information is unbalanced and skewed.

(2) There are clear distinctions between several database. The evaluation findings will vary from the real outcomes since evaluation algorithms within a single database are frequently not uniform. The complexity of generating and labeling various expressions varies at the same time, and some micro-expressions are frequently more challenging to recognize because they are challenging to detect. There is an imbalance because many recognitions can only identify the expression category when there are clear expression changes.

DOI: 10.54254/2755-2721/6/20230902

(3) Expression recognition accuracy and modal fusion technology are both subpar. The association between modes is frequently not thoroughly explored by current technology, and selecting too much data will reduce accuracy.

References

[1] Liu Bowen, Shuai Jianwei, Cao Yuping. Application of Facial expression recognition technology in the diagnosis and treatment of mental diseases. 2021 Chinese J. Be. Med. Br. Sci., 30 (10): 955- 960.

[2] Lai Dongsheng. Research and Application of Light scale Situation Recognition Algorithm based on Multi-feature fusion. 2022 Guangdong Univ. Tech.

[3] Xu Xiaokang. Research and Application of Expression Recognition Based on Deep Learning 2022, Donghua Univ.

[4] Wang Jin, Huang Xiaohua, Li Hang, Hong Jie. Application Research of Microexpression Recognition System in Low resolution Environment. 2022, Compute. Knowle. Tech., 18(20): 81-82+85.

[5] Lyons M J, Akamatsu S, Kama M, et al. Coding facial expressions with gabor wavelets 1998, Inter. Conf. Face & Gest. Rec., 14-16: 200-205.

[6] Lucey P, Cohn J F, Kande T, et al. The extended Cohn-Kanade dataset (CK+): a complete dataset for action unit and emotion-specified expression, 2010 Conf. Compute Vis. Pat. Rec., San Francisco, Jun 13-18: 94-101.

[7] Susskind J M, Anderson A K, Hinton G E. The Toronto face database, 2010, Toronto: Univ. Toronto.

[8] Zhao G, Huang X, Taini M, et al. Facial expression recognition from near-infrared videos, 2011, Ima. Vis. Comput. 2(9): 607-619.

[9] Yin L J, Wei X Z, Sun Y, et al. A 3D facial expression database for facial behavior research, 2006 7th IEEE Inter. Conf. Auto. Face Gest. Rec., 211-216.

[10] Savran A，Ala，Dibeklion H，et al. Bos phorus database for 3D face analysis, 2008 Euro. Biomet. Ident. Man.，Heidelberg：Springer, 47-56.

[11] Gross R, Matthews I, Conhn J, et al. Multi- PIE, Imag. Vis. Com., 2010, 28(5): 807-813.

[12] Dhall A,Goecke R, Lucey S, et al. Static facial expression analysis in tough conditions: data, evaluation protocol and benchmark 2011 IEEE Inter. Conf. Comp. Vis., Barcelona, Nov 6-13, 2011: 2106- 2112.

[13] Mistassini A, Hasa B, Mahoor M H. AffectN: a database for facial expression, valence, and arousal computing in the wild, 2019, IEEE Trans. Affi. Com., 10(1): 18-31.

[14] Goodfflow I J, Erhand D, Carrier P L, et al. Challenges in representation learning: a report on three machine learning contests, 2015, Neu. Net., 64: 59-63.

[15] Cootes T F，Taylo R C J，Coope R D H，et al. Active shape models-their training and application, 1995 Comp. vis. Image. Under. 61(1): 38-59.

[16] Tie Yun, Guan Ling. A deformable 3-D facial expression model for dynamic human emotional state recognition, 2013, IEEE trans. Cir. Sys. Vid. Tech, 23(1): 142-157.

[17] Lee T S. Image representation using 2D Gabor wavelets, 1996, IEEE trans. Pat. Anal. Mac. Intel., 18(10): 959-971.

[18] Zhu Y N, Li X Wu G H. Face expression recognition based on equable principal component analysis and linear regression classification, 2016 Inter. Conf. Sys. Infor., Nov 19-21: 876-880.

[19] Jiang Bo, Xie Lun, Liu Xin, et al. Microexpression Capture Based on Optical Flow Modulus Estimation, 2017, J. Zhejiang Univ., 51(3): 577-583, 589.

[20] Ahonet T, Hadida A, Pietik Inen M. Face recognition with local binary patterns. 2004, Compute. Vis., 469-481.

[21] Zhang F，Zhang T，mao Q，et al. Joint pose and expression modeling for facial expression

DOI: 10.54254/2755-2721/6/20230902

recognition 2018, Conf. compute vis. Pat. Rec, 3359-3368.
[22] Xuchao，Dong C，Feng Zhi, et al. Facial expression pervasive analysis based on Haar-like features and SVM 2012 Berlin Heidelberg: Springer, 521-529
[23] Viola P, Jones M. Rapid object detection using a boosted cascade of simple feature., 2011 Conf. Compute Vis. Pat. Rec.:511-518.
[24] Xie Lun, Lu Yannan, Jiang Bo, et al. Automatic Expression Recognition Based on Facial Motion Unit and Expression Relation Model, 2016, J. Beijing Ins. Tech., 36(2): 163-169.
[25] Gir R, Dong A J, Darr Llt T，et al. Rich feature hierarchies for accurate object detection and semantic segmentation 2014, Conf. compute Vis. Pat. Rec, 580-587.
[26] Hinton GE, Osinde R S, Teh YW. A fast-learning algorithm for deep belief nets. 2006, Neur. Comput., 18(7): 1527- 1554.
[27] Gong Qu, Ye Jianying, HUA TaoTao. Facial expression recognition based on improved LBP and LDP, 2013 Compute. Eng. Appl., 49(22):197-200.
[28] Wang S, Song J, Wang Meng, Wu S, Guan. Multi-feature fusion expression recognition algorithm based on referenced facial expression, 2021, Mod. Elec. Tech., 44(7):77-81.
[29] Xu Luhui. Facial Expression Recognition Based on the Fusion of ASM Different Texture Features and LDP Features.2015 Guangxi Normal Univ.
[30] YAO Lisha, XU Guoming, Zhao Feng. Expression Recognition Based on Local Feature Fusion of Convolutional Neural Network, 2017 Conf. Compute Vis. Pat. Rec 3259-3269.
[31] Chen Xinyi. Research on Multi-modal Fusion Emotion Recognition for Online Learning Scenarios. GuiLin Univ. Tech., 2022.