Journal on Multimodal User Interfaces
https://doi.org/10.1007/s12193-019-00308-9

ORIGINAL PAPER

Multi-modal facial expression feature based on deep-neural networks

Wei Wei[1] · Qingxuan Jia[2] · Yongli Feng[2] · Gang Chen[2] · Ming Chu[2]

Received: 24 July 2018 / Accepted: 10 July 2019

Abstract
Emotion recognition based on facial expression is a challenging research topic and has attracted a great deal of attention in the past few years. This paper presents a novel method, utilizing multi-modal strategy to extract emotion features from facial expression images. The basic idea is to combine the low-level empirical feature and the high-level self-learning feature into a multi-modal feature. The 2-dimensional coordinate of facial key points are extracted as low-level empirical feature and the high-level self-learning feature are extracted by the Convolutional Neural Networks (CNNs). To reduce the number of free parameters of CNNs, small filters are utilized for all convolutional layers. Owing to multiple small filters are equivalent of a large filter, which can reduce the number of parameters to learn effectively. And label-preserving transformation is used to enlarge the dataset artificially, in order to address the over-fitting and data imbalance of deep neural networks. Then, two kinds of modal features are fused linearly to form the facial expression feature. Extensive experiments are evaluated on the extended Cohn–Kanade (CK+) Dataset. For comparison, three kinds of feature vectors are adopted: low-level facial key point feature vector, high-level self-learning feature vector and multi-modal feature vector. The experiment results show that the multi-modal strategy can achieve encouraging recognition results compared to the single modal strategy.

Keywords Emotion recognition · Multi-modal feature · Convolutional neural networks · Support vector machine

1 Introduction

Emotion recognition based on facial expression has been one of the most extensively studied topics in real world applications due to its great potential in intelligent human computer interaction. Many methods have been proposed to address the facial expression recognition problems, but it remains a challenging and attractive research subject in computer vision. Extensive studies have been carried out on the facial expression recognition in static images for a long time in the past [1]. Given a static facial expression image, the technology is to estimate the correct emotional state, such as happiness, sadness, anger and so on. It mainly consists of two steps: feature extraction and classification. For feature extraction, traditional low-level empirical features are commonly used, such as gray features, geometric features or motion features. For classification, SVM is the most commonly used. Emotion recognition based on deep learning has also become an active research topic in computer vision.

In recent years, multi-modal facial expression feature extraction has become a new research topic and received more and more attention [2]. Different from the traditional single modal feature, the aim of multi-modal feature is intended to improve the robustness. The multi-modal feature contains not only low-level empirical feature, but also high-level self-learning feature, which, together, can further enhance recognition performance. Although the multi-modal feature is useful, there are still challenges regarding how to extract the feature reliably and robustly. For instance, the self-learning requires a computer to learn essential feature from sample data directly and autonomously. In order to

✉ Wei Wei
wei_wei@bupt.edu.cn

Qingxuan Jia
qingxuan@bupt.edu.cn

Yongli Feng
fengyongli@bupt.edu.cn

Gang Chen
chengang_zdh@bupt.edu.cn

Ming Chu
buptchuming@163.com

[1] College of Information Engineering, Beijing Institute of Petrochemical Technology, Beijing, China

[2] School of Automation, Beijing University of Posts and Telecommunications, Beijing 100876, China

Published online: 17 July 2019

generalize the feature, a high-level feature extraction model needs to be established. As multi-modal facial expression feature extraction strategies are different, it remains an open issue how high-level feature can be extracted effectively from each facial expression image.

In this paper, a new multi-modal facial expression feature extraction strategy is presented. It is motivated by the fact that accurate facial expression recognition depends on high quality feature representations. A good feature representation should be discriminative to the changes of facial expression while remaining robust to intra-personal variations. However, the feature representation composed by empirical feature is too shallow to differentiate the complex nonlinear variations in facial appearance [3]. To handle this problem, recent works have resorted to Convolutional Neural Networks (CNNs) [4, 5] to automatically learn effective features that are robust to the nonlinear variations on facial appearance images. However, the existing works extracted features via a single modal, and did not make good use of the complementary information contained in multiple modalities [6]. Inspired by the complementary information contained in multiple modalities and the recent progress of deep learning in various fields of computer vision, this paper presents a novel feature representation framework by means of linearly fusing the two kinds of modal features. This method consists of two stages: feature extraction and classification. In the feature extraction stage, it first extracts 2-dimensional coordinate of facial key points as low-level empirical feature. Then, it uses CNNs to automatically extract high-level self-learning feature. At last, it linearly fuses two modal features and obtains the facial expression feature. In the classification stage, it obtains a classification model based on SVM.

In summary, the contributions of the paper are as follows: (1) A more advanced multi-modal feature extraction strategy is used. In previous methods, single modal feature was exploited for facial expression recognition. In order to overcome this limitation, this paper proposes a multi-modal facial expression representation strategy. (2) Small filters are utilized for all convolutional layers. Owing to multiple small filters are equivalent of large filter, which can reduce the number of parameters to learn effectively. (3) The proposed method is evaluated in a database which contains 7 kinds of emotions. Moreover, comparison results are carefully analyzed and studied on whether to use the multi-modal feature. The rest of the paper is organized as follows: Sect. 2 gives an overview of related works on multi-modal feature extraction of facial expression and classification of emotions. Section 3 describes the method used. Section 4 verifies the proposed method through experiment and analyzes the experimental results. Section 5 concludes the paper.

2 Related work

The recognition performance is highly dependent on the feature extraction results. Many novel methods have been proposed to extract features of facial expression images. Popular facial expression feature can be broadly grouped into two categories: low-level feature and high-level feature. For the former, most of the existing works utilized various artificial features, including Local Binary Patterns (LBP) [7], Histogram of Oriented Gradient (HOG) [8], Gabor Wavelet (GW) [9], Scale-invariant feature transform (SIFT) [10], Active Appearance Model (AAM) [11], Active Shape Model (ASM) [12] and so on.

For high-level feature, most of the existing works utilized the self-learning based theory. Gavin's powerful representation ability, and deep learning based theory have also been employed in the feature extraction tasks [13–15] recently. [16] extracted a set of deep feature maps by a pre-trained CNNs model from the input images, where the local deep features were densely collected. Compared with the commonly used DL algorithms [17, 18]. Shi et al. [19] proposed deep polynomial network (DPN) algorithm not only shows superior performance on large scale data, but also has the potential to learn effective feature representation from a relatively small dataset. In the real world, the features based on the self-learning strategy are more adaptable than the artificial features. As the whole image is employed as input, every part of the face is treated equally regardless of whether it is closely related to the target facial expression. In order to make the features more adaptable to the real world, the deep learning theory has been employed for emotion recognition based on facial expression [20, 21].

Fusion of features is an important branch of feature representation. Many researchers created a number of fusion strategies to boost the classification performance [22]. Luo et al. [23] adopted the Principal Component Analysis (PCA) to extract the global feature of facial images and Local Directional Pattern (LDP) to extract local texture features of eyes and mouth. Then the global feature is combined with local texture feature to form a fusion feature. In the proposed method [24], four effective descriptors for facial expression representation were considered, namely Local Binary Pattern (LBP), Local Phase Quantization (LPQ), Weber Local Descriptor (WLD) and Pyramid of Histogram of Oriented Gradients (PHOG). Zavaschi et al. [25] used Gabor filter and Local Binary Pattern as two prominent facial expression feature sets to train their base classifiers. However, to the best of our knowledge, those fusion methods are based on various artificial features. There may be an upper bound of the performance if the features are extracted based on the self-learning strategy. If we try fusing the artificial feature with the self-learning features, an improved representation can be expected due to the strong representation ability of the CNNs [26].

For the classifier construction, SVM [27, 28] is the most common and effective method. Some modifications have been proposed to extend its applicability. Some other methods have also been proposed to construct classifiers to discriminate different expressions, such as Hidden Markov Model (HMM) [29], Random Forest (RF) [30], Neural Network (NN) [31], K nearest neighbor (KNN) [32] and so on. The classification model based on facial expression image has two important aspects: accuracy and efficiency. The latter is measured in terms of time complexity, computational complexity and space complexity. Studies show that these methods are extremely suitable for facial expression classification.

3 Methodology

This section first presents data augmentation strategies to address the small data quantity and data imbalance. Then, this paper describes the design of different extraction strategies for different modalities to achieve powerful representation.

3.1 Data augmentation

In our proposed CNNs framework, many parameters are to be adjusted and exhibit unequal distribution across the classes in the database. To address the over-fitting and data imbalance problems, the most common method is aggressive data augmentation. We make use of artificially enlarged datasets by using label-preserving transformation. The most common methods are transition and reflection [33]. Details of the strategies are described in the experiment section. Then, we use the augmented data to train our feature extraction and classification model.

3.2 Facial expression feature design

We propose a facial expression representation scheme as illustrated in Fig. 1. Under this framework, we consider two different types of feature for representing facial expressions: empirical feature, which captures the appearance of the most discriminative facial key points for expression recognition, and self-learning feature which is high-level feature extracted by means of CNNs. We also consider a fusion strategy of empirical feature and self-learning feature, which we refer to as the multi-modal facial expression feature. Besides, we strike a balance between recognition performance and efficiency.

3.2.1 Empirical feature of facial expression

In the paper, we use facial key points of each image as the artificial feature for emotion recognition based on facial expression. The changes in facial expression lead to slight different instant changes in individual facial muscles in facial appearance. A face has different rigidness in different areas. Fasel and Luettin [34] proposed that the generation of emotion brings about facial behaviour changes and that they are strongly linked to some specific areas, such as eyebrows, eyes, mouth, nose and tissue textures, rather than the whole face. In order to improve accuracy of recognition and avoid over-fitting, we reduce the feature points to a reasonable number. According to the principles above, we select 59 key points from the eyebrows, eyes, nose and mouth, as shown in Fig. 2. Each feature point is expressed as a 2-dimensional coordinate as follows: (x, y). Therefore, a 118-dimensional facial feature vector $\vec{f}^1$ can be obtained from each frame as follows: $\vec{f}^1 = (x_{1,1}, y_{1,1}, x_{1,2}, y_{1,2}, \dots, x_{1,59}, y_{1,59})$.

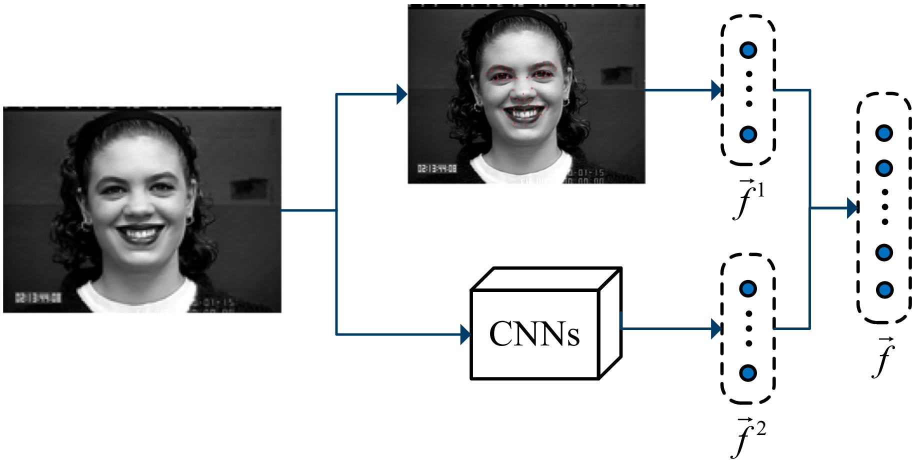

Fig. 1 Flowchart of the proposed multi-modal facial expression representation framework

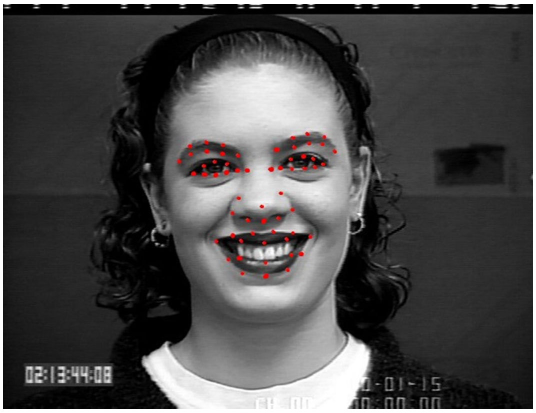

Fig. 2 59 key points of human face

3.2.2 Self-learning feature of facial expression

We design different structures for different modalities as a complex structure contains richer information. CNNs is used to extract high-level self-learning feature from the whole face image. In machine learning, CNNs is a type of feed-forward artificial neural network inspired by animal visual cortex tissue. It is a supervised multilayer neural network capable of extracting prominent and distinguishing feature from facial images. Such feature has desirable generalization abilities for unknown image. Besides, it is also known as shift-invariant or space-invariant artificial neural network (SIANN), mainly due to its special properties, such as weight sharing, local receptive fields, and spatial sub-sampling. Also, the weight sharing architecture can reduce the number of free parameters drastically and increase the generalization performance. In this fashion, feature is extracted from raw images through multiple layers of convolutional filtering and down sampling.

We illustrate the architecture of the CNNs framework for facial expression feature extraction as shown in Table 1. It contains 8 convolutional layers and 4 max-pooling layers. Small filters are utilized for all convolutional layers as small filters together are equivalent of a large filter that can enhance the discriminatory power of the model and reduce the number of filter parameters to learn effectively. Max-pooling can reduce the number of parameters and features, and ensure the invariance of translation, scaling and rotation of features. Therefore, a 120-dimensional facial feature vector $\vec{f}^2$ can be obtained from each frame as follows: $\vec{f}^2 = (z_1, z_2, \ldots, z_{120})$.

4 Experimental evaluation

In this section, extensive experiments on the Extended Cohn–Kanade Dataset (CK+) [35] are conducted to validate the effectiveness of the proposed method. Besides, we use python programs based on Keras and LIBSVM software packages. The data processing platform is a computer with Windows 7, Intel(R) Core(TM) i3-2120 CPU (3.30 GHz,) and 4.00 GB RAM.

Table 1 Details of the feature extraction model architecture for CNN

Name	Type	Input size	Filter number	Filter size
Conv 11	conv	128×96	2	3×3
Conv 12	conv	126×94	4	3×3
Pool 1	max pool	124×92		2×2
Conv 21	conv	62×46	2	3×3
Conv 22	conv	60×44	4	3×3
Pool 2	max pool	58×42		2×2
Conv 31	conv	29×21	4	3×3
Conv 32	conv	27×19	8	3×3
Pool 3	max pool	25×17		2×2
Conv 41	conv	13×9	4	3×3
Conv 42	conv	11×7	8	3×3
Pool 4	max pool	9×5		2×2

4.1 Experimental data

Patrick Lucey et al. presented the CK+ Dataset containing 593 sequences from 123 subjects, who are of both genders and have different cultural and education backgrounds, as shown in Fig. 3.

Each of the sequences incorporates images from onset (neutral frame) to the peak expression (last frame). However, only 327 of the 593 sequences were found to meet one of the criteria for discrete emotions. Therefore, 327 peak frames were selected and labeled.

Despite its popularity, the CK+ database contains limited number of images and subjects, so we enlarge the dataset artificially through label-preserving transformation. Firstly, the image pixel is unified to 160×120. Secondly, 4 patches with the size of 128×96 are cropped from the top left, top right, bottom left and bottom right corner of the image in order to enlarge data. Thirdly, images with label of contempt, fear or sadness are selected and horizontal mirroring is used as data augmentation to balance data, as shown in Fig. 4. Finally, 1592 images together compose the origin facial expression image dataset O. The detailed number of images of each discrete emotion is shown in Table 2.

4.2 Multi-modal facial expression feature

We conduct feature level fusion to obtain a single raw feature vector $\vec{f}$ for each facial image. Specifically, we denote the features extracted as $\vec{f} = \{\vec{f}^1, \vec{f}^2\}$. The feature vector $\vec{f}^1$ contains 2-dimensional coordinate of 59 key points, i.e. $\vec{f}^1 = (x_{1,1}, y_{1,1}, x_{1,2}, y_{1,2}, \ldots, x_{1,59}, y_{1,59})$. And the feature

Fig. 3 Examples of the CK+ database

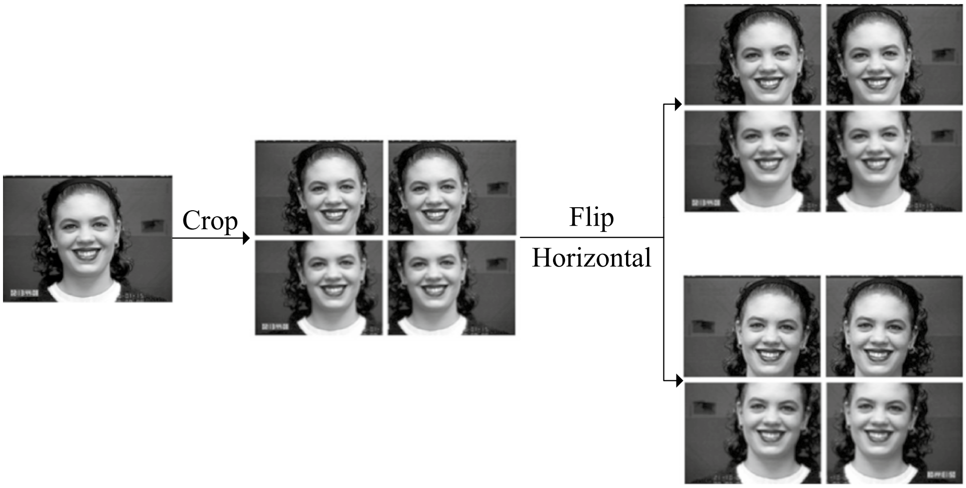

Fig. 4 The flow chart of data augmentation

Table 2 Detailed number of images of each discrete emotion in the dataset O

	Sample set	Training set	Test set
Angry	180	115	65
Contempt	144	115	29
Disgust	236	115	121
Fear	200	115	85
Happy	276	115	161
Sadness	224	115	109
Surprise	332	115	217

vector $\vec{f}^2$ is extracted by the CNNs framework, which is $\vec{f}^2 = (z_1, z_2, \dots, z_{120})$.

According to the principles above, the sample set S contains the feature vectors of 1592 facial expression images with seven discrete emotions. We adopt the stratification sampling method to generate the training set and test set. First, we divide the sample set S into 7 disjoint sets by the type of emotion. Then, we select the same number of feature vectors from each layer randomly and independently. The number is determined by the size of the minimal emotion set. In this paper, we set the number to 115 which is 80% of the size of the contempt emotion set. At last, all these selected feature vectors together compose the training set T with a size of 805, while the rest of the feature vectors together compose the test set V with a size of 787. The detailed number of feature vectors of each emotion is shown in Table 2.

4.3 Experiment contrast with different features

Finally, our experiment employs SVM due to its outstanding performance. In this study, we apply SVM by using a freely available package called LIBSVM with radial basis function (RBF) kernel. In this experiment, we improve the performance of the proposed multi-modal feature. For comparison, three kinds of feature vectors are adopted: low-level facial key point feature vector $\vec{f}^1$, high-level self-learning feature vector $\vec{f}^2$ and multi-modal feature vector $\vec{f}$. From the above, we built up three facial expression recognition models based on SVM which adopt three kinds of feature vectors. Same data sets are adapted to train and test the three models respectively. The numbers of correctly recognized facial expressions under three kinds of feature vectors are shown in Table 3. The average recognition rate of the model using the multi-modal feature is 93% is higher than that of the model using single modal feature, as shown in Table 3. Besides, data wrongly recognized under the feature vectors

Table 3 Number and average recognition rate of correctly recognized facial expression under three kinds of features

Emotion	Test set	$\vec{f}^1$	$\vec{f}^2$	$\vec{f}$
Angry	65	48	55	62
Contempt	29	12	17	24
Disgust	121	115	116	118
Fear	85	61	67	76
Happy	161	143	149	156
Sadness	109	73	87	98
Surprise	217	193	199	208
Average Recognition Rate	–	81.96%	87.67%	94.16%

Table 4 Average recognition rate of different feature extraction strategy

Feature extraction strategy	Average recognition rate (%)
LBP-TOP,CNN + SVM [36]	93
LBP-SP + SVM [37]	94.14
AUDN + SVM [38]	93.70
AlexNet + SVM [39]	92.94
Our strategy	94.41

$\vec{f}$ fall within data wrongly recognized under the feature vector $\vec{f}^1$ or $\vec{f}^2$. It is clear that the proposed multi-modal facial expression representation algorithm combines the advantages of the two kinds of single modal features and thus can achieve significantly outstanding results.

4.4 Comparison with other feature extraction strategies

To evaluate the efficiency of our feature extraction strategy we compared it to other feature extraction strategies set on the same database with the same classification method. Specifically, experimental results are not comparable across the different hardware and parameters. However, experimental results could partly reflect the feasibility of the proposed methods. Results in Table 4 show that the recognition rate of our model is superior to the other feature extraction strategies. By analyzing the structure of these models, it is shown that our model is more powerful for its small-scale and few parameters.

5 Conclusion

This paper proposes a new facial expression feature extraction strategy for emotion recognition. Firstly, it extracts both empirical feature and self-learning feature from facial expression images. Secondly, it enhances the ability of CNNs through data augmentation and small filters. Thirdly, it presents a linear feature level fusion strategy to obtain the final emotion feature. Extensive experiment results on the three kinds of feature vectors suggest that the approach based on the multi-modal feature vector has good performance in terms of the emotion recognition rate. In addition, the proposed approach also shows a very good performance when dealing with the dataset with many restrictions. Emotion recognition based on facial expression is still a challenging task in the real world. In the future, we will explore the multi-modal feature strategy that is applicable to poor-quality facial images without the whole key facial areas.

Acknowledgements This work was supported by the National Natural Science Foundation of China (No. 61573066, No. 61327806).

References

1. Zhao XM, Zhang SQ (2016) Outcome facial expression recognition: feature extraction and classification. IETE Tech Rev 33(5):505–517
2. Ding CX, Tao DC (2015) Robust face recognition via multimodal deep face representation. IEEE Trans Multimed 17(11):2049–2058
3. Shi XS, Guo ZH, Nie FP, Yang L, You J, Tao DC (2015) Neutral face classification using personalized appearance models for fast and robust emotion detection. IEEE Trans Image Process 24:2701–2711
4. Taigman Y, Yang M, Ranzato M, Wolf L (2014) Deepface: closing the gap to human-level performance in face verification. In: IEEE conference on computer vision and pattern recognition (CVPR), IEEE,Columbus, pp 1701–1708
5. Sun Y, Chen Y, Wang X, Tang X (2014) Deep learning face representation by joint identification-verification. In: Annual conference on neural information processing systems, NIPS, 2014, pp 1988–1996
6. Zhang FF, Yu YB, Mao QR, Gou JP, Zhan YZ (2016) Pose-robust feature learning for facial expression recognition. Front Comput Sci 10:832–844
7. Zhao GY, Pietikäinen M (2007) Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Trans Pattern Anal Mach Intell 29(6):915–928
8. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: IEEE computer society conference on computer vision and pattern recognition(CVPR), IEEE,San Diego, 2005, pp 886–893
9. Jones JP, Palmer LA (1987) An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. J Neurophysiol 58(6):1233–1258
10. Ren FJ, Huang Z (2015) Facial expression recognition based on AAM-SIFT and adaptive regional weighting. IEEJ Trans Electr Electron Eng 10(6):713–722
11. Gao XB, Su Y, Li XL, Tao DC (2010) A review of active appearance models. IEEE Trans Syst Man Cybern Part C Appl Rev 40(2):145–158
12. Sung JW, Kanada T, Kim DJ (2007) A unified gradient-based approach for combining ASM into AAM. Int J Comput Vis 75(2):297–310

13. Chen SZ, Wang HP, Xu F, Jin YQ (2016) Target classification using the deep convolutional networks for SAR images. IEEE Trans Geosci Remote Sens 58(8):4806–4817
14. Xu J, Luo XF, Wang GH, Gilmore H, Madabhushi A (2016) A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images. Neurocomputing 191:214–223
15. Chowdhury A, Kautz E, Yener B, Lewis D (2016) Image driven machine learning methods for microstructurerecognition. Comput Mater Sci 123:176–187
16. Shi BG, Bai X, Yao C (2016) Script identification in the wild via discriminativeconvolutional neural network. Pattern Recognit 52:448–458
17. Yu ZD, Zhang C (2015) Image based static facial expression recognition with multiple deep network learning. In: ACM international conference on multimodal interaction, ACM, Seattle, 2015, pp 435–442
18. Song H (2017) Facial expression classification using deep convolutional neural network. J Broadcast Eng 11:162–172
19. Shi J, Zhou SC, Liu X, Zhang Q, Lu MH, Wang TF (2016) Stacked deep polynomial network based representationlearning for tumor classification with small ultrasound imagedataset. Neurocomputing 194:87–94
20. Ranzato M, Susskind J, Mnih V, Hinton G (2011) On deep generative models with applications to recognition. In: IEEE conference on computer vision and pattern recognition (CVPR), IEEE, Colorado Springs, 2011, pp 2857–2864
21. Rifai S, Bengio Y, Courville A, Vincent P, Mirza M (2012) Disentangling factors of variation for facial expression recognition. In: European conference on computer vision (ECCV), Springer, Florence, 2012, pp 808–822
22. Panagakis Y, Nicolaou MA, Zafeiriou S, Pantic M (2016) Robust correlated and individual component analysis. IEEE Trans Pattern Anal Mach Intell 38(8):1665–1678
23. Luo Y, Zhang T, Zhang Y (2016) A novel fusion method of PCA and LDP for facial expressionfeature extraction. OPTIK 127(2):718–721
24. Turan C, Lam KM, He XJ (2015) Facial expression recognition with emotion-based feature fusion. In: Asia-Pacfic signal and information processing association annual summit and conference (APSIPA), IEEE, Hong Kong, 2015, pp 1–6
25. Zavaschi THH, Britto AS, Oliveira LES, Koerich AL (2013) Fusion of feature sets and classifiers for facial expression recognition. Expert Syst Appl 40(2):646–655
26. Krizhevsky A, Sutskever I, Hinton G (2012) Imagenet classification with deep convolutional neural networks. In: Annual conference on neural information processing systems, NIPS, Lake Tahoe, 2012, pp 1097–1105
27. Burges CJC (1998) A tutorial on support vector machines for pattern recognition. Data Min Knowl Discov 2(2):121–167
28. Zhong L, Liu Q, Yang P, Liu B, Huang J, Metaxas D (2012) A review on facial patches for expression analysis. IEEE conference on computer vision and pattern recognition (CVPR), IEEE, Providence, 2012, pp 2562–2569
29. Cohen I, Sebe N, Garg A, Chen LS, Huang TS (2003) Facial expression recognition from video sequences: temporal and static modeling. Comput Vis Image Underst 91(1–2):160–187
30. Breiman L (2001) Random forests. Mach Learn 45(1):5–32
31. Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid sceneanalysis. IEEE Trans Pattern Anal Mach Intell 20(11):1254–1259
32. Caruana R (1997) Multitask learning. Chine Learn 28(1):41–75
33. Liu ZH, Wang HZ, Yan Y, Guo GJ (2015) Effective facial expression recognition via the boosted convolutional neural network. Commun Comput Inf Sci 546:179–188
34. Fasel B, Luettin J (2003) Automatic facial expression analysis: a survey. Pattern Recognit 36(1):259–275
35. Patrick L, Jeffrey FC, Takeo K, Jason S, Zara A (2010) The extended Cohn–Kanade dataset (CK+): a complete dataset for action unit and emotion-specified expression. In: IEEE computer society conference on computer vision and pattern recognition-workshops, IEEE, San Francisco, 2010, pp 94–101
36. Happy SL, Routray A (2015) Automatic facial expression recognition using features of salient facial patches. IEEE Trans Affect Comput 6(1):1–12
37. Happy SL, Routray A (2015) Automatic facial expression recognition using features of salient facial patches. IEEE Trans Affect Comput 6(1):1–12
38. Liu MY, Li SX, Shan SG, Chen XL (2015) AU-inspired deep networks for facial expression feature learning. Neurocomputing 159:126–136
39. Vo DM, Sugimoto A, Le TH (2016) Facial expression recognition by re-ranking with global and local generic features. In: 23rd international conference on pattern recognition, IEEE, New York, 2016, pp 4118–4123

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.