J. Vis. Commun. Image R. 57 (2018) 228–233

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Facial expression video analysis for depression detection in Chinese patients ☆

Qingxiang Wang [a,*,1], Huanxin Yang [b,1], Yanhong Yu [c,*]

[a] College of Computer Science and Technology, Qilu University of Technology (Shandong Academy of Sciences), Jinan, China
[b] College of Bioengineering, Qilu University of Technology (Shandong Academy of Sciences), Jinan, China
[c] College of Traditional Chinese Medicine, Shandong University of Traditional Chinese Medicine, Jinan, China

ARTICLE INFO

Article history:
Received 25 September 2018
Revised 3 November 2018
Accepted 3 November 2018
Available online 3 November 2018

Keywords:
Depression detection
Facial expression
Video processing
Eye movement
Feature extraction

ABSTRACT

Emotional state analysis of facial expression is an important research content of emotion recognition. At the same time, in the medical field, the auxiliary early screening tools for depression are also urgently needed by clinics. Whether there are differences in facial expression changes between depressive patients and normal people in the same situation, and whether the characteristics can be obtained and recognized from the video images of depressive patients, so as to help doctors detect and diagnose potential depressive patients early are the contents of this study. In this paper, we introduce the videos collection process of depression patients and control group at Shandong Mental Health Center in China. The key facial features are extracted from the collected facial videos by person specific active appearance model. On the basis of locating facial features, we classified depression with the movement changes of eyes, eyebrows and corners of mouth by support vector machine. The results show that these features are effective for automatic classification of depression patients.

1. Introduction

Depression is the leading cause of ill health and disability worldwide and more than 300 million people are now living with depression. Depression will become the second leading causes of disability and death in humans in 2020. The 2015 Report of Disease Prevention and Control Progress in China showed that 4.376 million patients with psychosis had established health records by the end of 2014. At the same time, the misdiagnosis, missed diagnosis and recurrence rate are high. Because of the high incidence of depression, the auxiliary early screening tools are urgently needed for depression clinics.

At present, depression is mainly assessed by self-report scales or clinician's rating scales. self-report scales and inventories (Self-RIs, e.g., PHQ-9) are often used for self-assessment, but the Self-RIs are susceptible to subjective factors and are insufficient to support the diagnosis of depression [1]. The clinician's rating scales (e.g., HRSD) rely on clinical skills and professional knowledge, and special training is required. "There is no blood test" for depression [1]. A large number of literatures have studied depressive disorders in neuroanatomy, endocrinology, and physiology, but there are no methods that can be used as a diagnostic tool for depression, as mentioned in DSM-5 [2].

Although there is a large literature describing the neuroanatomy, neuroendocrinology, and neurophysiology of major depressive disorders, no laboratory tests have yielded sufficiently sensitive and specific results as a diagnostic tool for this disease. Therefore, the search for direct and accurate objective evaluation methods is still an important part of multi-medical and interdisciplinary research.

Eyes and expressions are important ways for human beings to express their emotions. People's emotions can be expressed and reflected by the state of the eyes and a group of interrelated expressions. For example, in daily communication, only 7% of the total information is transmitted by language, while 55% of the total information is transmitted by facial expressions [3]. Clinically, depressive episodes have some observable external manifestations. Experienced medical professionals can identify potential depressive patients by direct observation during contact. The Diagnostic and Statistical Manual of Mental Disorders (DSM-5) [2] believes that it is possible to infer the presence of depression from human facial expressions and behaviors in some individuals. This indicates that there are direct observable differences in facial expressions between depressive patients and normal people.

☆ This article is part of the Special Issue on TIUSM.

* Corresponding authors.
E-mail addresses: wangqx@qlu.edu.cn (Q. Wang), yhysdutcm@sdutcm.edu.cn (Y. Yu).

[1] Co-first authorship: The first and second listed authors contributed equally.

https://doi.org/10.1016/j.jvcir.2018.11.003
1047-3203/

Depression analysis of facial expression video is an important research content of emotion recognition, and has become one of the hot topics in recent years. Whether depressive patients have different facial expression changes in the same situation as normal people and whether they can be detected and identified from video images are the main contents of the study. The results can help doctors to identify potential depressive patients and make early diagnosis.

Affective computing is a highly integrated technology area. Up to now, some progress has been made in facial expression, posture analysis, speech emotion recognition and expression. Correspondingly, the number of publications on depression detection by visual cues has also increased year by year. Since 2013, the annual Audio/Visual Emotion and Depression Recognition Challenge (AVEC) competition has attracted more and more attention, indicating that the study of machine recognition of depression is emerging and has become an active research field.

Jeffrey F. Cohn et al. [4] have studied whether depression can be recognized by facial movement in 2009. They collected the video and audio data of patients, then calibrated the video through FACS and active appearance model respectively, and analyzed them with support vector machine. They drew a conclusion: the similar accuracy rate (about 79%) of depression detection could be achieved by using artificial FACS parameters or the active appearance model. Gordon [5] further divided AUs into three groups of RU, and used active appearance model to get AUs and analyze RU. Dhall et al. [6] proposed a temporally piece-wise Fisher vectors method to perform the LBPTOP descriptors of face videos. In AVEC 2014, Williamson et al. [7] used bimodal analysis of facial motion units and speech spectrum. Many other works also used AUs to analyze facial expression movements for depression detection [8–11].

There was also a lot of work focused on the proportion of specific expressions and eye characteristics. On expressions, Scherer [12] found that depressive patients had fewer smiles, and Girard [8] found that they had fewer frowns and more tight corners of the mouth. Alghowinem [13] concluded that depression had shorter blinking intervals by analysing eye movement and blinking rate and used Gaussian Mixture Models and Support Vector Machines to identify it. Lucas GM [14] believed that depression patients had fewer smiles and more frowns. Jeffrey F. Cohn [4] found that the classification method had higher recognition accuracy by using AU 14 (AU 14 for tightening the corners of the mouth). Song [15] combined the human behaviour primitives, such as AU, gaze direction, to detect depression.

Some studies combined facial features with head movements or sounds to identify depression [16]. Jyoti [17] analysed head postures of subjects on a dataset created by Gordon. The results showed that depressive patients had fewer head movements than normal. Scherer [12] found that the vertical eye gaze of depression patients is lower and Alghowinem [18] found that the duration of looking down is longer. Girard [8] found that the amplitude and speed of the head in severe depression decreased. Alghowinem [19] used head posture, eye focus, Paralinguistic to detect depression. Jyoti [20] used audio and video for multimodal analysis, LBP-TOP and spatio-temporal interest point to process video, and FFT to process the audio. Kaya [21] used correlation analysis and Jain [22] used Fisher Vector to construct the dual-mode frame of voice and video.

Most of the work choose support vector machine as the classification method, because SVM is more suitable for high-dimensional binary classification problem [1]. Others have used Nearest Neighbour, Gaussian Mixture Models, Neural Networks, Random Forest, HMM and so on. In the newer research, Song used CNN [15]. In their work, the database of AVEC 2016 (almost the largest database in published work) was used and deep CNN has not shown its obvious advantages because of the limited number of training data.

The boundaries between normality and pathology vary across cultures for specific types of behaviours. Thresholds of tolerance for specific symptoms or behaviours differ across cultures, social settings, and families [2]. The vast majority of the existing research databases used for automatic depression classification based on visual cues in current researches is based on specific groups of people.

The available databases mainly include Pittsburgh, BlackDog, DAICWOZ, ORI, ORYGEN, CHI-MEI, etc. [1]. Pittsburgh participants were Euro- or African-American [23]. Native Australian English speaking participants were in BlackDog [19]. The DAICWOZ database had been used in AVEC 2016 [24] and the subset of the SEWA database had been used in AVEC 2017 [25]. In DAICWOZ, All participants were fluent English speakers and all interviews were conducted in English [26]. SEWA database was the video chat recording of German subjects. ORI used the video corpus of Oregon Research Institute (ORI) in USA and all the subjects selected had white skin [27]. CHI-MEI collected the patients in Chi-Mei Medical Center [28], but their work aimed to distinguish the unipolar depression and bipolar disorder. There is still a lack of Chinese databases.

In this paper, we introduced the data collection process of depression patients and control group at Shandong Mental Health Center in China. And then the data was preliminarily processed. We extracted the key motion features of the eyes, eye brows and corners of mouth in the facial expression sequences, and used SVM to classify the features. The experimental results showed that our method had achieved good results in the recognition of depression.

2. Methodology

2.1. Participants and data acquisition

The study was approved by the hospital ethical committee. All participants provided written informed consent before entry to the study.

2.1.1. Participants

Our clinical video samples used in this paper were from 26 hospitalized patients (16 males and 10 females) who had been diagnosed with depression at "Shandong Mental Health Center" in China. All the participations are Chinese. The participations were diagnosed with depression according to diagnostic criteria of the World Health Organization (ICD-10) by psychiatric clinicians with many years of practical experience. They were also interviewed to assess symptom severity by using the Hamilton Rating Scale for Depression (HRSD $\geq$ 20) [29]. Their age was between 18 and 60. They had not been treated for nearly two weeks of antidepressant drugs. Although they had other diseases, the diseases should have no direct contact with the depression. At the same time, we had to rule out some people: They are pregnant, lactating women; they also suffered from other psychiatric disorders, such as bipolar disorder and schizophrenia; they currently suffered from other serious diseases. At the same time, 26 healthy people were recruited as control group. The participations were voluntarily participated in this study and signed a written informed consent by himself or by the legal guardian.

2.1.2. Data acquisition

In the experiment, a Canon 600D camera was used to record the facial expressions. The resolution of camera photo is 1509 W pixels (5184 $\times$ 2912). The resolution of camera video is 1080P, 25fps.

The experimental was conducted in a well-lit room. The participations were seated at a distance of 1 m in front of the camera. A monitor screen which will play the experiment content like basic facial expressions and emotional pictures to participations is placed in front of the subjects. The camera is beside the monitor and the deviation angle is less than 10 degrees. At the same time, the head of participations and the collection equipment are in the same height.

The experiment was divided into three stages. The first stage was to collect the basic facial expression. The second stage was to answer the specific questions. The third stage was to watch the emotional pictures. After the completion of each stage of the experiment, the clinicians gave the participations thanks and rewards.

Before the experiment, the clinicians explained the rules to the participations, and the participations understood the experiment. Then the participations entered the laboratory to adapt to the environment.

2.1.2.1. Collecting the basic facial expression. Studies have shown that the expression of basic emotions is universal [30]. Ekman studied the facial expressions of Westerners and New Guinean primitive tribesmen who had never had any contact with Western culture. He asked respondents to identify pictures of various facial expressions and use them to convey their identified emotional states. He found that certain basic emotional expressions were expressed in two cultures. There are six basic emotions: sadness, disgust, fear, surprise, anger and happiness. In our data collection, the patients and the controls group were shown the six basic expressions in turn. In order to compare, the neutral images of relaxation were also collected.

The clinicians played a neutral expression and the six basic expressions (including sadness, disgust, fear, surprise, anger and happiness) on the computer screen, and asked the participations to imitate their expressions with reference to neutral and basic facial expressions. At the same time, these seven expressions images were captured by the camera with a resolution of 5184×2912 and the experimental procedure was recorded in videos by the collection equipment. Some captured images are shown in Fig. 1.

2.1.2.2. Answering the specific questions. In this stage, two well trained clinicians tested the participations with HRSD to measure the degree of depression.

2.1.2.3. Watching the emotional pictures. We chose three types of pictures as stimuli from the International Affective Picture Library [31] to stimulate emotional expression. Each of the participations completed the International Affective Picture System Test. The pictures were selected from the International Affective Picture Library which was divided into positive emotional pictures, neutral emotional pictures and negative emotional pictures according to their grade. We selected 10 pictures from each type. In order to prevent excessive stimulation of patients, the lowest negative grade of the pictures were not used in our experiment. A set of negative, neutral and positive pictures was shown in Fig. 2. In each emotional induction process, each picture was presented on the screen for six seconds, and each of pictures was randomly presented one time. Thus the time of each kind of the emotional stimulation was one minute. At the same time, the collection equipment (Canon camera) was used to record the video of participations' facial expressions.

2.2. Experimental method

The first step of our process was to locate the face and facial landmarks. Landmarks refer to points that define the shape of permanent facial features, such as the eyes. The feature points of the facial expression images are artificial marked according to FACS encoding system. After marked, the image set was trained by person specific Active Appearance Models (AAM) [32]. Secondly, the video frames of the participations were automatically matched by using the person specific AAM model. Then we selected some features of the eyebrows, eyes and corners of mouth from the

Fig. 1. Facial expressions images.

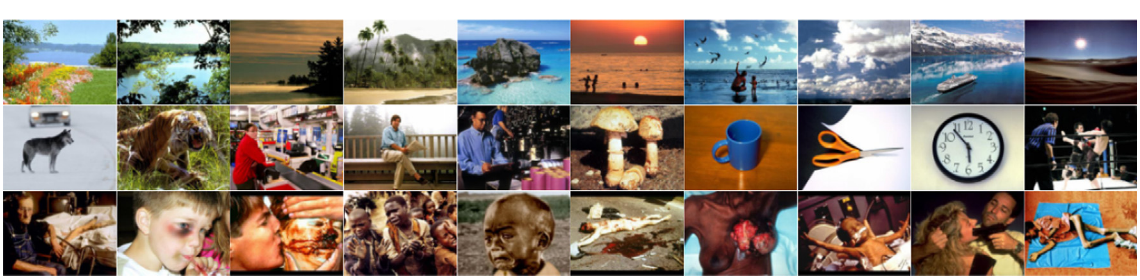

Fig. 2. A set of positive, neutral and negative pictures used in our experiment.

matched results of AAM and measured their location changes. At last, we trained the classifier with these features.

2.2.1. Basic facial features and facial feature points in facial expressions of depression patients

Each subject had a neutral expression and the six basic expressions. The feature points of these facial images of the participations were artificial marked with 68 points which were defined in Xm2vts [33] frontal face data, as shown in Fig. 3. After the marking step, the images were aligned to the basis of the feature points. Next, the images and marked points were trained by AAM, so as to generate a person specific AAM which included all the participations to improve matching accuracy.

2.2.2. Facial feature points tracking in video

In this step, we dealt with each frame of the videos. First, the Viola-Jones face detector was used to detect the human face region in the frame which was used for the initialization of the person specific AAM. Then, the current positions of the feature points were obtained by using the AAM on each frame by Eq. (1)

$$\arg\min(\sum_{x\in S_0}\left[A_0(x)+\sum_{i=1}^{m}q_iA_i(x)-I(N(W(x;p);u)\right]^2 \quad (1)$$

where S_0 is the mean shape, $A_0(x)+\sum_{i=1}^{m}q_iA_i(x)$ is the estimated texture, $I(N(W(x;p);u)$ is the texture obtained by mapping the input image with the currently estimated shape parameters p to the basis of the feature points, u is the translation and rotation scaling parameters.

Fig. 3. Feature points.

The matching results are shown in Fig. 4.

We had treated a total of 52 videos, including depression and control groups segments. The format of the matching results in each frame were {frame number, x_0, y_0, …, x_n, y_n}, (x_n, y_n) represented the transverse and vertical coordinates of the N point (Origin is in the upper left corner, a total of 68 points. Left and right pupils are 31 and 36 point respectively).

2.2.3. Feature extraction

Facial expressions and eyes feature points are selected to detect the depression, including eye pupil movement, blinking frequency, and movement changes of bilateral eyebrows and corners of mouth.

Eyes: The distance between the left pupil (31, the position of each point can be seen in Fig. 3) and left inner eye corner point (29), The distance between the right pupil (36) and the right inner eye corner point (34); The blink frequency of eyes in one minute in the video which is counted manually.

Eyebrows: the distances between the medial three feature points (23, 24, 25) on left eyebrow and the left inner eye corner point (29) which is invariant relative to the face; the distances of the medial three feature points (17, 18, 19) on the right eyebrow and the right inner eye corner point (34); The distance between the feature point on the left side of the eyebrow (21) and left inner eye corner point (29); The distance between the feature point on the right side of the eyebrow (15) and right inner eye corner point (34).

Corners of mouth: the distances between the nose tip point 67 and six characteristic points (48, 49, 59, 53, 54, and 55) around the corners of mouth.

The maximum, minimum and standard deviation of these distances mentioned above are used to measure the degree of facial expression changes. At the same time, in order to eliminate the difference between different faces and the projection difference caused by the distance from the camera, we divide those maximum, minimum and standard deviation distance values by the mean distance.

A total of 49 statistical features were extracted, which are:

Standard deviation/mean, maximum/mean and minimum/mean of the 8 points on eyebrows (3×8)
Standard deviation/mean, maximum/mean and minimum/mean of 2 eye pupils (3×2)
Standard deviation/mean, maximum/mean and minimum/mean of the 6 points on eyebrows (3×6)
Blink frequency of eyes (1×1)

2.2.4. Classification method

We hope to identify whether a participant is depressed. It is a binary classification problem to classify the participant by the selected features. In our work, the feature vectors are classified by support vector machine (SVM) which is suited for facial expression classification [34]. The SVM algorithm is to find the best sep-

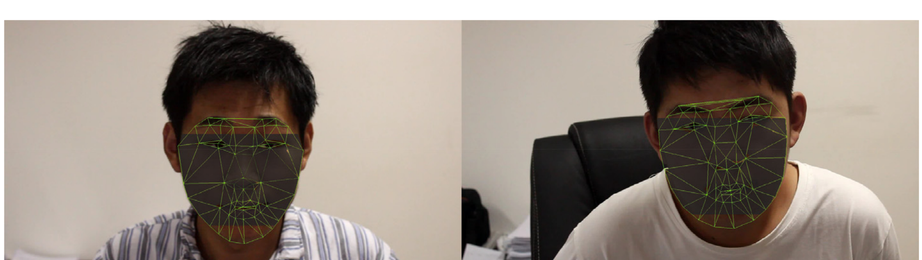

Fig. 4. Matching results (two patient video frames).

aration hyperplane in the feature space to maximize the interval between positive and negative samples in the training set.

The Radial Basis Function (RBF) is used in the SVM classifier and Chih-Jen Lin's Libsvm tools [35] are used for parameter optimization.

The kernel is given by:

$$K(u, v) = \exp(-\gamma^* \| u - v \|^2) \tag{2}$$

3. Results

The extracted vectors of eyebrows, pupils, corners of mouth and blinking frequency of depression group and control group are input into SVM model and a binary classifier is trained for the classification.

All the data are randomly divided into two groups, each group consisting of 26 participations, half of which are patients and half are controls. We train SVM classifier with these two sets of data respectively and validate them with another group. The results are shown in Table 1.

The performance metrics are calculated by the next equations:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

where TP is the number of true positives that depression patients are diagnosed as depression, TN is the number of true negatives that control volunteers are not diagnosed as depression, FP is the number of false positives that control volunteers are diagnosed as depression, and FN the number of false negatives that depression patients are not diagnosed as depression.

$$recall = \frac{TP}{TP + FN} \tag{4}$$

Positive predictive value (precision) is defined as a ratio of true positives to the total number of positives predicted by the model and specificity is defined as a ratio of true negatives to the total number of control group. F_1 is the F-measure when the parameter is 1.

$$precision = \frac{TP}{TP + FP} \tag{5}$$

$$specificity = \frac{TN}{TN + FP} \tag{6}$$

$$F_1 = 2 * \frac{precision * recall}{precision + recall} \tag{7}$$

Table 1
SVM classification results.

Classification	First group		Second group	
	Depression	Control	Depression	Control
Positive	10	3	11	3
Negative	3	10	2	10
Total	13	13	13	13

Table 2
Performance metrics results.

SVM performance metrics	Results
Accuracy	0.7885
Recall (sensitivity)	0.8077
positive predictive value (precision)	0.7775
Specificity	0.7692
F_1	0.792

The average performance metrics of the two groups are shown in Table 2.

The accuracy is 78.85% and recall is 80.77% and F1 is 0.792.

A leave-one-out validation is also used in the test to verify the effectiveness of our method for depression detection. That is, each sample is used as test sample one by one, while other samples are used as training set. After automatically parameter optimization by Libsvm, the training set is trained as SVM model to classify the test sample. Because there is only one patient in the test set, the result is only one possible, either positive or negative. The leave one classification results show that 23 of 26 depressive patients are correctly identified with the SVM model trained by the other 51 samples, 3 are incorrectly identified as normal, while 20 of 26 control volunteers are correctly identified and 6 are incorrectly identified as depression.

4. Conclusion

Depression is a serious mental illness, and the current diagnosis process still needs to be conducted by a specially trained psychiatrist or psychologist, usually using a scale and careful observation in communication, which depends on the doctor's experience. And it's hard for non-psychiatrists to diagnose and treat depression. A study has found that about two-thirds of cases were missed [36].

Capturing videos including eyes and faces, extracting and recognizing the features of the captured videos may help patients themselves or community doctors to detect and diagnose potential depressive patients early or to improve the diagnostic rate of depression. This is what we expect to achieve in our research.

At present, most of the studies by the visual direction are targeted at specific culture groups, but due to cultural differences, there are certain differences in the external manifestations of depression among different cultural groups [37,38]. This study starts with the Chinese domestic case data, which is helpful to find the characteristics of auxiliary diagnosis suitable for the Chinese domestic population.

In this paper, we analyze facial expressions and eye movement features for depression detection. In addition to eyebrow and corners of mouth which are closely related to facial expression, we also use pupil movement, blink frequency to identify depression to detect the depression. The results show that these features can achieve the detection accuracy rate at 78.85% and recall at 80.77%.

Relative to the number of depressed people, the number of samples is still relatively small. In the future, we will continue to collect cases and test more feature selection and classification methods.

As mentioned in DSM-5, depressive patients have some external manifestations, including agitation and retardation, such as inability to sit still, rub hands, and pull objects, reduced speech, and slow body activity, and so on. Some work also shows that the fusion of a variety of features can improve the detection accuracy of depression, such as voice, posture, etc. In the next work, we will take these factors into consideration and try to analyse and fuse the movement of body posture, audio and other facial expression changes.

Conflict of interest

There is no conflict of interest.

Acknowledgements

This work was supported by the Shandong Provincial Natural Science Foundation, China (Grant: ZR2016FM14), the National Natural Science Foundation of China (Grant: 81573829, 81703941)

References

[1] A. Pampouchidou, P. Simos, K. Marias, F. Meriaudeau, F. Yang, M. Pediaditis, M. Tsiknakis, Automatic assessment of depression based on visual cues: A systematic review, IEEE Trans. Affective Comput. (2017), 1–1.

[2] American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders DSM-V Fifth Edition, 5th ed., June 2013, pp 160–165.

[3] A. Mehrabian, J.A. Russell, An Approach to Environmental Psychology, MIT Press, Cambridge, 1974.

[4] J.F. Cohn, T.S. Kruez, I. Matthews, Y. Yang, M.H. Nguyen, M.T. Padilla, F. Zhou, F. D.L. Torre, Detecting depression from facial actions and vocal prosody, in: 3rd International Conference on ACII 2009, 2009, pp. 1–7.

[5] G. McIntyre, R. Gocke, M. Hyett, M. Green, M. Breakspear, An approach for automatically measuring facial activity in depressed subjects, in: 3rd International Conference on ACII 2009, 2009, pp. 223–230.

[6] A. Dhall, R. Goecke, A temporally piece-wise fisher vector approach for depression analysis, in: International Conference on Affective Computing and Intelligent Interaction, 2015, pp. 255–259.

[7] J.R. Williamson, T.F. Quatieri, B.S. Helfer, G. Ciccarelli, D.D. Mehta, Vocal and facial biomarkers of depression based on motor incoordination and timing, in: 4th International Workshop on Audio/Visual Emotion Challenge ACM, 2014, pp. 65–72.

[8] J.M. Girard, J.F. Cohn, M.H. Mahoor, S.M. Mavadati, Z. Hammal, D.P. Rosenwald, Nonverbal social withdrawal in depression: Evidence from manual and automatic analyses, Image Vision Comput 32 (10) (2014) 641–647.

[9] G. Stratou, S. Scherer, J. Gratch, L.P. Morency, Automatic nonverbal behavior indicators of depression and PTSD: the effect of gender, J. Multimodal User In. 9 (1) (2015) 17–29.

[10] S. Poria, A. Mondal, P. Mukhopadhyay, Evaluation of the intricacies of emotional facial expression of psychiatric patients using computational models, in: M. Mandal, A. Awasthi (Eds.), Understanding Facial Expressions in Communication, Springer, New Delhi, 2015.

[11] Michel Valstar, Jonathan Gratch, Björn Schuller, Fabien Ringeval, Denis Lalanne, Mercedes Torres Torres, Stefan Scherer, Giota Stratou, Roddy Cowie, Maja Panti, AVEC 2016: depression, mood, and emotion recognition workshop and challenge, in: 6th International Workshop on Audio/Visual Emotion Challenge, 2016, pp. 3–10.

[12] S. Scherer, G. Stratou, G. Lucas, M. Mahmoud, J. Boberg, J. Gratch, A. Rizzo, L.P. Morency, Automatic audiovisual behavior descriptors for psychological disorder analysis, Image Vision Comput. 32 (10) (2014) 648–658.

[13] S. Alghowinem, R. Goecke, M. Wagn.er, G. Parker, Eye movement analysis for depression detection, in: IEEE International Conference on Image Processing, 2014, pp. 4220–4224.

[14] G.M. Lucas, J. Gratch, S. Scherer, J. Boberg, Towards an affective interface for assessment of psychological distress, in: International Conference on Affective Computing and Intelligent Interaction, IEEE, 2015, pp. 539–545.

[15] S. Song, L. Shen, M. Valstar, Human behaviour-based automatic depression analysis using hand-crafted statistics and deep learned spectral features, in: IEEE International Conference on Face and Gesture Recognition, 2018, pp. 158–165.

[16] J.M. Girard, J.F. Cohn, Automated audiovisual depression analysis, Curr. Opin. Psychol. 4 (2015) 75–79.

[17] Jyoti Joshi, Roland Goecke, Gordon Parker, Michae Breakspear, Can body expressions contribute to automatic depression analysis?, in: 10th IEEE International Conference on Automatic Face and Gesture Recognition, 2013, pp 1–7.

[18] S. Alghowinem, R. Goecke, M. Wagner, G. Parker, M. Breakspear, Head pose and movement analysis as an indicator of depression, in: Humaine Association Conference on Affective Computing and Intelligent Interaction, 2013, pp. 283–288.

[19] S. Alghowinem, R. Goecke, M. Wagner, J. Epps, M. Hyett, G. Parker, et al., Multimodal depression detection:fusion analysis of paralinguistic, head pose and eye gaze behaviors, IEEE Trans. Affective Comput. 99 (2016), 1-1.

[20] J. Joshi, R. Goecke, S. Alghowinem, et al., Multimodal assistive technologies for depression diagnosis and monitoring, J. Multimodal User Interfaces 7 (3) (2013) 217–228.

[21] H. Kaya, A.A. Salah, Ensemble cca for continuous emotion prediction, in: 4th International Workshop on Audio/Visual Emotion Challenge, ACM, 2014, pp. 19–26.

[22] V. Jain, J.L. Crowley, A. Dey, A. Lux, Depression estimation using audiovisual features and fisher vector encoding, in: 4th International Workshop on Audio/Visual Emotion Challenge, ACM, 2014, pp. 87–91.

[23] H. Dibeklioglu, Z. Hammal, J.F. Cohn, Dynamic multimodal measurement of depression severity using deep autoencoding, IEEE J. Biomed. Health Inform. 99 (2017), 1–1.

[24] F. Ringeval, M. Pantic, B. Schuller, M. Valstar, J. Gratch, R. Cowie, et al., AVEC 2017: real-life depression, and affect recognition workshop and challenge, in: The Audio/Visual Emotion Challenge and Workshop, 2017, pp. 3–9.

[25] M. Valstar, J. Gratch, F. Ringeval, D. Lalanne, M.T. Torres, S. Scherer, G. Stratou, R. Cowie, M. Pantic, In: Avec 2016: depression, mood, and emotion recognition workshop and challenge, 2016, pp. 3–10.

[26] J. Gratch, R. Artstein, G.M. Lucas, G. Stratou, S. Scherer, A. Nazarian, R. Wood, J. Boberg, D. DeVault, S. Marsella, D.R. Traum, The distress analysis interview corpus of human and computer interviews, in: Language Resources Evaluation Conference, 2014, pp. 3123–3128.

[27] N.C. Maddage, R. Senaratne, L.S. Low, M. Lech, N. Allen, Video-based detection of the clinical depression in adolescents, in: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE Engineering in Medicine and Biology Society Annual Conference, 2009, pp. 3723–3726.

[28] T.H. Yang, C.H. Wu, K.Y. Huang, M.H. Su, Coupled HMM-based multimodal fusion for mood disorder detection through elicited audio–visual signals, J. Ambient Intell. Hum. Comput. 8 (6) (2016) 895–906.

[29] M. Hamilton, A rating scale for depression, J. Neurol Neurosurg. Psychiatry 23 (1) (1960) 56–62.

[30] P. Ekman, Universals and Cultural Differences in Facial Expressions of Emotion, University of Nebraska Press, Lincoln, 1971.

[31] P.J. Lang, M.M. Bradley, B.N. Cuthbert, International Affective Picture System (IAPS): Affective Ratings of Pictures and Instruction Manual. Technical Report A-8, University of Florida, Gainesville, FL, 2008.

[32] T.F. Cootes, G.J. Edwards, C.J. Taylor, Active appearance models, IEEE T. Pattern Anal. 23 (6) (2001) 681–685.

[33] K. Messer, J. Matas, J. Kittler, K. Jonsson, XM2VTS: the extended M2VTS database, in: Second International Conference on Audio- and Video-Based Biometric Person Authentication, 2000, pp. 72–77.

[34] M.S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, Recognizing facial expression: machine learning and application to spontaneous behaviour, IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit. (2005) 568–573.

[35] C.C. Chang, C.J. Lin, LIBSVM: a library for support vector machines, Acm T. Intel Syst. Tec. 2 (3) (2011) 1–27.

[36] M. Cepoiu, J. McCusker, M.G. Cole, M. Sewitch, E. Belzile, A. Ciampi, Recognition of depression by non-psychiatric physicians - a systematic literature review and meta-analysis, J. Gen. Intern. Med. 23 (1) (2008) 25–36.

[37] Z. Zhang, P. Luo, C.L. Chen, X. Tang, From facial expression recognition to interpersonal relation prediction, Int. J. Comput. Vision 126 (5) (2016) 550–569.

[38] K. Zhang, Y. Huang, Y. Du, L. Wang, Facial expression recognition based on deep evolutional spatial-temporal networks, IEEE Trans. Image Process. 99 (2017), 1-1.

Qingxiang Wang received the M.S. degrees in computer software and theory from Donghua University, Shanghai, China, and the Ph.D. degree in computer software and theory from Shandong University, Jinan, China. Currently, he is an associate professor in College of Computer Science and Technology, Qilu University of Technology (Shandong Academy of Sciences), Jinan, China. His research interests include computer vision and pattern recognition.

Huanxin Yang Doctor of Shandong University of Traditional Chinese Medicine, major is the basic theory of Chinese medicine. He is a university lecturer of Qilu University of Technology (Shandong Academy of Sciences) College of Bioengineering.The current research direction is the modernization of traditional Chinese medicine and the research on the extraction, purification and detection of traditional Chinese medicine.

Yanhong Yu, Ph.D, Master's supervisor of basic theory of traditional Chinese medicine. She is currently an associate professor in College of Traditional Chinese Medicine, Shandong University of Traditional Chinese Medicine, Shandong, China. Her research interests include emotional expression recognition and Emotional disease.